



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO – UNIRIO
Centro de Ciências Humanas e Sociais – CCH
Programa de Pós-Graduação em Gestão de Documentos e Arquivos – PPGARQ
Mestrado Profissional em Gestão de Documentos e Arquivos – MPGA

Carolina de Oliveira

Um estudo de caso sobre *datasets* do Ministério da Justiça: dados brutos ou documentos arquivísticos?

Rio de Janeiro

2015

Carolina de Oliveira

Um estudo de caso sobre *datasets* do Ministério da Justiça: dados brutos ou documentos arquivísticos?

Produto técnico-científico apresentado ao Programa de Pós-Graduação em Gestão de Documentos e Arquivos, da Universidade Federal do Estado do Rio de Janeiro – Unirio – como requisito parcial para obtenção do título de Mestre.

Área de concentração: Gestão de arquivos na Arquivologia contemporânea.

Linha de pesquisa: Gestão da informação arquivística.

Orientador: Prof. Dr. Daniel Flores

Rio de Janeiro

2015

Oliveira, Carolina de
OL48 Um estudo de caso sobre *datasets* do Ministério da Justiça: dados brutos ou documentos arquivísticos? / Carolina de Oliveira. - 2015.
115 p. : il. ; 30cm. Revisão gramatical: Iraneide A. Fontes Cidreira.

Dissertação (Mestrado Profissional em Gestão de Documentos e Arquivos) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

Orientador: Prof. Dr. Daniel Flores

1. Arquivologia 2. Gestão de documentos 3. Gestão de dados 4. Dados governamentais abertos. Governo Aberto 5. Ministério da Justiça I. Flores, Daniel II. Título.

CDD 025.21

Carolina de Oliveira

Um estudo de caso sobre *datasets* do Ministério da Justiça: dados brutos ou documentos arquivísticos?

Produto técnico-científico apresentado ao Programa de Pós-Graduação em Gestão de Documentos e Arquivos, da Universidade Federal do Estado do Rio de Janeiro – Unirio – como requisito parcial para obtenção do título de Mestre.

Área de concentração: Gestão de arquivos na Arquivologia contemporânea.

Linha de pesquisa: Gestão da informação arquivística.

Aprovado em: ___/___/2015

BANCA EXAMINADORA

Prof. Dr. Daniel Flores
Universidade Federal do Estado do Rio de Janeiro – UNIRIO
Orientador

Profa. Dra. Ana Celeste Indolfo
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

Profa. Dra. Gisele da Silva Craveiro
Universidade de São Paulo – USP

Profa. Dra. Aline Lopes de Lacerda
Universidade Federal do Estado do Rio de Janeiro – UNIRIO
Suplente

Profa. Dra. Fernanda Baião
Universidade Federal do Estado do Rio de Janeiro – UNIRIO
Suplente

Rio de Janeiro, 21 de agosto de 2015.

Dedicado à sociedade brasileira e aos
arquivistas atuantes nas esferas públicas.

AGRADECIMENTOS

À minha família, por serem minha base sólida no plano terreno e a Deus, por me inspirar na busca pelo discernimento de qual caminho seguir.

Aos meus amiguinhos Raone, Zé Márcio, Brenda, Cláudia Teresa, Antônio Laurindo e Bia que tiveram a paciência de ouvir meu chororô ao longo desses últimos quatro semestres.

À Claudia Lacombe e Rosely C. Rondinelli por, generosamente, concederem uma parte de seu tempo para compartilharem comigo o conhecimento e a experiência que têm sobre Arquivologia e documentos arquivísticos digitais. Fizeram a diferença nessa etapa da minha vida acadêmica.

À Autoridade Pública Olímpica, nas pessoas de Rafael Lucindo e Iggor Costa, que acompanharam essa etapa da minha vida profissional.

Ao professor Sérgio Albite pelas palavras certas, nas horas certas e ao meu orientador, prof. Dr. Daniel Flores.

Aos meus colegas de turma Zenaide, Leandro, Bianca, Karina, Duala, Raquel, Fabrício, Kíssila e Vanessa: se não a melhor turma, o melhor grupo! Agora, sim: “- Vamos para Paquetá”!

[...] sob condições de dominação autoritária, os ideais liberais clássicos, que são também expressos por Marx, Bakunin e por todos os verdadeiros revolucionários, não podem ser realizados. Em outras palavras, os seres humanos não serão livres para informar-se e criar, para desenvolver a plenitude de suas potencialidades; o trabalhador continuará a ser um fragmento de uma ferramenta humana degradada no processo produtivo dirigido de cima para baixo.

(O governo no futuro, de Noam Chomsky)

RESUMO

No século XX, ao dado digital, especialmente, tem sido agregado valor por ele ser a base de informações geradas num ambiente laboral e pelo seu reuso para outros fins. Esse reuso de dados tem sido reconhecido por governos de diferentes continentes como um recurso para dar transparência a suas ações, para a participação e colaboração dos cidadãos na definição de políticas públicas. Parte da premissa de que dados governamentais estão contidos em documentos arquivísticos e analisa se um *dataset* publicado numa seção específica do sítio do Ministério da Justiça é um dado bruto ou um documento arquivístico, pois demandam diferentes procedimentos para sua gestão e preservação. Aborda a gestão de documentos por compreender a necessidade de ampliação de seu conceito técnico para o também político-administrativo. Reflete sobre as fases da gestão de documentos auxiliarem no entendimento do macroprocesso da gestão de dados estudada pela Computação e averigua semelhanças, mesmo que os objetos de estudos sejam diferentes: o documento arquivístico pela Arquivologia e o dado pela área de Computação. Adota o método do estudo de caso nesta pesquisa de natureza aplicada, abordagem qualitativa e objetivo exploratório e utiliza a pesquisa bibliográfica, a documental, a observação direta e o Serviço de Informação ao Cidadão, do Poder Executivo federal, para a coleta de dados. Chega ao resultado de que o *dataset* é um dado bruto. Confirma que os documentos arquivísticos também são fontes de dados governamentais. Faz a proposta de um instrumento de *benchmark* com linhas de base para identidade e integridade de documentos arquivísticos digitais, contribuindo para a produção de documentos autênticos. Demonstra que a gestão de documentos de acordo com os princípios, conceitos e métodos da Arquivologia apoia a iniciativa de Governo Aberto aderida pelo governo brasileiro.

Palavras-chaves: Arquivologia. Gestão de documentos. Gestão de dados. Dados governamentais abertos. Ministério da Justiça.

ABSTRACT

In the 20th century, value has been aggregated for digital data for the business transactions, not only for the use of data but for the reuse of it too. The reuse of data has been adopted by governments around the world as an instrument to increase public transparency, accountability and citizen participation at public policy decisions. For this research, the premise is that government data is embedded in records created by the government bodies, so that the main purpose is to analyze a dataset published at Brazilian Ministry of Justice's website is a data or a record, that may require different actions and procedures for management and preservation of each one. For the theoretical review in this research, the Brazilian records management concept is being discussed because of the necessity of increasing its focus on political and administrative aspects aside from technical practices. Another point is data management, an object of study by Computer Science. The macro process of data management has similarities to records management macro process. The research methodology adopted is the case study for an exploratory purpose. The data collection that is chosen, for instance: bibliographical references, documents, direct observation and the Serviço de Informação ao Cidadão (SIC) – the Brazilian information system for public information access – from federal Executive Power. Based on the analysis of data collected, the result is that Ministry of Justice's dataset is data. Beyond that, it is possible to confirm that records are also one of the sources of government data, so a benchmark instrument is suggested to apply record's creation phase with the purpose to support authentic records. Finally, this research demonstrates that records management based on principles, concepts and methods from Archival Science stands for the Open government initiatives agreed by the Brazilian government.

Key-words: Archival Science. Records management. Data management. Open government data. Ministry of Justice.

LISTA DE ILUSTRAÇÕES

Figura 1 – A gestão arquivística de documentos apoiando a disponibilização de dados e informações	25
Figura 2 – Conexão entre dado, informação, documento e documento arquivístico ...	34
Figura 3 – Representação de dado, informação e documento em ambiente digital	36
Figura 4 – Macroprocesso da gestão de dados	48
Figura 5 – Atividades englobadas pela Gestão de Dados	49
Figura 6 – Macroprocesso da gestão de documentos	51
Figura 7 – Um arcabouço conceitual sobre a qualidade de dado	53
Figura 8 – Linha do tempo de algumas ações do GOV.BR	69
Figura 9 – Membros do CIGA, da Parceria para Governo Aberto	70
Figura 10 – Membros do GE, do CIGA	71
Figura 11 – Membros do Comitê Gestor da INDA	78
Figura A-1 – Níveis de um objeto digital	90

LISTA DE QUADROS

Quadro 1 – Conceitos de dados e suas semelhanças	28
Quadro 2 – Características e partes constituintes do documento arquivístico digital ...	39
Quadro 3 – Qualidades de um documento arquivístico digital	42
Quadro 4 – Equiparação dos macroprocessos de gestão de dados e documentos arquivísticos	52
Quadro 5 – Qualidade de dados: categoria e fatores	53
Quadro 6 – Dimensões de qualidade de dados	54
Quadro 7 – Equiparação de dimensões de qualidade de dados	55
Quadro 8 – Qualidade de dado e de documento arquivístico digital	56
Quadro 9 – Membros do CIGA e do GE do Governo Aberto, e do Comitê Gestor da INDA	79
Quadro 10 – Campos do <i>dataset</i> “Lista de terceirizados do Ministério da justiça”	96
Quadro 11 – Proposta de instrumento de <i>benchmark</i>	103

LISTA DE SIGLAS

AN – Arquivo Nacional

APF – Administração Pública Federal

CDI/MJ – Coordenação de Documentação e Informação, do Ministério da Justiça

CGMA/MJ – Coordenação-Geral de Modernização e Administração, do Ministério da Justiça

CGRH/MJ – Coordenação Geral de Recursos Humanos, do Ministério da Justiça

CGTI/MJ – Coordenação Geral de Tecnologia da Informação, do Ministério da Justiça

CGU – Controladoria-Geral da União

CPAD – Comissão Permanente de Avaliação de Documentos

CTDE – Câmara Técnica de Documentos Eletrônicos

DAMA-DMBOK – Data Management/ Data Management Book of Knowledge

ERIC – Educational Resources Information Center

eSIC – Sistema eletrônico do Serviço de Informação ao Cidadão

GE – Grupo Executivo, do Comitê Interministerial para Governo Aberto

GM/MJ – Gabinete do Ministro, do Ministério da Justiça

GOV.BR – Programa de Governo Eletrônico brasileiro

IBM – International Business Machine

INDA – Infraestrutura Nacional de Dados Abertos

InterPARES – International Project on Permanent Authentic Records Electronic Systems

LAI – Lei de Acesso à Informação

MJ – Ministério da Justiça

MP – Ministério do Planejamento, Orçamento e Gestão

OA – Open Access

OGP – Open Government Partnership

OKF – Open Knowledge Foundation

PDA/MP – Plano de Dados Abertos, do Ministério do Planejamento, Orçamento e Gestão

PDAE/MJ – Plano de Dados Abertos e Espaciais, do Ministério da Justiça

PEN – Processo Eletrônico Nacional

PI – Sistema Protocolo Integrado

SE/MJ – Secretaria-Executiva, do Ministério da Justiça

SIC – Serviço de Informação ao Cidadão

SIGA – Sistema de Gestão de Arquivos

SIGAD – Sistema Informatizado de Gestão Arquivística de Documentos

SINAR – Sistema Nacional de Arquivos

SLTI – Secretaria de Logística e Tecnologia da Informação

TICs – Tecnologias de Informação e Comunicação
Unicamp – Universidade Estadual de Campinas

SUMÁRIO

1	INTRODUÇÃO	15
2	GESTÃO DE DOCUMENTOS	20
2.1	DADO, INFORMAÇÃO E DOCUMENTO: UMA CONEXÃO POSSÍVEL	25
2.2	O DOCUMENTO ARQUIVÍSTICO DIGITAL COMO FONTE DE DADOS DIGITAIS	37
2.2.1	As qualidades de um documento arquivístico digital	41
2.3	REPOSITÓRIO DIGITAL CONFIÁVEL PARA DOCUMENTOS ARQUIVÍSTICOS	43
3	GESTÃO DE DADOS	47
3.1	O DADO DIGITAL E ALGUMAS DE SUAS QUALIDADES	52
4	OPEN DATA = DADOS ABERTOS	59
4.1	DADOS GOVERNAMENTAIS ABERTOS	63
4.2	PARTICIPAÇÃO CIDADÃ NAS DECISÕES GOVERNAMENTAIS	65
5	O PROJETO “DADOS ABERTOS GOVERNAMENTAIS”, DO PODER EXECUTIVO FEDERAL BRASILEIRO	68
5.1	LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011, A “LEI DE ACESSO À INFORMAÇÃO PÚBLICA”	72
5.2	MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO (MP)	74
5.3	INFRAESTRUTURA NACIONAL DE DADOS ABERTOS (INDA)	77
5.4	“PORTAL BRASILEIRO DE DADOS ABERTOS”	80
5.5	MINISTÉRIO DA JUSTIÇA (MJ)	82
6	METODOLOGIA	84
7	ANÁLISE DIPLOMÁTICA DO DATASET “LISTA DE TERCEIRIZADOS DO MINISTÉRIO DA JUSTIÇA”	88
8	ANÁLISE E DISCUSSÃO DOS RESULTADOS	94
9	PROPOSTA DE INSTRUMENTO DE BENCHMARK	102
10	CONSIDERAÇÕES FINAIS	106
	REFERÊNCIAS	110
	ANEXO A – INTERPARES PROJECT 2: DIPLOMATIC ANALYSIS TEMPLATE	120
	ANEXO B – INTERPARES PROJECT 2: BENCHMARK REQUIREMENTS SUPPORTING THE PRESUMPTION OF AUTHENTICITY OF ELECTRONIC RECORDS	124
	ANEXO C – INTERNATIONAL RECORDS MANAGEMENT TRUST: GOVERNO ABERTO E DOCUMENTOS ARQUIVÍSTICOS CONFIÁVEIS – INSTRUMENTO DE REFERÊNCIA PARA MARCOS REGULATÓRIOS/INSTITUCIONAIS E CAPACIDADES	127

1 INTRODUÇÃO

O Poder Executivo federal tem avançado nas iniciativas para a disponibilização de informações públicas ao cidadão brasileiro. Dentre suas ações está o Projeto “Dados Abertos Governamentais”, cujo produto é o Portal Brasileiro de Dados Abertos. No entanto, cabe ressaltar que a gestão arquivística de documentos também contribui para que informações públicas estejam acessíveis e, sobretudo, sejam autênticas e confiáveis.

Em uma administração pública, a adoção de recursos de tecnologias da informação e comunicação (TICs) proporciona celeridade aos processos de negócio. No ano de 2000, foi oficializado o Programa de Governo Eletrônico brasileiro (GOV.BR), tendo como pilares a melhoria do gerenciamento interno da administração pública, da prestação de serviços aos cidadãos e do relacionamento com fornecedores.

Dentre os projetos integrantes do Programa GOV.BR, encontra-se o Projeto “Dados Abertos Governamentais”. Em seu escopo, ele é visto como uma *metodologia* utilizada para publicação e reúso dos dados governamentais pelos cidadãos, com o objetivo de incentivar a participação da sociedade brasileira nas ações governamentais no que se refere à transparência, melhoria dos serviços públicos, definição de políticas públicas, dentre outras necessidades e interesses da sociedade.

Ao longo de sua implantação, o Projeto adquiriu relevância no âmbito do Poder Executivo federal por também ser um recurso no cumprimento da Lei de Acesso à Informação, a Lei nº 12.527, de 18 de novembro de 2011 – a LAI. Ao objetivo dessa lei acrescenta-se a Parceria para Governo Aberto – uma iniciativa internacional, lançada também em 2011, de adesão voluntária, que busca o compromisso dos governos com relação à transparência na administração pública, ao combate à corrupção e à prestação de contas aos cidadãos. Partindo desse mesmo enfoque, toma corpo outra iniciativa, a dos dados governamentais abertos.

O dar a transparência governamental tem demandado dos órgãos públicos uma atuação diversa daquela em que o secreto era a regra, especialmente em países de regime democrático pós um período de ditadura, como o Brasil. Porém, o cumprimento da LAI e das ações da Parceria para Governo Aberto demandam novo comportamento dos agentes do fronte governamental.

É pertinente salientar que os ideais de “governo eletrônico” e “governo aberto”, ainda que minimamente, se diferenciam. No primeiro, prevalece o uso das TICs, como os sítios na Internet, para transparência das ações governamentais; já no segundo, também através do uso das TICs, o que se pretende é a colaboração cidadã na construção das ações e políticas de governos.

Em ambas iniciativas, de governo eletrônico e de governo aberto, ao cidadão é oferecido um “lugar” para inserir-se nas discussões governamentais, a fim de apontar sua visão para políticas públicas apropriadas, no entanto, qual o nível de colaboração da sociedade civil nas decisões governamentais faz-se relevante conhecer. Um caminho para isso são os dados governamentais abertos.

O tema dos dados governamentais abertos (ou dados abertos governamentais), delimitado para esta pesquisa, envolve a publicação de dados/informações governamentais sem restrições a quem tem interesse, utilizando principalmente as TICs, salvo situações de sigilo ou em referência a pessoas. O diferencial dessa iniciativa está no reuso dos dados coletados, ou seja, são utilizados para outros fins, principalmente por meio do desenvolvimento de aplicativos. Daí se poder falar da existência de um potencial empoderamento da sociedade, já que a sua participação e controle das ações governamentais podem ser estabelecidos através de uma rede colaborativa entre governo, sociedade civil e empresas privadas, em consonância com a iniciativa de governo aberto.

A fim de tornarem-se mais transparentes, participativos e colaborativos, os governos dos EUA, Reino Unido, Canadá e Nova Zelândia foram os primeiros a adotarem dados abertos, trazendo a temática a uma posição de destaque a partir dos anos de 2009 (OPEN KNOWLEDGE FOUNDATION, 2012) e, desde então, vem afirmando-se nas agendas políticas de diferentes governos, inclusive o brasileiro.

Para que qualquer pessoa tenha condições plenas de reutilizar os dados governamentais publicados na Internet pelos órgãos públicos responsáveis, faz-se necessário que os *datasets*¹ tenham qualidades para serem abertos, com acesso sem custos e fontes de origem confiáveis.

1 *Dataset* é um conjunto de dados estruturados ou semiestruturados. Nesta pesquisa, opta-se por adotar o termo em inglês porque facilita a compreensão do objeto e, na literatura brasileira de Computação e Informática, seu uso é recorrente.

Após a qualificação do projeto desta pesquisa, em novembro de 2014, reavaliou-se qual seria o seu objeto de investigação e chegou-se ao *dataset* – anteriormente, a abordagem havia sido o Projeto “Dados Abertos Governamentais” propriamente dito. Devido a essa restrição na proposta inicial de estudo, a partir de então, parte-se da premissa de que os documentos arquivísticos digitais também são fontes de dados governamentais.

A mudança do objeto de pesquisa, por sua vez, deu origem a uma nova problemática: como identificar se um *dataset* publicado na Internet é um dado bruto ou um documento arquivístico? Na busca por essa resposta, o objetivo geral da investigação passou a ser analisar se o *dataset* publicado na seção destinada aos dados governamentais abertos, no sítio do Ministério da Justiça (MJ), é um dado bruto ou um documento arquivístico.

A seleção do MJ como campo empírico fez-se por ele ser o órgão ao qual se vincula o Arquivo Nacional (AN), instituição arquivística onde a pesquisadora atua profissionalmente, correspondendo, portanto, à proposta de mestrados profissionais: o desenvolvimento de pesquisas científicas cujos produtos técnicos são aplicados ao mercado laboral. Uma conciliação da teoria com a prática.

Para atingir o objetivo geral, mais especificamente, houve a necessidade de:

- selecionar os *datasets* disponibilizados pelo MJ sobre funcionários terceirizados;
- identificar quais qualidades são inerentes a um dado;
- aplicar um modelo de análise diplomática ao *dataset*; e
- propor a criação de um instrumento de *benchmark* como linha de base para a identidade e integridade de documentos arquivísticos digitais, independentemente de TICs, utilizados como uma das fontes para formação de *datasets*.

Cabe ressaltar que a identificação de um *dataset* como um dado bruto ou um documento arquivístico é relevante para decisões relacionadas ao tratamento a ser estabelecido para tais objetos, principalmente em ambiente digital.

A partir do entendimento de que um dado bruto está contido nos documentos arquivísticos – os quais exigem, para sua gestão e preservação, os princípios da Arquivologia, a fim de que permaneçam orgânicos, únicos, autênticos, confiáveis e acessíveis pelo tempo que se fizerem necessários –, pode-se afirmar que será com base na gestão de documentos que tais dados também terão suas qualidades mantidas.

Como o registro das ações governamentais para garantir direitos aos cidadãos

brasileiros tem sido feito, sobretudo por meio do ambiente digital, esta investigação sobre dados governamentais abertos, gestão de dados e gestão de documentos justifica-se pela escassez de estudos acerca do tema no cenário arquivístico brasileiro.

De acordo com a proposta de estudos aqui desenvolvida, uma gestão adequada dos dados governamentais a serem abertos à sociedade e dos documentos arquivísticos, como uma de suas fontes, manterá as qualidades essenciais para o seu reuso e, assim, o controle por parte da sociedade, garantindo que a governança, a transparência e *accountability*² das ações governamentais sejam cumpridas de forma legítima no atendimento à demanda dos cidadãos brasileiros.

No que se refere à estrutura da dissertação, além desta Introdução, apresentam-se três seções para reflexão teórica, uma seção dedicada ao campo empírico e outra à apresentação da metodologia e análise dos dados coletados. Como produto técnico-científico, desenvolve-se a proposta de criação de um instrumento de *benchmark* para identidade e integridade de documentos arquivísticos digitais. Por fim, são feitas as considerações sobre a pesquisa realizada.

Na seção 2, propõe-se uma revisitação do termo “gestão de documentos”, adotado no âmbito da Arquivologia, mas originário da Administração e caracterizado por uma abordagem mais técnica do que político-administrativa. A ampliação do conceito envolvendo política, gestão e preservação de documentos arquivísticos favorece, principalmente, aqueles produzidos e mantidos em ambiente digital.

Na seção 3, desenvolve-se uma discussão sobre o que é a gestão de dados e quais qualidades lhes são pertinentes. A qualidade de dados é uma temática recorrente em pesquisas, mormente da área de Computação. Os documentos arquivísticos também se caracterizam por qualidades específicas estudadas na Arquivologia. Partindo-se da premissa de que os primeiros estão contidos nos últimos, ilumina-se um debate entre essas duas áreas do conhecimento para que se complementem.

Na seção 4, para abordar a temática dos dados abertos e dados governamentais abertos, faz-se uma visita, em vez de uma revisitação, da literatura existente, já que ela ainda era vista com pouca intimidade por parte da pesquisadora. Assim, discorre-se sobre a

2 “[...] podemos afirmar ainda que *accountability* nasce com a assunção por uma pessoa da responsabilidade delegada por outra, da qual se exige a prestação de contas, sendo que a análise dessas contas pode levar à responsabilização [...]”. PINHO, José Antonio Gomes de; SACRAMENTO, Ana Rita Silva. *Accountability: já podemos traduzi-la para o português?* Disponível em: <<http://www.scielo.br/pdf/rap/v43n6/06.pdf>>.

origem da iniciativa para dados abertos, a noção adotada pelos estudiosos para sua compreensão e a motivação dos países de diferentes continentes na adoção de tais recursos, no sentido de dar transparência às ações governamentais.

Na seção 5, apresenta-se o campo empírico pesquisado: o Ministério da Justiça, do Poder Executivo federal. Além dele, o Ministério do Planejamento, Orçamento e Gestão (MP), a Lei de Acesso à Informação (lei nº 12.527, de 18/11/2011), a Infraestrutura Nacional de Dados Abertos e o Portal Brasileiro de Dados Abertos também estão contemplados, tendo em vista a influência direta de suas ações na publicação de dados governamentais abertos.

A seção 6, por sua vez, faz o detalhamento de como esta pesquisa, de natureza aplicada, abordagem qualitativa e objetivo exploratório, foi desenvolvida. Além disso, a análise dos dados coletados também é apresentada nesta seção.

Na seção 7, propõe-se a criação de um instrumento de *benchmark*, cuja finalidade seria indicar linhas de base para aprimorar o desempenho de uma organização, pública ou privada, em determinada área ou atividade específica, como no caso da publicação de *datasets* com qualidade para uso e reúso do consumidor de dados. Nessa proposta, as linhas de base são para identidade e integridade de documentos arquivísticos digitais, que apoiarão a qualidade de um *dataset*.

Por fim, a seção 8 destina-se às considerações finais sobre a temática de dados abertos e a contribuição da gestão de documentos, sob a ótica da Arquivologia, para que *datasets* sejam publicados com qualidades de uso e reúso pelos consumidores de dados.

2 GESTÃO DE DOCUMENTOS

A afirmação da Arquivologia como ciência traz benefícios, tanto para essa área do conhecimento quanto para a sociedade na qual se insere, porque os resultados das pesquisas científicas possibilitam outras reflexões ou mesmo revisões e mudanças do que era aceito como consenso ou verdade imutável.

O ato de registrar uma ação praticada remonta às antigas civilizações localizadas no Oriente Médio. Depreende-se que a intenção era deixar para gerações futuras os atos e o conhecimento gerado em diferentes épocas e contextos sociais. De acordo com Silva (2009, p. 45) “[...], A importância da escrita para a actividade humana levou [...] à consciência de que era preciso conservar tais registos, tendo em vista uma posterior utilização”.

As informações geradas pela humanidade até a contemporaneidade têm sido registradas em diferentes suportes e formas que, à medida que novas tecnologias vão sendo descobertas e desenvolvidas, também passam por mudanças.

O uso de tecnologias para produção, uso, manutenção, preservação e acesso de documentos em ambiente digital estimulou revisitações sobre o campo teórico da Arquivologia. Na gestão de documentos, por exemplo, passaram a ser implantados sistemas informatizados para o cumprimento de sua finalidade, nos quais

[...] é essencial a inclusão de requisitos de natureza arquivística desde o momento da concepção desses sistemas, tornando-se imprescindível o cuidadoso controle ao longo de todo o seu ciclo vital, de modo a garantir a autenticidade, a fidedignidade³, a integridade e a acessibilidade. (INDOLFO, 2013, p. 54).

O conceito de gestão de documentos, nos anos de 1940, foi cunhado com uma “visão mais administrativa do que arquivística” em países de origem anglo-saxônica. Na mesma década, Philip C. Brooks⁴ acrescenta ao conceito o ciclo de vida dos documentos, que, a partir de então, passa a ser difundido e aplicado nas práticas de tratamento. (INDOLFO, 2013, p. 44-46).

3 O termo fidedignidade é adotado por alguns autores brasileiros como tradução para *reliability*. Após uma revisão conceitual pela Câmara Técnica de Documentos Eletrônicos (CTDE), do Conarq, o termo sinônimo adotado é confiabilidade.

4 Philip Coolidge Brooks (1906-1977) foi arquivista, gestor de documentos e professor, atuando no Arquivo Nacional, na *Society of American Archivists* e na *American University* (Washington D.C.). Disponível em: <<http://www.lib.utexas.edu/taro/utcah/02711/cah-02711.html>>. Acesso em: 25 set. 2014.

Segundo a UNESCO (apud INDOLFO, 2013, p. 48), gestão de documentos é definida como a “área da administração geral com vistas a assegurar a economia e a eficácia das operações desde a criação, manutenção e utilização, até a destinação final dos documentos”.

A gestão de documentos tem por finalidade estudar os fluxos de trabalho para a criação de documentos, racionalizar a produção de documentos, otimizar o trabalho corporativo, definir a guarda permanente ou eliminação de documentos. No entanto, não há um posicionamento único dos especialistas sobre o início e o fim da gestão de documentos, nem sobre as funções que a integram, ocasionando diferentes abordagens acerca do tema. (HEREDIA HERRERA, 2007, p. 92).

Na mesma linha de raciocínio, Indolfo diz que

[...] não se pode falar de gestão de documentos como um conceito único e de aplicação universal, uma vez que na sua elaboração e no seu desenvolvimento contribuíram diversos fatores, em que se destacam a tradição arquivística, e também a administrativa, e um contexto histórico e institucional. Não se pode falar, ainda, de um modelo de gestão de documentos, mas de diversidades que se produziram com as práticas em diferentes países. (INDOLFO, 2013, p. 46).

Na legislação norte-americana, a definição de gestão de documentos engloba, além do ciclo vital dos documentos, algumas funções administrativas que apoiam essa gestão, tais como “direção” e “capacitação”:

O planejamento, o controle, a direção, a organização, a capacitação, a promoção e outras atividades gerenciais relacionadas com a criação de documentos, sua manutenção, uso e eliminação, incluindo o manejo de correspondência, formulários, diretrizes, informes, documentos informáticos, microformas, recuperação de informação, fichários, correios, documentos vitais, equipamentos e materiais, máquinas reprográficas, técnicas de automação e elaboração de dados, preservação e centros de arquivamento intermediários ou outras instalações para armazenagem⁵. (JARDIM, 1987, p. 35).

No Brasil, a Lei nº 8.159, de 8 de janeiro de 1991, em seu art. 3º, define gestão de documentos como “o conjunto de procedimentos e operações técnicas referentes à sua produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, visando a sua eliminação ou recolhimento para guarda permanente” (BRASIL, 1991). Neste trabalho, compreende-se que essa definição tem uma abordagem estritamente técnica para o tratamento do conjunto de documentos arquivísticos correntes e intermediários.

5 No artigo não foi dada a referência da lei norte-americana que o autor utilizou.

Pensar a gestão de documentos para além do tratamento técnico é uma forma de ampliar sua implantação em um ambiente corporativo. Essa visão macro da gestão de documentos vem ao encontro da ideia defendida por Heredia Herrera, em que:

[...] A gestão de documentos não é sinônimo de tratamento documental, ainda que esse esteja incluído nela. Se o tratamento não tinha mais o controle do arquivista e sua execução não ultrapassava as paredes do Arquivo, então a gestão de documentos pula o muro e muitas das funções arquivísticas se preveem ou se ultrapassam⁶. (HEREDIA HERRERA, 2007, p. 93).

O conceito de gestão de documentos teve sua gênese a partir de uma visão administrativa e não arquivística. Essa visão ainda é mantida, principalmente, nas áreas de Administração e Informática quando da decisão pelo desenvolvimento de sistemas informatizados para gestão de documentos, de informações ou de dados em ambiente digital. O que se considera nessas áreas é a economia de espaço físico (iniciativas “papel zero”) e a agilidade na busca e acesso à informação. Ressalte-se, ainda, que, desses sistemas informatizados, poucos contemplam requisitos mínimos aplicáveis aos documentos arquivísticos.

Para que documentos arquivísticos, destacadamente os digitais, mantenham-se confiáveis, autênticos e acessíveis pelo tempo em que forem necessários, é essencial que esses sistemas informatizados contemplem os seguintes requisitos arquivísticos apresentados no e-ARQ Brasil: Modelo de Requisitos para Sistemas de Gestão Arquivística de Documentos⁷:

- captura, armazenamento, indexação e recuperação de todos os tipos de documentos arquivísticos;
- captura, armazenamento, indexação e recuperação de todos os componentes digitais do documento arquivístico como uma unidade complexa;
- gestão dos documentos a partir do plano de classificação para manter a relação orgânica entre os documentos;
- implementação de metadados associados aos documentos para descrever os contextos desses mesmos documentos (jurídico-administrativo, de proveniência, de procedimentos, documental e tecnológico);
- integração entre documentos digitais e convencionais;
- foco na manutenção da autenticidade dos documentos;
- avaliação e seleção dos documentos para recolhimento e preservação daqueles considerados de valor permanente;
- aplicação de tabela de temporalidade e destinação de documentos;
- transferência e recolhimento dos documentos por meio de uma função de exportação; e
- gestão e preservação dos documentos. (CONSELHO NACIONAL DE

6 [...] La gestión documental no es sinónimo de tratamiento documental, aunque este esté incluido en ella. Si aquel non tenía más control que el del archivero y su realización no sobrepasaba las paredes del Archivo, la gestión documental salta la tapia y muchas de las funciones archivísticas se prevén o se adelantan.

7 Disponível em: <<http://www.conarq.arquivonacional.gov.br>>.

ARQUIVOS, 2011, p. 11).

Neste momento, cabe destacar que o e-ARQ Brasil

[...] foi elaborado pela CTDE tendo como base os modelos e padrões internacionais. Foram tomadas como ponto de partida para esse trabalho a norma ISO 15.489 de Gestão de Documentos e duas iniciativas internacionais de especificação de requisitos funcionais: Modelo de Requisitos Funcionais (MoReq), da União Europeia e a especificação de requisitos do Departamento de Defesa norte-americano (DoD), que foi adotada pelo Arquivo Nacional dos Estados Unidos para ser aplicada na administração federal. Esses documentos foram confrontados com as normas e diretrizes no âmbito nacional relacionadas à gestão, manutenção, acesso e preservação de documentos arquivísticos, a fim de adaptar a definição dos requisitos à legislação e práticas arquivísticas brasileiras. (ROCHA, 2008, p. 114).

Como se pode observar, Rocha (2008) esclarece que a elaboração do modelo de requisitos para sistemas informatizados de gestão arquivística de documentos (SIGADs), para adoção no âmbito brasileiro, acompanha um movimento internacional cuja compreensão é de que, para gestão e preservação de documentos arquivísticos digitais confiáveis, autênticos e acessíveis, é necessária a implementação de um SIGAD, conjuntamente a uma política e programas arquivísticos.

Diante disso, faz-se necessária uma revisitação do termo e do conceito de gestão de documentos para ampliar sua acepção. Com relação ao termo – do inglês *records management* –, encontram-se na literatura brasileira outras traduções, como administração de documentos, gestão arquivística de documentos e gestão de documentos arquivísticos, sendo esta última mais próxima do termo em inglês.

Neste trabalho, entretanto, supõe-se apropriado qualificar a gestão como *arquivística*, o que remete à adoção dos princípios, métodos e processos preconizados pela Arquivologia para gestão e preservação dos documentos produzidos e recebidos por uma entidade coletiva, pessoa ou família no desenvolvimento de suas atividades, sejam eles digitais e não-digitais.

Uma outra perspectiva quanto ao conceito de gestão de documentos consiste em incluir a terceira idade⁸ do arquivo, o arquivo permanente, pois

Alguns autores e mesmo a lei brasileira de arquivos considera a gestão como sendo os procedimentos, as medidas e/ou rotinas destinadas ao controle e uso dos documentos, incluindo a sua produção, tramitação, arquivamento, uso até a avaliação ou destinação final. Estes conceitos implicam afirmar que a gestão

8 Os arquivos têm três idades: corrente, intermediária e permanente. Na idade corrente, os arquivos são consultados frequentemente devido ao seu valor administrativo, fiscal ou jurídico em vigor. Na idade intermediária mantêm esses valores, porém é pouco consultado, e são guardados por razões legais. Já na idade permanente, é atribuído um valor histórico, informativo, servindo também a outros interesses que não aqueles para os quais foram criados.

acontece até o arquivo intermediário, não estando compreendido o gerenciamento dos arquivos permanentes [...]. (SILVA; MARTINS, 2007, p.76).

Na Arquivologia nacional, já estão consagrados o termo e o conceito de gestão de documentos, mas há estudiosos optando por gestão de documentos arquivísticos e gestão arquivística de documentos. Isso pode ser considerado um indicativo de que gestão e preservação de documentos arquivísticos, especialmente em ambiente digital, precisam ser contempladas desde a primeira idade dos arquivos.

Quando Silva (2009) discorre sobre o “fenômeno do *records management*”, aponta que

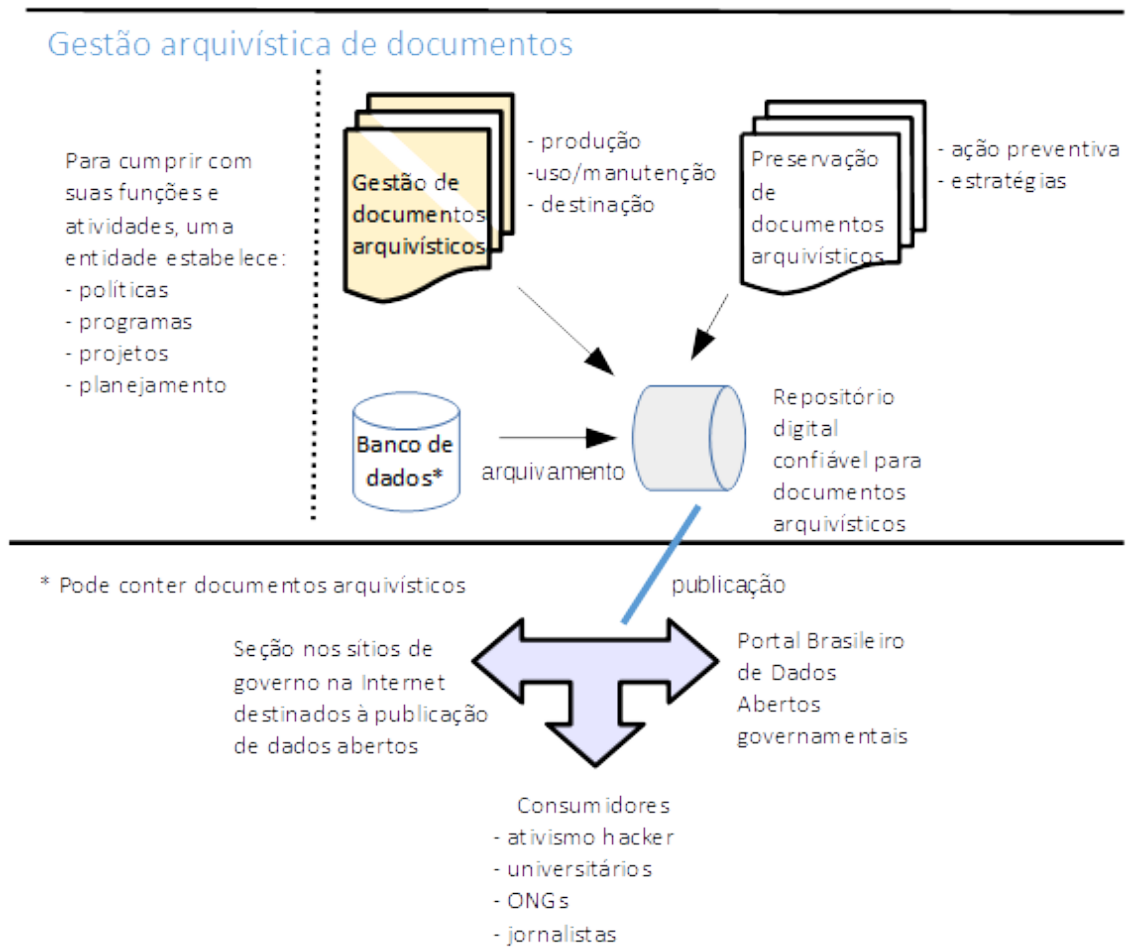
[...] Embora seja inevitável, em muitos casos, a separação material dos documentos e, em cada fase da vida dos arquivos haja operações específicas a ter em conta, isto não contraria a necessidade de uma metodologia integradora, sem a qual não serão inteligíveis os mecanismos internos que explicam a essência dos arquivos, enquanto sistemas orgânicos. (SILVA, 2009, p. 136, grifo nosso).

Aqui, a noção de *gestão arquivística de documentos* suscita indagações quanto à sua pertinência. Porém, se afirmada, seria possível a gestão arquivística da informação ou de dados, especialmente em um ambiente digital?

Devido ao fato de esses objetos de estudo serem abordados por diferentes áreas do conhecimento e suas respectivas teorias e práticas, torna-se inviável fazer tal afirmação, mas, especialmente em ambiente digital, a gestão arquivística de documentos apoia o acesso e a disponibilização de informações e dados com qualidades, tais como credibilidade, acurácia e autenticidade, já que eles estão contidos em documentos arquivísticos que são arquivados em repositórios digitais confiáveis (Figura 1).

Esse cenário possibilita melhorias ao acesso à informação pública por meio de transparência ativa (quando os órgãos publicam suas informações em sites) ou de transparência passiva (quando o cidadão faz um pedido de acesso ao órgão).

Figura 1 – A gestão arquivística de documentos apoiando a disponibilização de dados e informações.



Por fim, a gestão (arquivística) de documentos, quando aplicada a um ambiente digital, mantém-se importante para otimização dos recursos materiais e financeiros, assim como contribui para a qualidade de trabalho dos colaboradores envolvidos nas atividades e funções de um órgão público ou entidade privada.

2.1 DADO, INFORMAÇÃO E DOCUMENTO: UMA CONEXÃO POSSÍVEL

De acordo com Duranti (2001), a adoção de tecnologias por indivíduos e instituições para o desenvolvimento de suas atividades tem gerado mais informações agora do que em décadas anteriores. Com isso, também cresce a preocupação com os documentos que são produzidos e arquivados em ambiente digital. A autora acrescenta ainda que:

O maior desafio que os sistemas digitais nos apresentam é a produção e manutenção de dados confiáveis e a preservação de sua autenticidade ao longo do tempo. Sobretudo, o problema é mais sério não com relação aos dados ou informações ou documentos eletrônicos, mas sim com os documentos arquivísticos eletrônicos⁹. (DURANTI, 2001, p. 272, tradução nossa).

Com relação ao fato de o documento arquivístico digital também ser fonte de dados é pertinente apresentar a diferença conceitual entre dado, informação e documento, a partir dos estudos de pesquisadores das áreas de Arquivologia, Engenharia Eletrônica, Ciência da Informação e da Computação.

Na literatura brasileira, o estudo realizado por Rondinelli (2013, p. 19-101) apresenta os conceitos de documento, informação e dado no âmbito da Ciência da Informação e da Arquivologia. Nele, elucida-se uma “convergência conceitual” entre informação e documento. Tal convergência centra-se na materialidade e na funcionalidade detidas por ambos.

Duranti apresenta também “o entendimento do termo 'dado' no âmbito da arquivologia” (p. 97) que, “[...] à exceção do InterPares 3 Project, o termo é definido da mesma maneira pela maioria dos *autores da ciência da informação* [...], ou seja, como *matéria-prima da informação*” (p. 98, grifos nossos).

Por razões estruturais na elaboração deste trabalho, serão abordados separadamente os conceitos dos três objetos – dado, informação e documento – porém, parte-se da compreensão da existência de uma conexão entre eles, isto é, uma ligação conceitual entre os diferentes objetos cujo resultado seria o aprimoramento das partes envolvidas.

a) Conceito de dado

Os estudos sobre os dados produzidos em áreas corporativas, científicas, estatísticas, censitárias, entre outras, têm avançado progressivamente, devido ao entendimento por parte dos gestores de que tais dados são relevantes para o sucesso dos negócios empreendidos. Para que o uso seja profícuo, os produtores de dados precisam estar atentos a eles como ativos da entidade. Mas o que é um dado?

A definição de dado apresenta nuances conforme a área do conhecimento em que

9 The greatest challenge with which digital systems present us is the creation and maintenance of reliable data and the preservation of their authenticity over time. Moreover, this problem is most serious not with regard to electronic data or information or documents, but with regard to electronic records.

está sendo abordado, porém, em sentido amplo, é compreendido como um valor bruto ou a menor parte da informação, cujo significado não é abrangente para o leitor, por isso pode ser considerado como fonte de informação para os indivíduos que dele façam uso.

A ideia acima é reforçada quando Duranti (2005) afirma que

Um dado é a unidade elementar de determinado conteúdo, isto é, sua menor parte com significado, uma indivisível unidade de informação [...] do ponto de vista arquivístico, dados são as menores partes do conteúdo de um documento. Eles são geridos, comunicados e mantidos no contexto de outros dados, então tornando-se informação, a qual, manifestada numa certa forma e fixada num suporte, constitui um documento¹⁰. (DURANTI, 2005, p. 109, tradução nossa).

Rousseau e Couture (1998) aproximam-se de Duranti ao dizerem que

o “dado” pode ser definido como a mais pequena representação convencional e fundamental de uma informação (facto, noção, objecto, nome próprio, número, estatística, etc.) sob a *forma analógica ou digital*, que permita efetuar o seu tratamento manual ou automático (informático). (ROUSSEAU; COUTURE, 1998, p. 137, grifo nosso).

Em outra definição, dado é a “sequência de símbolos ou valores, representados em algum meio, produzidos como resultado de um processo natural ou artificial”, adotada pela Secretaria de Logística e Tecnologia da Informação (SLTI) em sua Instrução Normativa nº 4 (BRASIL, 2012).

O autor Valdemar W. Setzer define “dado como uma sequência de símbolos quantificados ou quantificáveis”. Tal definição exige mais explicações do próprio autor:

[...] um texto é um dado. De fato, as letras são símbolos quantificados, já que o alfabeto por si só constitui uma base numérica. Também são dados imagens, sons e animação, pois todos podem ser quantificados a ponto de alguém que entra em contato com eles ter eventualmente dificuldade de distinguir a sua reprodução, a partir da representação quantificada, com o original. É muito importante notar-se que qualquer texto constitui um dado ou uma sequência de dados, mesmo que ele seja ininteligível para o leitor. [...] Como são símbolos quantificáveis, dados podem obviamente ser armazenados em um computador e processados por ele.

Em nossa definição, um dado é necessariamente uma entidade matemática e, desta forma, puramente sintática. Isto significa que os dados podem ser totalmente descritos através de representações formais, estruturais. Dentro de um computador, trechos de um texto podem ser ligados virtualmente a outros trechos, por meio de contiguidade física ou por "ponteiros", isto é, endereços da unidade de armazenamento sendo utilizada. Ponteiros podem fazer a ligação de um ponto de um texto a uma representação quantificada de uma figura, de um som, etc. (SETZER, 1999, grifo nosso).

10 A datum is the elementary unit of such content, that is, its smallest meaningful part, an indivisible unit of information [...] from an archival point of view, data are the smallest part of a record's content. They are managed, communicated and kept in the context of other data, thereby becoming information, which, manifested in a given form and affixed to a medium constitutes a document.

Aldo Barreto, cientista da informação, apresenta também uma outra contribuição acerca da definição de dado. Para ele, “dado é um conjunto de símbolos quantificados e/ou qualificados”¹¹ (apud ZINS, 2007, p. 480, tradução nossa). Cabe observar que essa definição assemelha-se àquela feita por Setzer, em 1999.

Para Meadow e Yuan (1997 apud RONDINELLI, 2013), dado é um “conjunto de símbolos com pouco ou nenhum significado para o receptor”. Nessa acepção, aparece a figura do receptor do dado, o que não ficou explícito nos demais conceitos apresentados.

William Alves (2009, p. 22-23) diz que há uma diferença entre dado e informação. Para o autor, “[...] Informação é qualquer fato ou conhecimento do mundo real e que pode ou não ser registrado/armazenado. Dado é a representação da informação, que pode estar registrado em papel, num quadro de aviso ou no disco rígido do computador. [...] e] devemos lembrar que um computador armazena e processa dados, não informações.”

A título de ilustração, foram selecionadas aqui algumas definições de *dado* apresentadas por arquivistas (Duranti, Rousseau e Couture), engenheiro eletrônico (Setzer), cientistas da informação (Aldo Barreto, Meadow e Yuan) e informáticos (Alves, equipe da SLTI). Trata-se, portanto, de autores de diferentes áreas do conhecimento abordando um mesmo objeto: o dado (Quadro 1).

Quadro 1 – Conceitos de dado e suas semelhanças

Autor	Definição	Semelhanças
Duranti (arquivista)	Dado é a unidade elementar de determinado conteúdo, isto é, sua menor parte com significado, uma indivisível unidade de informação.	<ul style="list-style-type: none"> - Conjunto de símbolos - Quantificados - Representação da informação - Com significado
Rousseau e Couture (arquivistas)	Representação convencional e fundamental de uma informação (facto, noção, objecto, nome próprio, número, estatística, etc.).	
Setzer (engenheiro eletrônico)	Sequência de <i>símbolos quantificados</i> ou quantificáveis; entidade matemática.	
Aldo Barreto (cientista da informação)	Conjunto de <i>símbolos quantificados</i> e/ou qualificados.	
Meadow e Yuan (cientista da informação)	Conjunto de <i>símbolos</i> com pouco ou nenhum significado para o receptor	
Secretaria de Logística e Tecnologia da Informação (cientista da computação/informática)	Sequência de <i>símbolos</i> ou valores, representados em algum meio, produzidos como resultado de um processo natural ou artificial	

¹¹ Is a symbol set that is quantified and/or qualified.

Alves (cientista da computação/informática)	Dado é a representação da informação (fato ou conhecimento do mundo real e que pode ou não ser registrado/armazenado), que pode estar registrado em papel, num quadro de aviso ou no disco rígido do computador.	
---	--	--

Fonte: elaboração própria com base em DURANTI (2005), ROUSSEAU E COUTURE (1998), SETZER (1999), BARRETO (apud ZINS, 2007), MEADOW E YUAN (1997 apud RONDINELLI, 2013), SLTI (2012), ALVES (2009).

Para fins teórico-metodológicos na explanação deste trabalho, compreende-se aqui que dados também estão contidos em um documento arquivístico, em conformidade com a linha de pensamento dos autores Duranti (2005) e Rousseau e Couture (1998).

b) Conceito de informação

O conceito de informação foi (e ainda é) discutido por autores de diferentes áreas do conhecimento, principalmente da Ciência da Informação, por isso, a abordagem adotada nesta pesquisa não será de cunho extensivo, já que a informação, como se sabe, não é o objeto de estudo da Arquivologia. O que se pretende é levantar subsídios para uma compreensão mais detalhada do que seja informação, a fim de diferenciá-la do termo dado.

Em relevante contribuição, Rondinelli (2013) desenvolve o estudo filológico e entitativo do termo *informação*, baseando-se em autores como Rafael Capurro, Félix Fernández Sagredo, José Maria Arroyo Izquierdo, Lena Pinheiro, dentre outros cientistas da informação, assim como Blanca Rodríguez Bravo, Ana Maria Camargo, Trevor Livelton, entre outros arquivistas.

De acordo com suas interpretações comparativas sobre os estudos do conceito de informação na Ciência da Informação e na Arquivologia, Rondinelli afirma que

[...] dados, ideias ou elemento referencial devidamente registrados, ou organizados, ou estruturados, ou ainda modelados, têm por função comunicar conhecimento, testemunhar ações, o que evoca imediatamente a abordagem etimológica e semântica da entidade *informação*. (RONDINELLI, 2013, p. 99, grifo do autor).

Essa funcionalidade da informação de “comunicar conhecimento” e “testemunhar ações” foi otimizada através do uso das tecnologias da informação e comunicação, o que nos conduz à reflexão sobre se isso contribuiu para a generalização do termo informação, sendo adotado também como sinônimo de documento e dado, ou ainda

[...] para denotar coisas distintas que pouco têm a ver entre si, ou seja, a linguagem corrente o dotou de um amplo conteúdo semântico [...] abarca desde fatos compatíveis com o conhecimento ou o significado da mensagem, até aspectos relativos à importância ou à verdade do mesmo. A consequência dessa situação é a ambiguidade do termo, bem como a pobreza e a confusão conceitual. (RODRÍGUEZ BRAVO, 2002, p. 43 apud RONDINELLI, 2013, p. 96).

A diversidade acerca do conceito de informação foi mapeada por Chaim Zins (2007), assim como de dado e conhecimento. Dentre as acepções de informação, Barreto (apud ZINS, 2007, p. 480, tradução nossa) diz que “é um conjunto de símbolos com significados que tem a habilidade de criar conhecimento”¹². Meadow e Yuan (1997 apud RONDINELLI, 2013) apresentam um conceito semelhante, afirmando que informação é um “conjunto de símbolos com significado para o receptor”.

Para Albrechtsen (apud ZINS, 2007, p. 480, tradução nossa), a informação está relacionada à intenção humana e que “em sistemas computacionais informação é o conteúdo de base de dados, da *web* etc. [...]”¹³. Tal conteúdo é que tem sido publicado nos sítios de órgãos governamentais para cumprimento das ações de governo aberto, tanto em âmbito nacional quanto internacional.

Assim, com base nos autores citados, tende-se a concordar com a ideia de que a informação é o conjunto de dados que tem algum significado para o seu receptor, diferentemente do dado bruto quando representa parte do todo de um texto completo.

c) Conceito de documento

O registro de uma informação num suporte constitui um documento. Trata-se, portanto, de uma visão reducionista a respeito desse objeto de estudo para diferentes áreas do conhecimento, como a Documentação, a Diplomática, a Ciência da Informação e a Ciência da Computação, aqui destacadas, já que seus arcabouços teóricos se aproximam do objeto de estudo da Arquivologia – o documento arquivístico.

Neste ponto, faz-se necessário um retorno ao trabalho de Rondinelli (2013), no qual a autora também faz uma abordagem filológica (sob as vertentes semântica e etimológica) e entitativa do termo *documento* quando estudado pela Ciência da Informação e pela

12 [...] is a set of significant signs that has the ability to create knowledge.

13 [...] In computational systems information is the contents of databases, the web, etc [...].

Arquivologia. A autora, ao fazer suas “interpretações comparativas” (RONDINELLI, 2013, p. 52), identifica a materialidade e a funcionalidade do documento como pontos de convergência nas duas áreas, sendo que

[...] Por meio da funcionalidade o documento detém as funções de ensino e aprendizagem, registro e comunicação da informação, testemunho de fatos e de ações, numa noção mais coerente com a etimologia e a semântica do termo documento. (RONDINELLI, 2013, p. 53)

De acordo com Otlet (1907 apud RAYWARD, 1997, p. 299, tradução nossa), os “[...] Documentos consistem no que quer que seja que represente ou expresse um objeto, um fato, uma impressão por meio de qualquer sinal (escrita, pintura, diagrama, símbolos)¹⁴” - os códigos binários existentes num ambiente digital enquadram-se na categoria de símbolos.

A criação de documentos em ambiente digital foi uma inovação para os arquivistas no sentido de que os registros sobre um fato ou um objeto de estudo somente seriam acessíveis de forma indireta – por meio de *softwares* e *hardwares* (RONDINELLI, 2013, p. 25). Oportunidade esta para uma aproximação com as áreas de exatas, como a Ciência da Computação ou Sistemas de Informação, por exemplo.

A funcionalidade de um documento melhor o caracteriza quando estudado em ambientes digitais, isso porque

Diferentes tecnologias possuem diferentes capacidades, diferentes restrições. Se nós sustentamos a visão funcional de um documento, podemos esperar que os documentos tenham diferentes formas no contexto de diferentes tecnologias e, então, podemos esperar que a escala do que pode ser considerado um documento diferencie em ambientes digitais e em papel. O algoritmo para gerar um logaritmo, como um brinquedo educativo mecânico, pode ser visto como um tipo de documento dinâmico oposto a um documento comum em papel, mas ainda de acordo com a origem etimológica de “docu-ment”, no sentido de ensino – ou, na verdade, evidência, algo a partir do qual alguém aprende alguma coisa¹⁵. (BUCKLAND, [1998], tradução nossa).

Dentre os conceitos apresentados por Rondinelli, a partir de diferentes fontes, destaca-se aqui, como o mais representativo, o de Rodríguez Bravo (2002 apud

14 Documents consist of whatever represents or expresses and (sic) object, a fact, an impression by means any sign whatever (writing, picture, diagrams, symbols).

15 Each different technology has different capabilities, different constraints. If we sustain the functional view of what constitutes a document, we should expect documents to take different forms in the contexts of different technologies and so we should expect the range of what could be considered a document to be different in a digital and paper environments. The algorithm for generating logarithms, like a mechanical educational toy, can be seen as a dynamic kind of document unlike ordinary paper documents, but still consistent with the etymological origins of "docu-ment", a means of teaching - or, in effect, evidence, something from which one learns.

RONDINELLI, 2013, p. 54), em que documento é o “suporte portador de uma mensagem emitida com intenção comunicativa e potencialmente informativa para o receptor”.

Para fins teóricos, compreende-se aqui, neste trabalho, que informação registrada em suporte é documento, o que “não significa o atrelamento da entidade documental a determinado tipo de suporte, mas à necessidade de estabilidade e permanência” de seu conteúdo (RONDINELLI, 2013, p. 53). Na produção de documentos através do uso de tecnologias da informação e comunicação (TICs), é válido ter esse entendimento como referência.

De acordo com Zins (2007), “muitos acadêmicos dizem que dado, informação e conhecimento são partes de uma ordem sequencial. Dados são o material bruto para a informação e informação é o material bruto para o conhecimento [...]”¹⁶ (ZINS, 2007, p. 479, tradução nossa). O autor não inclui, porém, o documento nessa sua hierarquia; no entanto, vale ressaltar que tais dados e informações são registrados em algum suporte.

d) Definição de *dataset*

Revisitar os conceitos de dado, informação e documento – a partir da ótica arquivística – se faz relevante para a análise do que consiste um conjunto de dados (*dataset*) publicado segundo a iniciativa de dados abertos adotada em âmbitos governamentais: se é um dado, uma informação, um documento ou um documento arquivístico. Contudo, é pertinente dizer que o termo *dataset* (ou *data set*) apresenta diferentes conotações, dependendo do grupo de usuários que trabalha com esse objeto, daí a dificuldade de se encontrar uma única acepção (PEPLER, 2008 apud RENEAR; SACCHI; WICKETT, 2010).

O engenheiro de *software* Leigh Dodds, em 2013, fez um levantamento das diferentes definições de *dataset* adotadas de acordo com o propósito para o qual determinado conjunto de dados foi estabelecido. Segundo esse levantamento, ele pode ser de uso geral, de uso pela lei de acesso à informação, de uso estatístico, de uso científico e em pesquisas, de uso em *Resource Description Framework* (RDF) e SPARQL¹⁷. Para o autor, “a razão pela qual

16 Many scholars claim that data, information, and knowledge are part of a sequential order. Data are the raw material for information, and information is the raw material for knowledge.

17 RDF e SPARQL são, respectivamente, um modelo padrão para intercâmbio de dados na Web e um conjunto de especificações para linguagens e protocolos de *query* e manipulação de conteúdos grafos em RDF na Web ou em RDF *store*. Disponível em: <www.w3.org>.

um *dataset* é formado e as definições de seu conteúdo são importantes para apoiar o seu correto reuso” (DODDS, 2013, tradução nossa)¹⁸. No âmbito da iniciativa de dados abertos, entende-se, aqui, a relevância do apontamento feito por Dodds.

O termo *dataset* é usual nas áreas de Computação e Informática. No glossário da *International Business Machine* (IBM), é definido como “uma coleção de dados, normalmente na forma de linhas (registros) e colunas (campos) e contidos num arquivo ou tabela de banco de dados”¹⁹ (INTERNATIONAL BUSINESS MACHINE, 2015, tradução nossa).

No glossário da *Society of American Archivists*, *dataset* é definido como “uma coleção de informação relacionada, especialmente informação formatada para análise por um computador”. Na nota do mesmo verbete consta que “no termo fica subentendido que a informação tenha sido estruturada para facilitar a análise”²⁰ (PEARSE-MOSES, 2005, tradução nossa).

Para as autoras Renear, Sacchi e Wickett (2010), da Escola de Biblioteconomia e Ciência da Informação, Illinois, Estados Unidos, *dataset* é uma noção²¹:

5. Está claro a partir do exposto aqui que a noção de “dataset” encontrada na literatura não pode por si mesma conter uma definição formal precisa, ou seja, esta noção geral deve ser substituída por uma família inter-relacionada de conceitos específicos²². (RENEAR; SACCHI; WICKETT, 2010, p. 3, tradução nossa, grifo do autor).

Após a exposição de algumas definições para o termo *dataset*, questiona-se se, nas seções dos sítios de órgãos públicos na Internet destinadas aos dados abertos são publicados:

1) dados brutos (símbolos com pouco significado); 2) informações (símbolos com significado); 3) documentos (representação de fatos, num suporte, por meio de sinais); ou 4) documentos arquivísticos (instrumento ou subproduto de uma atividade prática).

18 The means by which a dataset is collected and the definitions of its contents are important for supporting proper re-use.

19 A collection of data, usually in the form of rows (records) and columns (fields) and contained in a file or database table.

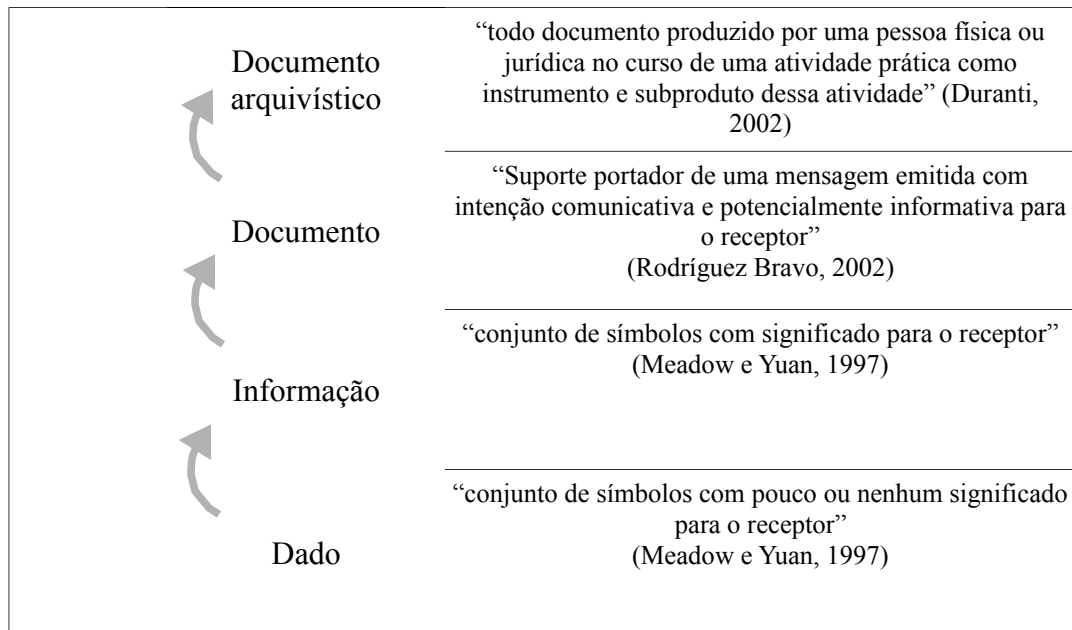
20 (Noun) A collection of related information, especially information formatted for analysis by a computer. [...] The term implies that the information has been structured to facilitate analysis.

21 “[...] por noção entendemos aqueles elementos de uma teoria que ainda não apresentam clareza suficiente e são usados como 'imagens' na explicação do real”. MINAYO, M. C. S. Fase exploratória da pesquisa. In: _____. **O Desafio do Conhecimento: Pesquisa Qualitativa em Saúde**. São Paulo – Rio de Janeiro: Hucitec/ Abrasco, 1992, p. 89-104.

22 It is clear from the forgoing that the notion of “dataset” found in the literature cannot itself be provided with a precise formal definition, but that this general notion must be replaced by an interrelated family of more specific concepts.

Ao se compreender a possibilidade de uma conexão entre dado, informação, documento e documento arquivístico, pois um dado está contido numa informação e essa num documento que, por sua vez, sendo o registro de uma atividade é reconhecido como um documento arquivístico, observa-se o limite de cada objeto de estudo (Figura 2).

Figura 2 – Conexão entre dado, informação, documento e documento arquivístico.



Fonte: elaboração própria, com base em RONDINELLI, 2013, p. 100-101.

A conexão entre dado, informação, documento e documento arquivístico pode ser estabelecida e melhor compreendida teoricamente. Porém, na busca por uma resposta ao questionamento feito sobre o que é um *dataset* publicado numa seção específica dos sítios de órgãos públicos na Internet, elaborou-se uma representação de dado, informação e documento a partir do tipo documental, “demonstrativo de pagamento de pessoal”, que foi analisado pela equipe da Universidade Estadual de Campinas (Unicamp) integrante do Projeto InterPARES 3 (International Research on Permanent Authentic Records in Electronic Systems) (Figura 3).

O Projeto InterPARES²³, com sua base na universidade canadense de British Columbia (UBC), tem como objetivo desenvolver pesquisas acerca da preservação em longo prazo dos documentos arquivísticos digitais. Sua terceira fase concluiu-se em 2012, quando foi apresentado o relatório final do Estudo de Caso BR07.

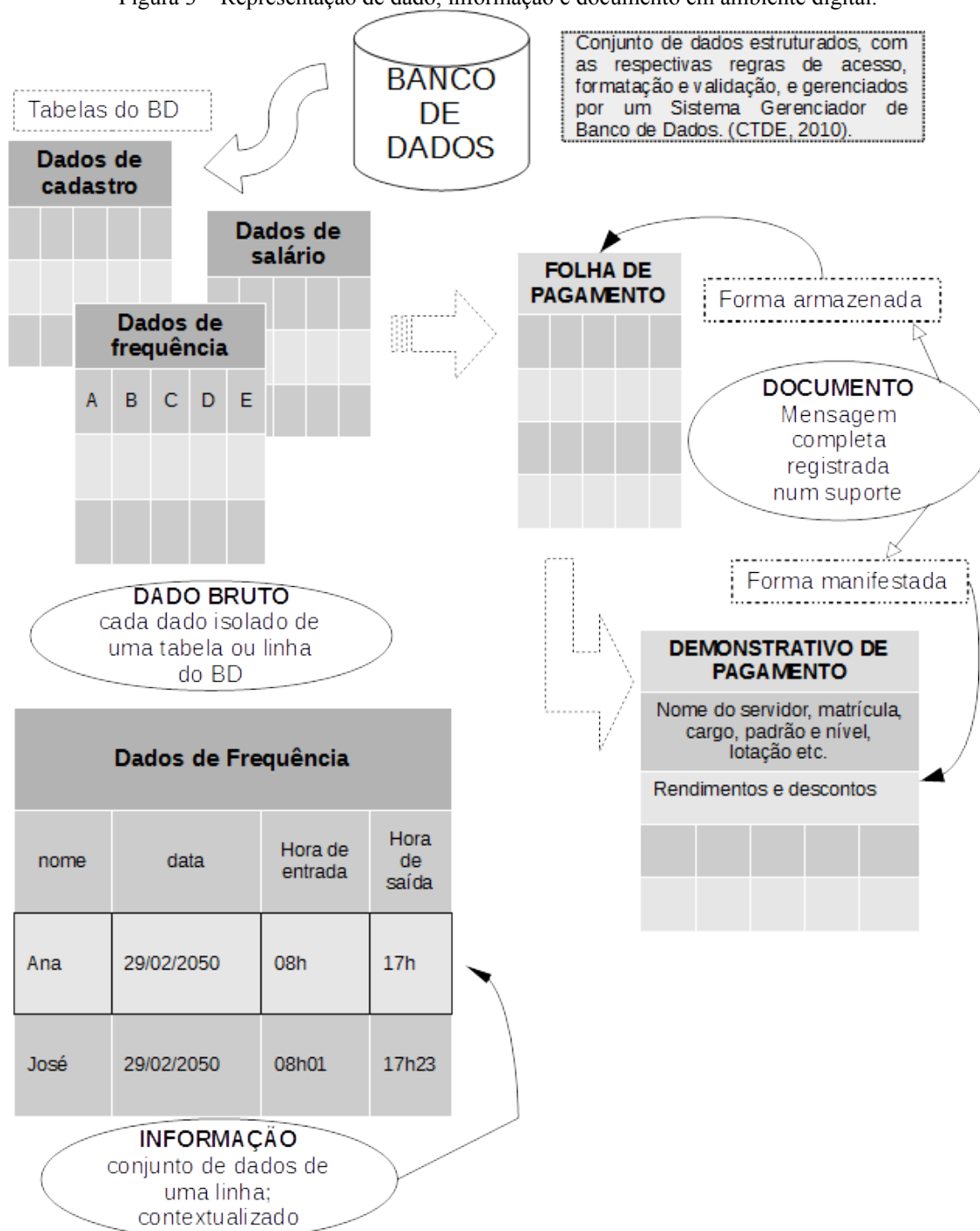
23 www.interpares.org

De acordo com o Estudo de Caso BR07, “[...] Os dados correspondentes ao pagamento do mês [aos funcionários da Unicamp] são mantidos em tabelas do banco de dados (Oracle)” (PROJETO INTERPARES 3, p. 9). Tais tabelas são constituídas por campos que, segundo Alves (2009),

o campo de um registro é a menor unidade destinada ao armazenamento de valores existente em um arquivo ou tabela de um banco de dados e pode ser referenciado por um programa aplicativo. Isso significa que os dados armazenados são separados em pequenos fragmentos. Cada campo somente pode conter um tipo de dado. (ALVES, 2009, p. 37).

A partir desse entendimento, cada campo das tabelas contém um dado bruto (com pouco significado).

Figura 3 – Representação de dado, informação e documento em ambiente digital.



Fonte: elaboração própria, com base no Estudo de Caso BR07 (PROJETO INTERPARES 3, 2012). Colaboração de Claudia Lacombe Rocha, mestre em Sistemas de Computação, do Arquivo Nacional.

A identificação de um *dataset* como um dado bruto, informação, documento ou documento arquivístico auxilia no seu gerenciamento, a fim de atingir o objetivo de publicização de dados governamentais apropriados para reuso dos cidadãos.

Ao se retomar a definição ampla de que *dataset* é um conjunto de dados estruturados/semiestruturados, em geral na forma de linhas e colunas, contidos num arquivo digital ou tabela de banco de dados, e aplicar na representação da figura 3, didaticamente, obtém-se que:

- a) cada campo de uma tabela contém um *dado bruto*, isolado;
- b) uma linha de uma tabela ou a tabela completa é uma *informação*, com significado para o receptor;
- c) o conjunto de informações presentes nas tabelas, registrado num suporte e com uma mensagem completa, é um *documento*. Esse documento, no ambiente digital, pode estar disperso em várias linhas de diferentes tabelas (os componentes digitais de um documento na sua forma manifestada, como é o caso do demonstrativo de pagamento do Estudo de Caso BR 07). Ele também pode ser entendido como um conjunto de tabelas (um documento na sua forma armazenada, como é o caso da folha de pagamento mês a mês); e
- d) o conjunto de tabelas que registra e apoia uma atividade-meio ou fim desenvolvida por um órgão é um *documento arquivístico*.

A exposição acima buscou elucidar, com um exemplo prático, as conexões entre dado, informação e documento. Esse exercício faz-se relevante para que arquivistas e informáticos encontrem os recursos teóricos e práticos mais adequados para apoiar a publicação de dados governamentais abertos nas seções dos sítios de órgãos públicos na Web.

2.2 O DOCUMENTO ARQUIVÍSTICO DIGITAL COMO FONTE DE DADOS DIGITAIS

Duranti define o termo *records* (documento arquivístico) como “todo documento produzido por uma pessoa física ou jurídica no curso de uma *atividade prática* como *instrumento e subproduto* dessa atividade”. (DURANTI, 2002 apud RONDINELLI, 2013, p. 185, grifos nossos).

Duranti e Rondinelli, a partir de seus estudos, apresentam as explicações sobre os termos “atividade prática” e “como instrumento e subproduto” na definição de documento arquivístico, a saber:

- a) atividade prática: “conjunto de atos cujo objetivo é o cumprimento da atividade em si, onde atos são manifestações da vontade com um fim específico” (DURANTI, 2002

apud RONDINELLI, 2013, p. 186);

b) instrumento: “se refere a todos os documentos arquivísticos que viabilizam uma ação ou sem o qual a ação não se efetiva. Trata-se do que a diplomática chama de documentos dispositivos” (RONDINELLI, 2013, p. 186); e

c) subproduto: “diz respeito a todos os demais documentos arquivísticos que provam [...], apoiam [...], facilitam [...], orientam [...] ou possibilitam uma ação, mas não a condicionam” (RONDINELLI, 2013, p. 186).

O documento arquivístico digital (*digital record*) tem sido objeto de estudo de arquivistas nacionais e internacionais devido às especificidades que tal documento exige em sua produção, uso, manutenção, destinação e preservação. Por isso,

um documento digital produzido no decorrer de atividades desempenhadas por pessoas físicas ou jurídicas, e cuja análise diplomática demonstra que, além das características de forma fixa e conteúdo estável, as demais – relação orgânica, contexto identificável, ação, e cinco pessoas (autor, redator, destinatário, originador e produtor), ou ao menos as três primeiras – também se fazem presentes, é documento arquivístico, e como tal deverá ser submetido aos procedimentos de gestão e preservação preconizados pela arquivologia. (RONDINELLI, 2013, p. 267).

Duranti e Rondinelli, dentre outros autores, analisam o documento arquivístico digital com base nos preceitos da Diplomática; para as autoras, tal documento deve ter forma fixa e conteúdo estável. Segundo Rondinelli,

[...] Em um primeiro momento, tais características [de forma fixa e conteúdo estável] podem parecer incompatíveis com o dinamismo do ambiente digital. Entretanto, um estudo desse ambiente e dos objetos ali existentes, à luz da ciência da computação, da diplomática e da arquivologia, resultou numa taxonomia dos documentos digitais pela qual estes são classificados em estáticos, interativos não dinâmicos e interativos dinâmicos. (RONDINELLI, 2013, p. 259).

As características e partes constituintes do documento arquivístico digital estão representadas no quadro 2.

Quadro 2 – Características e partes constituintes do documento arquivístico digital.

<p>Características: forma fixa, conteúdo estável, relação orgânica, contexto identificável, ação e cinco pessoas (autor, redator, destinatário, originador e produtor).</p> <p>Partes constituintes: forma documental, anotações, contexto, suporte, atributos e componentes digitais.</p> <p>FORMA DOCUMENTAL: elementos intrínsecos e extrínsecos.</p> <ul style="list-style-type: none"> • elementos intrínsecos: cinco pessoas (autor, redator, destinatário, originador e produtor), data cronológica, data tópica, indicação e descrição da ação ou assunto e atestação; • elementos extrínsecos: <ul style="list-style-type: none"> apresentação geral: texto, imagem, som, gráfico; apresentação específica: layout, cor, hiperlink, indicação de anexo; outros: assinatura digital, marcas d'água, logomarca etc. <p>ANOTAÇÕES: indicação de prioridade (urgente), data e hora do envio e do recebimento, indicação de anexo etc.</p> <p>CONTEXTO: jurídico administrativo, de proveniência, de procedimentos, documental e tecnológico (hardware, software e formatos).</p> <p>SUPORTE: constitui o contexto tecnológico, mais especificamente na categoria hardware.</p> <p>ATRIBUTOS: nome do autor, do destinatário, formato, direitos autorais etc.</p> <p>Obs.: Na maioria das vezes, os atributos encontram-se expressos em metadados.</p> <p>COMPONENTES DIGITAIS: dados de forma, conteúdo e composição.</p>
--

Fonte: RONDINELLI (2013, p. 243).

O estabelecimento de forma fixa, conteúdo estável, relação orgânica, contexto identificável, ação e pessoas na criação de documentos em ambientes digitais dota-os de características que os diferem de dados e informações. Essa diferenciação é relevante para a identificação dos documentos arquivísticos, especialmente os digitais, porque seu tratamento ocorre conforme os preceitos da Arquivologia.

No momento em que são conhecidas as características e partes constituintes do documento arquivístico digital, torna-se pertinente a seguinte questão: como explicitar que tal documento é semântico ao ser humano? Pesquisadores do Projeto InterPARES²⁴ apresentam como resposta a “forma manifestada” do documento arquivístico.

Um documento arquivístico digital manifestado é aquele que, com o processamento

24 Disponível em: <<http://www.interpares.org>>.

de seus componentes digitais²⁵, fica compreensível ao ser humano ou a outro sistema informatizado. A forma manifestada do documento consiste na sua apresentação via *links* entre os diferentes arquivos que compõem o objeto digital em questão: um texto, um áudio, uma fotografia, uma planilha de cálculo, uma planta de arquitetura etc.

Rondinelli (2013, p. 241-260) esclarece que o componente digital é uma parte constituinte exclusiva do documento digital; a cadeia de bits (*bit strings*) desse último é “processada por um *software* que a transforma em documento passível de leitura e compreensão” (p. 241) – a forma manifestada – após ter sido solicitada sua exibição por um humano, por exemplo. Ainda segundo a mesma autora,

as cadeias de bits constituem os chamados componentes digitais, os quais compreendem dados que determinam a forma do documento e os que definem seu conteúdo. Há ainda um terceiro grupo de dados, aqueles que regulam o processamento dos dois primeiros, quando necessário, bem como os metadados relacionados ao documento. São os chamados dados de composição. Assim, *um componente digital pode conter dados de forma e/ou de conteúdo e/ou de composição*. (RONDINELLI, 2013, p. 241-242, grifos nossos).

Nesse sentido, este trabalho (não impresso) que está sendo lido neste momento é um exemplo de documento digital na sua forma manifestada, ou seja, um arquivo digital com um componente digital, contendo dados de forma e conteúdo, que é o arquivo *portable document format* (PDF).

Há, ainda, o documento arquivístico armazenado, dotado das características citadas no quadro 2 e que tem de ser interpretado para ser compreensível. São documentos na forma armazenada, por exemplo, os “programas conhecidos como *patches*” para interação de um instrumento musical com um computador ou as tabelas de um banco de dados de folha de pagamento de funcionários. (RONDINELLI, 2013).

De acordo com Duranti e Thibodeau (2008),

os componentes digitais e o documento gerado a partir desses componentes podem constituir documentos arquivísticos relacionados, mas diferentes: o documento arquivístico armazenado e o “manifestado” [...] A principal razão para a manutenção do documento arquivístico armazenado é possibilitar gerar o documento arquivístico manifestado, o qual é preservado para comunicar a informação para pessoas ou outros sistemas²⁶. (DURANTI; THIBODEAU, 2008, p. 384, tradução

25 Componente digital é o “objeto digital que é parte de um ou mais documentos digitais, e os metadados necessários para ordenar, estruturar ou manifestar seu conteúdo e forma, que requer determinadas ações de preservação”. Glossário da Câmara Técnica de Documentos Eletrônicos (CTDE). Disponível em: <http://www.documentoseletronicos.arquivonacional.gov.br/media/publicacoes/glossario/2010glossario_v5.1.pdf>.

26 the digital components and the document reproduced from these components may constitute, that is, related

nossa, grifos do autor).

O que se discutiu visa a apresentar os documentos arquivísticos digitais também como fontes para a formação de *datasets* publicados nas seções dos sítios dos órgãos e entidades governamentais destinadas a dados abertos. A seguir, apresentam-se as qualidades do documento arquivístico digital, porque elas colaboram para com as qualidades dos dados que se fazem essenciais ao reuso de quaisquer interessados.

2.2.1 As qualidades de um documento arquivístico digital

Os documentos arquivísticos constituem-se em evidências por registrarem as atividades desenvolvidas por uma pessoa física ou jurídica e, para isso, precisam apresentar determinadas qualidades, tais como ser “íntegro, fiel, adequado, completo e significativo, compreensível e utilizável, autêntico, inalterado”²⁷ (Barran et al. [2008], p. 11, tradução nossa). Duranti (1994) também elenca as seguintes características (reconhecidas ainda como qualidades) do documento arquivístico: autenticidade, naturalidade, inter-relacionamento, unicidade, imparcialidade.

O Projeto InterPARES 2 (2010), em suas “Diretrizes para o produtor”, apresenta outras qualidades do documento arquivístico digital, como fixidez (forma fixa e conteúdo estável), variabilidade limitada, confiabilidade, acessibilidade, acurácia e completeza.

No glossário da Câmara Técnica de Documentos Eletrônicos (CTDE), do Conselho Nacional de Arquivos, constam as definições das seguintes qualidades do documento arquivístico: acessibilidade, autenticidade, completeza, confiabilidade, organicidade e integridade. (CONSELHO NACIONAL DE ARQUIVOS, 2010).

Neste trabalho, ao elencarem-se 11 qualidades de um documento arquivístico digital (Quadro 3), busca-se uma comparação com as qualidades de um dado.

but distinct records: the digitally stored record(s), and the “manifested record” [...] The primary purpose of keeping the stored record is to be able to reproduce the manifest record, while the manifest record is preserved to communicate information to persons or other systems.

27 integral, fiel, adecuado, completo y significativo, comprensible y utilizable, auténtico, inalterado.

Quadro 3 – Qualidades de um documento arquivístico digital.

<p>Acessibilidade: Facilidade no acesso ao conteúdo e ao significado de um objeto digital.</p>	<p>Autenticidade: Qualidade de um documento ser o que diz ser e que está livre de adulteração ou qualquer outro tipo de corrupção.</p>	<p>Confiabilidade: Credibilidade de um documento arquivístico como fonte de uma afirmação do fato. Existe quando um documento arquivístico pode sustentar o fato ao qual se refere, e é estabelecida pelo exame da completeza, da forma do documento e do grau de controle exercido no processo de sua produção</p>
<p>Fixidez: Qualidade de um documento arquivístico que assegura a forma fixa e o conteúdo estável.</p>	<p>Imparcialidade: As razões por que os documentos são produzidos (para desenvolver atividades) e as circunstâncias de sua criação (rotinas processuais) asseguram que não são escritos na intenção ou para a informação da posteridade.</p>	<p>Inter-relacionamento (relação orgânica): Vínculos que os documentos arquivísticos guardam entre si e que expressam as funções e atividades da pessoa ou organização que os produziu.</p>
<p>Naturalidade: Diz respeito à maneira como os documentos se acumulam no curso das transações, de acordo com as necessidades da matéria em pauta.</p>	<p>Unicidade: Cada documento assume um lugar único na estrutura documental do grupo ao qual pertence.</p>	<p>Variabilidade limitada: Qualidade de um documento arquivístico que assegura que suas apresentações documentais são limitadas e controladas por regras fixas e um armazenamento estável do conteúdo, da forma e da composição, de modo que a mesma interação, pesquisa, busca ou atividade por parte do usuário sempre produza o mesmo resultado.</p>
<p>Acurácia: Grau de precisão, correção, verdade e ausência de erros e distorções existentes nos dados contidos nos materiais.</p>	<p>Completeza: Atributo de um documento arquivístico que se refere à presença de todos os elementos intrínsecos e extrínsecos exigidos pela organização produtora e pelo sistema jurídico-administrativo a que pertence, de maneira a ser capaz de gerar consequências.</p>	

Fonte: elaboração própria, com base em DURANTI (1994), GLOSSÁRIO CTDE (2010) e PROJETO INTERPARES 2 (2010).

Os estudos de diplomatistas e arquivistas sobre as qualidades de um documento arquivístico digital demonstram sua relevância, pois elas “[...] tornam a análise dos registros documentais o método básico pelo qual se pode alcançar a compreensão do passado tanto

imediatamente quanto histórico, seja com propósitos administrativos ou culturais” (DURANTI, 1994, p. 4). Na atualidade, em que o imediatismo tem sido preponderante e as iniciativas de acesso aberto têm-se expandido, as qualidades de um documento arquivístico são úteis também para a compreensão do presente.

Cabe apontar que, na Ciência da Computação e em Sistemas de Informação, são realizados estudos sobre qualidade de dados. Isso demonstra outra oportunidade de estudos conjuntos entre a Arquivologia e as ciências citadas acima.

Após a identificação das qualidades dos documentos arquivísticos digitais, faz-se essencial a adoção de recursos e procedimentos para que elas sejam mantidas. Um recurso a ser adotado, por exemplo, seria a implantação de um repositório digital confiável, o que também contribuirá para que os *datasets* sejam disponibilizados para qualquer cidadão interessado em (re)utilizá-los.

2.3 REPOSITÓRIO DIGITAL CONFIÁVEL PARA DOCUMENTOS ARQUIVÍSTICOS

O documento arquivístico digital, para cumprir sua função primária, assim como servir de fonte de pesquisa histórica e informacional, exige, para sua gestão, preservação e acesso, procedimentos fundamentados na Arquivologia, recursos tecnológicos e políticas institucionais. Partindo-se dessa perspectiva, far-se-á aqui uma abordagem sobre o repositório digital e o repositório digital confiável para documentos arquivísticos.

Segundo Viana, Márdero Arellano e Shintaku (2006, p. 3 apud CAMARGO; VIDOTTI, 2009), “um repositório digital é uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos de tempo e prover o acesso apropriado”. Assim, um repositório digital pode ser implementado para o gerenciamento de diferentes materiais digitais, como os bibliográficos, os artísticos, os científicos e culturais, os arquivísticos, dentre outros. Para fins desta pesquisa, deter-se-á sobre o armazenamento do documento arquivístico digital.

A Câmara Técnica de Documentos Eletrônicos (CTDE) publicou, em 2014, as “Diretrizes para implementação de repositórios digitais confiáveis de documentos arquivísticos”²⁸ cujos públicos-alvo são os órgãos e entidades integrantes do Sistema Nacional

²⁸ Anexo da Resolução nº 39, do Conselho Nacional de Arquivos (Conarq). Disponível em: <http://www.conarq.arquivonacional.gov.br/media/publicacoes/resol_conarq_39_repositorios.pdf>.

de Arquivos (SINAR). Nessas Diretrizes, a CTDE faz o alerta de que “um repositório digital não se resume a uma solução informatizada para armazenamento (*storage*), que é apenas um componente do repositório”, e que, além dos “elementos de *hardware*, *software* e metadados”, incluem-se “infraestrutura organizacional e procedimentos normativos e técnicos” (CONSELHO NACIONAL DE ARQUIVOS, 2014, p. 9). Dessa forma, seria possível afirmar que um repositório digital é o próprio órgão que o detém e por isso um repositório institucional?

Acerca disso, cabem algumas informações sobre repositório institucional, objeto de estudo especialmente da área de Biblioteconomia. A expressão pode induzir ao entendimento de que é um repositório no âmbito de uma instituição, porém,

Repositório institucional foi o termo utilizado para representar um novo serviço bibliotecário cuja ênfase é constituir, gerenciar e, sobretudo, disseminar amplamente coleções digitais de informação científica, de modo que a comunicação, acesso e uso de resultados de pesquisa fossem expandidos. (COSTA; LEITE, 2009, p. 163).

A partir do que afirmam Costa e Leite, compreende-se que um repositório institucional não é o órgão ou entidade que definiu em sua política institucional o estabelecimento desse tipo de repositório ou *serviço bibliotecário*.

Outra observação a ser feita diz respeito ao fato de que um repositório digital diferencia-se de um banco de dados, pois, esse último é

um conjunto lógico e ordenado de dados que possuem algum significado, e não uma coleção aleatória sem um fim ou objetivo específico. [e] é construído e povoado com dados que têm um determinado objetivo, com usuários e aplicações desenvolvidas para manipulá-los. (ALVES, 2009, p. 22-23).

O repositório digital abrange a infraestrutura organizacional com suas políticas, procedimentos técnico-normativos e parque tecnológico.

Quando um repositório digital tem a capacidade de manter autênticos, preserváveis e acessíveis os materiais digitais nele armazenados, qualifica-se como confiável, sobretudo se estiver de acordo com os critérios do documento denominado *Auditoria e Certificação de Repositórios Confiáveis: critérios e lista de conferência* (TRAC, da sigla em inglês²⁹) e obter a certificação interna ou de terceiros. Outras iniciativas para certificação de repositórios digitais são a “*Audit And Certification of Trustworthy Digital Repositories* (ACTDR), a *Catalogue of Criteria for Trusted Digital Repositories da Network of Expertise in long-term*

²⁹ *Trustworthy Repositories Audit & Certification: Criteria and Checklist.*

STORage (NESTOR) e a *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA)” (SANTOS; FLORES, 2015, p. 210).

Um repositório digital confiável, para receber a certificação, significa ter sido atestado em conformidade com as melhores práticas e requisitos quanto a gestão e preservação de objetos digitais. De acordo com o *Center for Research Library* e o *Online Computer Library Center* (2007),

Nesse novo ambiente de informação e responsabilidades digitais, todos os repositórios precisarão estabelecer seus estatutos de confiáveis. Adotar esta lista de conferência [TRAC] – iniciando com uma auto-avaliação – é um dos mecanismos que um repositório pode usar para entender suas capacidades, como se posicionar diante de possíveis ameaças e quaisquer outros riscos inerentes aos seus sistemas [...] Finalmente, receber o estatuto de confiável não é um comprometimento de via única – conquistado e esquecido. Para manter tal estatuto, um repositório precisará conduzir um ciclo regular de auditoria e/ou certificação³⁰. (CENTER FOR RESEARCH LIBRARY; ONLINE COMPUTER LIBRARY CENTER, 2007, p. 3-4).

Para armazenamento e gerenciamento de documentos arquivísticos em suas idades corrente, intermediária e permanente, quando implementado, cabe a um repositório digital:

- gerenciar os documentos e metadados de acordo com as práticas e normas da Arquivologia, especificamente relacionadas à gestão documental, descrição arquivística multinível e preservação; e
- proteger as características do documento arquivístico, em especial a autenticidade (identidade e integridade) e a relação orgânica entre os documentos. (CONSELHO NACIONAL DE ARQUIVOS, 2014, p. 9).

Por interpolação das concepções de repositório digital, repositório digital confiável e documento arquivístico, chega-se ao repositório digital confiável para documentos arquivísticos, ambiente onde é feita a captura, o armazenamento, a gerência e o acesso aos documentos produzidos e recebidos por uma entidade no curso de suas atividades.

O uso de um repositório digital confiável para documentos arquivísticos tem estimulado estudos e desenvolvimento de *softwares* que protegem as características desses documentos, sendo reconhecidos como “repositórios arquivísticos digitais confiáveis”; têm-se como exemplos em *software* livre o Archivematica e o Repositório de Objetos Digitais

30 In this new environment of digital information and digital responsibilities, all repositories will need to establish their trusted status. Use of this audit checklist—beginning with self assessment— is one mechanism a repository can use to understand its capabilities, where it stands against potential threats, and any other risks inherent in its systems [...] Finally, attaining trusted status is not a one-time accomplishment —achieved and forgotten. To retain trusted status, a repository will need to undertake a regular cycle of audit and/or certification.

Autênticos (RODA) (FLORES, [2014]). Para o autor,

[...] Um repositório arquivístico digital é um ambiente de preservação digital, com funcionalidades, com estratégias de preservação implícitas, com requisitos arquivísticos, com navegação multinível e garantia da autenticidade e relação orgânica dos documentos ali armazenados, gerenciados e preservados. (FLORES, [2014], p. 16).

Na seção 2.2 desta pesquisa, argumentou-se que o documento arquivístico digital também é uma fonte de dados. Por isso, entende-se que a implantação de um repositório digital confiável para documentos arquivísticos é um recurso para a manutenção de documentos digitais que servem de fonte de dados digitais primários, completos, atuais e acessíveis.

3 GESTÃO DE DADOS

A relevância da gestão de dados nos ambientes corporativos, sejam públicos ou privados, é crescente, principalmente porque a produção e o armazenamento de dados por meio das tecnologias de informação e comunicação têm sido realizados em escalas de gigabytes, terabytes, petabytes (MAYER-SCHÖNBERGER; CUKIER, 2013). O aumento dessa produção de dados tem sido acompanhado pela preocupação de como estabelecer o equilíbrio entre volume, qualidade, uso e infraestrutura tecnológica.

Mas um certo otimismo é apresentado por adeptos da tecnologia *big data*³¹, pois com um grande volume de dados é possível chegar a melhores resultados do que com amostras reduzidas de dados sobre determinado assunto ou objeto de estudo (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 10) e, aliado a isso, a redução dos custos de produção dos equipamentos eletrônicos.

Nesse cenário, é pertinente a indagação sobre como se devem recuperar aqueles dados relevantes para uma tomada de decisão, prestação de contas ou transparência nos negócios, ou para o reúso por uma sociedade. A Gestão de Dados, como disciplina, oferece o arcabouço necessário para esse desafio.

Reconhecer e reutilizar dados que agregam valor aos processos de negócios da organização é possível por meio de planejamento, controle e administração dos procedimentos adequados para que tais dados sejam gerenciados com eficiência, eficácia e efetividade.

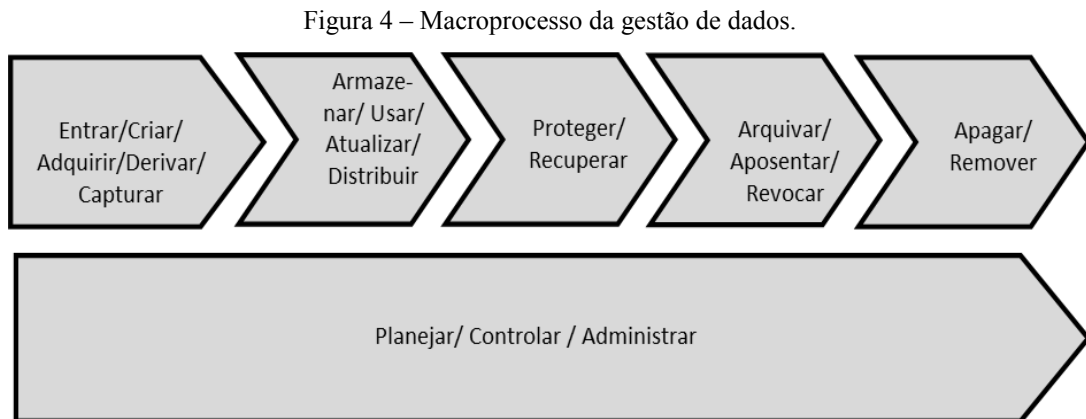
A definição de gestão de dados é o “processo administrativo pelo qual o dado solicitado é adquirido, validado, armazenado, protegido e processado, cujas acessibilidade, confiabilidade e atualidade são garantidas para satisfazer as necessidades dos usuários”³² (BUSINESS DICTIONARY, [201-], tradução nossa). Segundo Rêgo (2013, p. 25), a gestão é uma função “[...] responsável por definir, planejar, implantar e executar: estratégias, procedimentos e práticas necessárias para gerenciar de forma efetiva os recursos de dados e informações das organizações”. Pode-se observar uma complementariedade entre as

31 Noção cuja tradução livre é “grande dado”. A essência para sua compreensão consiste em ter-se um volumoso *corpus* de informação que permita conexões, as quais não eram antes identificadas numa amostra pouco volumosa. (MAYER-SCHÖNBERGER; CUKIER, 2013).

32 Administrative process by which the required data is acquired, validated, stored, protected, and processed, and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users.

definições, pois o primeiro direciona-se aos dados e o segundo à gestão propriamente dita. Já de acordo com Mosley (2008), a gestão de dados tem como função “o controle efetivo e a otimização dos ativos de dados” de uma organização.

Com base nessas definições, apresenta-se o macroprocesso da gestão de dados na figura 4.



Fonte: BAIÃO (2014).

Ao retomar-se a função da gestão de dados adotada por Mosley, faz-se necessária a reflexão sobre de que maneira é possível ter o controle e otimização dos ativos de dados de uma organização. Pode-se, preliminarmente, obter como resposta a ampliação e a atualização da infraestrutura tecnológica, com ferramentas de armazenamento e distribuição de dados. No entanto, essa decisão não é suficiente; há que se compreender o arcabouço teórico proposto pela disciplina Gestão de Dados.

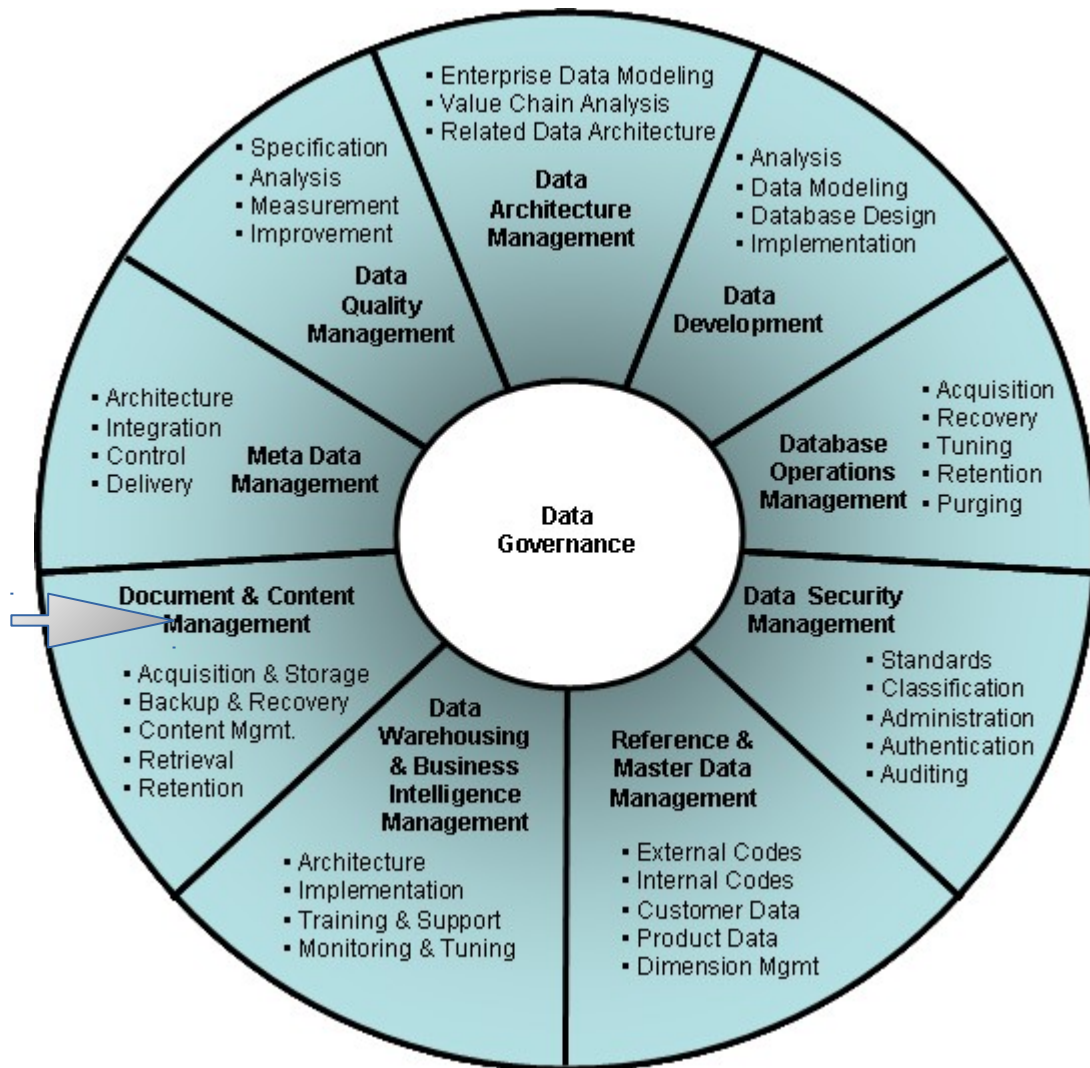
Entende-se a Gestão de Dados como uma disciplina composta por matérias curriculares; no entanto, no *Data Management Book of Knowledge*, do *The Data Management Association* (DAMA-DMBOK), ela é estudada como uma função, tendo, em seu escopo, as seguintes atividades:

governança de dados, gestão da arquitetura de dados, gestão do desenvolvimento dos dados, gestão de operação de dados, gestão da segurança dos dados, gestão de dados mestres e dados de referência, gestão de *data warehousing* e *business intelligence*, gestão da documentação e conteúdo, gestão de metadados e gestão da qualidade dos dados. (RÊGO, 2013, p. 27-28, grifos do autor).

A Gestão de Dados envolve diferentes atividades, como as representadas na figura 5, cujos profissionais das áreas de Informática e de Negócios devem ter conhecimento. Cabe ressaltar que esses profissionais, com base em suas experiências teóricas e práticas, têm

responsabilidades compartilhadas nas tomadas de decisões quanto à gestão de dados em uma organização.

Figura 5 – Atividades englobadas pela Gestão de Dados.



Fonte: MOSLEY (2008, p. 10). A tradução dos termos em inglês não foi realizada para evitar equívocos conceituais.

Dentre as atividades abordadas pela Gestão de Dados, está a Gestão de Documentos e Conteúdo (*Document and Content Management*), que consiste nas “atividades de planejamento, implementação e controle para armazenar, proteger e acessar dados contidos em arquivos eletrônicos e registros físicos (incluindo textos, gráficos, imagem, áudio, vídeo)”³³ (MOSLEY, 2008, p. 11, tradução nossa). Em detalhe, as atividades são:

33 Planning, implementation and control activities to store, protect and access data found within electronic files and physical records (including text, graphics, image, audio, video).

1. Planejar o gerenciamento de documentos físicos e eletrônicos.
2. Definir e manter a taxonomia da corporação.
3. Implementar e manter os sistemas de armazenamento de documentos.
4. Adquirir e armazenar documentos.
5. Indexar conteúdos informacionais dos documentos.
6. Fazer *backup* e recuperar documentos.
7. Apoiar a análise de conteúdos documentais.
8. Apoiar a atualização, tramitação e acesso aos documentos.
9. Monitorar e ajustar o desempenho da localização de documentos.
10. Arquivar, localizar e eliminar documentos.
11. Auditar a gestão de documentos e conteúdo³⁴. (MOSLEY, 2008, p. 18, tradução nossa).

As organizações contemporâneas têm passado por mudanças quanto à realização de seus negócios devido a inovações tecnológicas e ao comportamento das pessoas envolvidas nos negócios: as empresas não trabalham mais isoladas; dependem do contexto em que estão inseridas. (JOHANNESSON, 2007).

O estudo da gestão de dados, além do campo da Ciência da Computação e de Sistemas de Informação, é uma oportunidade para estabelecer relações interdisciplinares (MADNICK et al. [2012]) em que os corpos teórico e metodológico das ciências humanas podem colaborar com as ciências exatas. É preponderante que arquivistas e informáticos sejam conscientes de que os documentos digitais produzidos e recebidos por uma organização têm caráter de evidência³⁵, administrativo, informativo e, possivelmente, histórico. Por isso, devem ser geridos com base nos preceitos teóricos da Arquivologia, conjuntamente aos da Ciência da Computação e Sistemas de Informação.

A adoção das tecnologias de informação e comunicação como recurso no desenvolvimento das funções e atividades numa organização fez com que muitos arquivistas ficassem deslumbrados a ponto de não distinguirem um documento interlocutório de um documento arquivístico, nem diferenciar estes últimos de simples dados. (DURANTI, 1994).

Ao realizar-se a gestão de documentos digitais, faz-se necessário também que arquivistas e informáticos identifiquem o contexto, ou seja, o ambiente em que ocorreu uma determinada ação, pois,

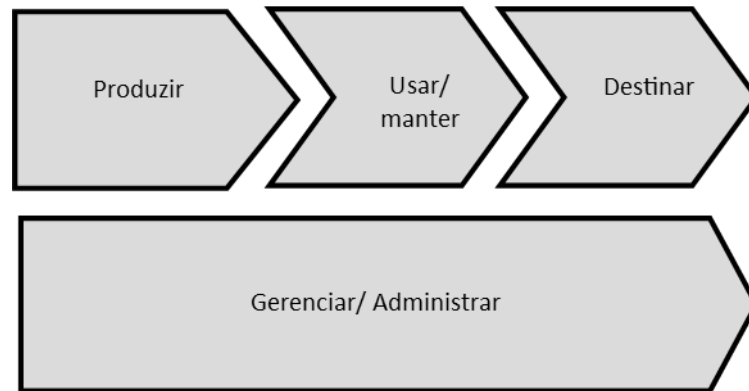
34 1. Plan for Managing Electronic & Physical Documents; 2. Define & Maintain Enterprise Taxonomies; 3. Implement & Maintain Document Storage Systems; 4. Acquire & Store Documents; 5. Index Document Information Contents; 6. Backup & Recover Documents; 7. Support Document Content Analysis; 8. Support Document Access, Circulation & Update; 9. Monitor & Tune Document Retrieval Performance; 10. Archive, Retrieve & Purge Documents; 11. Audit Document & Content Management.

35 Optou-se aqui pelo termo evidência, em detrimento do termo probatório, por concordar-se com o conceito arquivístico de evidência defendido por Meehan (2006, apud RONDINELLI, 2013, p. 219-223).

sem a clara evidência do contexto, não existem garantias circunstanciais de fidedignidade, nenhuma possibilidade de autenticação, nenhuma possibilidade de identificação do documento original e nenhuma justificativa para a ausência daquela documentação que deveria existir. (DURANTI, 1994, p.59).

Com essa afirmação de Duranti, pode-se dizer que, para os estudos sobre os contextos jurídico-administrativo, de proveniência, de procedimentos, documental e tecnológico (DURANTI; THIBODEAU, 2008, p. 351), os arquivistas colaboram com os profissionais de tecnologia da informação, que, por sua vez, dão suporte à gestão de documentos arquivísticos digitais. Para fins de representação, ilustram-se as fases integrantes dessa gestão na figura 6.

Figura 6 – Macroprocesso da gestão de documentos.



Fonte: elaboração própria.

A partir da equiparação entre as figuras 5, “Macroprocesso da gestão de dados”, e 6, “Macroprocesso da gestão de documentos”, é possível observar algumas semelhanças. No entanto, a que se devem tais semelhanças? Uma das possíveis respostas está na adoção do termo *gestão* – cuja definição já foi apresentada – aplicado ao objeto de estudo de cada área específica do conhecimento (Quadro 4).

Quadro 4 – Equiparação dos macroprocessos de gestão de dados e documentos arquivísticos.

GESTÃO		
	DADOS	DOCUMENTOS ARQUIVÍSTICOS
Atividade	<i>Planejar, controlar, administrar</i>	<i>Gerenciar, administrar</i>
FASES	Entrar, criar, adquirir, derivar, capturar	Produzir, capturar
	Armazenar, usar, atualizar, distribuir	Usar, manter
	Proteger, recuperar	Usar, manter
	Arquivar, aposentar, revocar	Destinar
	Apagar, remover	Destinar

Fonte: elaboração própria.

3.1 O DADO DIGITAL E ALGUMAS DE SUAS QUALIDADES

De acordo com Fisher (2009), as organizações estão dando maior significado à qualidade de um dado. No entanto, o que significa essa qualidade? Quais os atributos que qualificam um dado como sendo “bom” para uso de um consumidor final?

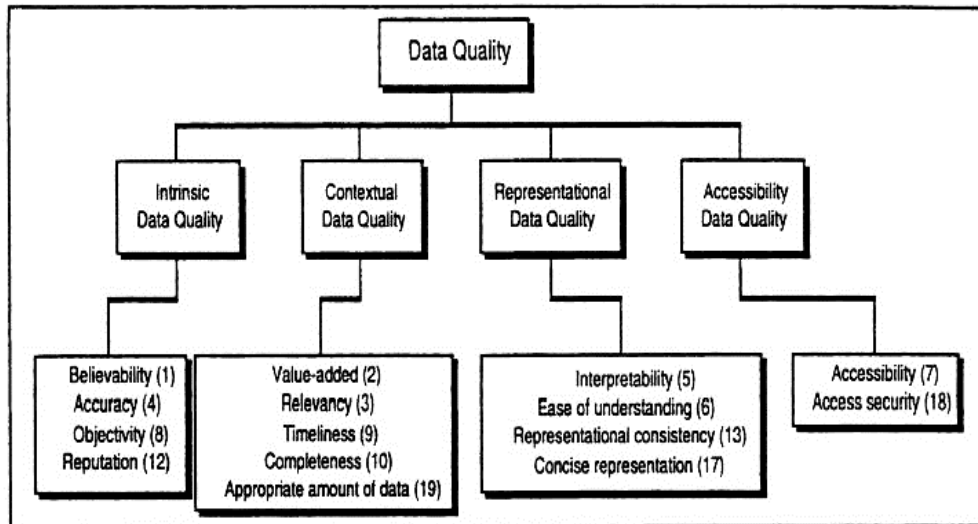
Diferentes estudos têm sido realizados para identificar quais características são pertinentes a um dado, de maneira que tenha qualidade para o consumidor final (SADIQ; YEGANEH; INDULSKA, 2011). Tal motivação deve-se ao valor que tem sido agregado ao dado, por ele ser a base das informações geradas pelas pessoas nos processos de negócios corporativos.

A qualidade de dados é relevante para a precisão nas tomadas de decisões, pois se elas são concretizadas a partir da análise de dados imprecisos ou desatualizados, por exemplo, esse procedimento pode acarretar prejuízos à organização. (MARTINS, 2009).

Na NBR ISO 9000 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2000), qualidade é definida como grau no qual “o conjunto de propriedades diferenciadoras inerentes” satisfazem a requisitos (necessidades). Dentre as pesquisas desenvolvidas para identificar as características quanto à qualidade de um dado, Wang e Strong (1996) definiram um arcabouço conceitual (Figura 7) a partir dos resultados obtidos através da realização de entrevistas, cujo público-alvo foi o consumidor de dados. A pesquisa é uma das principais

referências internacionais sobre esse tema e também norteará o estudo aqui desenvolvido.

Figura 7 – Um arcabouço conceitual sobre a qualidade de dado.



Fonte: WANG; STRONG (1996).

Martins (2009), em sua dissertação, propôs uma tradução (Quadro 5) do arcabouço conceitual elaborado por Wang e Strong (1996), a qual será adotada neste trabalho com a ressalva de que, na categoria “Contextual”, consta o fator “Utilidade” que não existe no original de Wang e Strong.

Quadro 5 – Qualidade de dados: categoria e fatores.

Qualidade de dados	
<i>Categoria de qualidade</i>	<i>Fatores de qualidade</i>
Intrinseca	Credibilidade, Precisão, Objetividade, Reputação
Contextual	Valor Agregado, Relevância, Utilidade, Atualidade, Completeza, Volume de Dados
Representacional	Interpretabilidade, Facilidade de Entendimento, Consistência Representacional, Representação Concisa
Acessibilidade	Acessibilidade, Segurança de Acesso

Fonte: MARTINS (2009).

O estudo de Sadiq, Yeganeh e Indulska (2011) feito para a compreensão do panorama atual das pesquisas sobre qualidade de dados apresenta uma taxonomia das palavras-chave que remetem à qualidade de dados. Essa taxonomia foi elaborada para análise dos 31.701 artigos selecionados pelos autores como objeto de estudo. Da nomenclatura adotada pelos

autores, destaca-se no quadro 6 o grupo “dimensões de qualidade de dados”.

Quadro 6 – Dimensões de qualidade de dados.

Dimensões de qualidade de dados <i>Data quality dimensions</i>	Usabilidade <i>Usability</i>
	Compreensão da informação <i>Information understandability</i>
	Relevância da informação <i>Information relevance</i>
	Adequado para uso <i>Fitness for use</i>
	Completeza do dado <i>Data completeness</i>
	Confiabilidade do dado <i>Data reliability</i>
	Acurácia do dado <i>Data accuracy</i>
	Utilidade da informação <i>Information usefulness</i>
	Atualidade do dado <i>Data currency</i>
	Consistência do dado <i>Data consistency</i>
	Incerteza do dado > banco de dados probabilísticos <i>Data uncertainty > probabilistic databases</i>
	Privacidade do dado <i>Data privacy</i>

Fonte: elaboração própria, com base em SADIQ; YEGANEH; INDULSKA (2011, tradução nossa).

Ao compararem-se as “dimensões de qualidade de dados” do quadro 6 ao arcabouço conceitual da figura 7, pode-se observar que são comuns *accuracy*, *completeness* e *information relevance* (Quadro 7). Optou-se por não traduzir os termos em inglês para que não houvesse comprometimento na compreensão das definições feitas pelos autores.

Quadro 7 – Equiparação de dimensões de qualidade de dados.

WANG; STRONG, 1996	SADIQ; YEGANEH; INDULSKA, 2011
Segurança de acesso <i>Access security</i>	
Acessibilidade <i>Accessibility</i>	
Acurácia/precisão <i>Accuracy</i>	Acurácia do dado <i>data Accuracy</i>
Volume de dados <i>Appropriate amount of data</i>	
Credibilidade <i>Believability</i>	
Completeza <i>Completeness</i>	Completeza do dado <i>data Completeness</i>
Representação concisa <i>Concise representation</i>	
Facilidade de entendimento <i>Easy of understanding</i>	Compreensão da informação <i>Information understandability</i>
Interpretabilidade <i>Interpretability</i>	
Objetividade <i>Objectivity</i>	
Relevância <i>Relevancy</i>	Relevância da informação <i>Information relevance</i>
Consistência representacional <i>Representational consistency</i>	Consistência do dado <i>data Consistency</i>
Reputação <i>Reputation</i>	
Atualidade <i>Timeliness</i>	Atualidade do dado <i>data Currency</i>
Valor agregado <i>Value-added</i>	
	Adequado para uso <i>Fitness for use</i>
	Utilidade da informação <i>Information usefulness</i>
	Privacidade do dado <i>data Privacy</i>
	Confiabilidade do dado <i>data Reliability</i>
	Incerteza do dado > banco de dados probabilísticos <i>Uncertainty (data) > probabilistic databases</i>
	Usabilidade <i>Usability</i>

Fonte: elaboração própria, com base em WANG; STRONG (1996) e SADIQ; YEGANEH; INDULSKA (2011), tradução nossa.

É pertinente também mencionar a ISO 8000, norma internacional (em desenvolvimento) cuja finalidade é estabelecer os parâmetros de qualidade de dados. A seguir, a divisão da norma em partes e respectivos números: “Introdução-100, Proveniência-120, Acurácia-130 e Completeza-140”³⁶. (BENSON, 2009, tradução nossa).

Há autores, como Peter R. Benson, Yang W. Lee, Carlo Batini, que também desenvolvem estudos acerca do tema qualidade de dados, mas com aprofundamento sobre

36 Overview – 100, Provenance – 120, Accuracy – 130 and Completeness – 140.

métricas, avaliação e tipos de conhecimento necessários para garantir a qualidade dos dados a serem utilizados numa organização. No entanto, neste trabalho, optou-se por se deter na identificação das qualidades necessárias a um dado e, assim, equipará-las com as de um documento arquivístico digital, conforme apresentado na seção 2.2.1.

As qualidades de um objeto digital são importantes para satisfazer as necessidades de indivíduos ou organizações que dele faz uso. A revisão teórica feita conduziu à identificação de que acessibilidade, precisão/acurácia, completeza e confiabilidade são qualidades prementes a um dado e também serem inerentes aos documentos arquivísticos digitais (Quadro 8).

Quadro 8 – Qualidade de dado e de documento arquivístico.

WANG; STRONG, 1996	SADIQ; YEGANEH; INDULSKA, 2011	DURANTI (1994); GLOSSÁRIO CTDE (2010); PROJETO INTERPARES (2010)
Segurança de acesso <i>Access security</i>		
Acessibilidade <i>Accessibility</i>		Acessibilidade
Acurácia/precisão <i>Accuracy</i>	Acurácia do dado data <i>Accuracy</i>	Acurácia
Volume de dados <i>Appropriate amount of data</i>		
Credibilidade <i>Believability</i>		
Completeza <i>Completeness</i>	Completeza do dado data <i>Completeness</i>	Completeza
Representação concisa <i>Concise representation</i>		
Facilidade de entendimento <i>Easy of understanding</i>	Compreensão da informação <i>Information understandability</i>	
Interpretabilidade <i>Interpretability</i>		
Objetividade <i>Objectivity</i>		
Relevância <i>Relevancy</i>	Relevância da informação <i>Information relevance</i>	
Consistência representacional <i>Representational consistency</i>	Consistência do dado data <i>Consistency</i>	
Reputação <i>Reputation</i>		
Atualidade <i>Timeliness</i>	Atualidade do dado data <i>Currency</i>	
Valor agregado <i>Value-added</i>		
	Adequado para uso <i>Fitness for use</i>	
	Utilidade da informação <i>Information usefulness</i>	
	Privacidade do dado data <i>Privacy</i>	
	Confiabilidade do dado data <i>Reliability</i>	Confiabilidade
	Incerteza do dado > banco de dados probabilísticos Uncertainty (data) > probabilistic databases	
	Usabilidade <i>Usability</i>	
		Autenticidade
		Fixidez
		Imparcialidade
		Inter-relacionamento (relação orgânica)
		Naturalidade
		Unicidade
		Variabilidade limitada

Fonte: elaboração própria, com base em WANG e STRONG (1996); SADIQ, YEGANEH e INDULSKA (2011); DURANTI (1994); GLOSSÁRIO CTDE (2010); PROJETO INTERPARES (2010).

O arcabouço teórico acerca do tema “qualidade de dados” é igualmente significativo, pois pesquisas têm sido desenvolvidas desde os anos de 1990 até recentemente e, ainda assim, continuam sendo necessários novos estudos ou revisitações de trabalhos anteriores, a fim de encontrar respostas para os questionamentos acerca do uso e reúso de dados nos ambientes corporativos, públicos ou privados.

Assim, foi possível observar que há uma proximidade teórica entre o estudo da gestão de documentos arquivísticos pela Arquivologia e o estudo da gestão de dados pelas áreas de Computação e Sistemas da Informação, caracterizando uma oportunidade para aproximações a fim de estabelecer correlações entre suas teorias e práticas, além de demonstrar que as áreas de Humanas e Exatas podem colaborar umas com as outras, resguardadas suas respectivas especificidades.

4 OPEN DATA = DADOS ABERTOS

Nas administrações públicas, seja no cenário nacional ou internacional, o uso de TICs para agilizar seus processos de negócios e dar transparência informacional aos cidadãos contribuiu para a constituição dos governos eletrônicos que, dentre suas diferentes concepções, têm como objetivo principal a melhoria dos serviços públicos e a participação cidadã nas políticas públicas.

A transparência da informação pública apoiada, sobretudo, no uso de TICs é insuficiente para uma sociedade democrática em que a participação dos cidadãos faz-se relevante. Em 21 de janeiro de 2009, por exemplo, Barack Obama, presidente dos Estados Unidos, publica o *Memorandum for the Heads of Executive Departments and Agencies – Transparency and Open Government*. Com isso, intensifica-se a ideia de “governo aberto”.

A diretiva desse *Memorandum* é que há um comprometimento da administração de Obama para com a abertura das ações governamentais: “[...] Nós trabalharemos juntos para assegurar a confiança pública e estabelecer um sistema de transparência, participação pública e colaboração. A abertura fortalecerá nossa democracia e promoverá eficiência e eficácia no Governo”³⁷. (THE WHITE HOUSE, 2009, tradução nossa).

A fim de tornarem-se mais transparentes, participativos e colaborativos, os governos dos EUA, Reino Unido, Canadá e Nova Zelândia foram os primeiros a adotarem dados abertos, trazendo a temática a uma posição de destaque a partir dos anos de 2009 (OPEN KNOWLEDGE FOUNDATION, 2012) e, desde então, vem afirmando-se nas agendas políticas de diferentes governos, inclusive o brasileiro. Mas qual é a compreensão que se tem de “aberto”? E o que é “dado aberto”?

Na literatura internacional consultada para esta pesquisa, a definição do termo aberto (*open*), adotado no contexto do “movimento aberto”, não está delimitada. O que ocorre é uma transposição de ideias compartilhadas em diferentes segmentos, como aponta a *Open Knowledge Foundation* (OKF):

A *Open Definition* apresenta o significado de “aberto” com relação a conhecimento, promovendo um espaço em que qualquer pessoa pode participar, e a interoperabilidade é maximizada. [...] Esse significado, essencialmente, condiz com aquele adotado para *software* em *Open Source Definition* e é sinônimo de “livre” ou

³⁷ We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.

“libre”, como em *Definition of Free Cultural Works*. A *Open Definition* inicialmente foi derivada da *Open Source Definition* que, por sua vez, derivou-se das *Debian Free Software Guidelines*³⁸. (OPEN DEFINITION, 2015, tradução nossa, grifos nossos).

Essas derivações apontadas pela OKF refletem o ideário de uma sociedade civil engajada em movimentos de cultura livre, cultura digital, *web 2.0* visando à construção do conhecimento com liberdade e de forma compartilhada.

O ato de publicar um conjunto de dados para reuso por parte de qualquer pessoa vai além do *upload* do arquivo digital na rede de Internet: há de se optar por uma licença³⁹, a fim de autorizar aquilo possível de ser feito a partir desse conjunto de dados. Conforme a *Open Definition* (2015), as seguintes condições devem ser atendidas para que uma “obra” seja considerada aberta para sua distribuição:

- 1) adotar licença aberta, dispor de acesso total e usar formato aberto;
- 2) satisfazer condições quanto à permissão de uso, redistribuição, modificação, separação de partes do todo, compilação, sem discriminação de pessoas ou grupos, propagação, uso para qualquer propósito e sem cobrança de taxas; e
- 3) possuir condições aceitáveis quanto a atribuições de direitos, integridade, compartilhamento sob a mesma licença, notificação sobre direitos autorais, modificação de fonte, proibição de restrição técnica e *non-aggression*.

No que se refere ao ideal desse “movimento aberto”, o que se depreende é que qualquer obra intelectual publicada, principalmente em ambiente digital, não deveria conter restrições que pudessem cercear a ampliação do conhecimento nas diferentes áreas de interesse dos indivíduos de uma sociedade. Nessa mesma trilha, está outra iniciativa, a do acesso aberto (*open access*).

Segundo Peter Suber (2013), as principais e mais influentes referências sobre o movimento *open access* (OA) são *The Budapest OA Initiative* (2002-2012), *The Bethesda Statement on OA* (junho de 2003) e a *Berlin Declaration* (outubro de 2003). O ponto comum dessas referências é que publicações científicas a serem disponibilizadas para usuários na

38 The Open Definition makes precise the meaning of “open” with respect to knowledge, promoting a robust commons in which anyone may participate, and interoperability is maximized. [...] This essential meaning matches that of “open” with respect to software as in the *Open Source Definition* and is synonymous with “free” or “libre” as in the *Definition of Free Cultural Works*. The Open Definition was initially derived from the *Open Source Definition*, which in turn was derived from the *Debian Free Software Guidelines*.

39 DE ROBBIO, Antonella. Dati aperti nella Pubblica Amministrazione tra crescita e trasparenza.

Internet não seriam taxadas e as barreiras relacionadas a permissão e custo seriam removidas, em conformidade com o ideal do movimento OA.

De acordo com a linha do tempo estabelecida por Peter Suber (2015), a iniciativa para acesso aberto teve seu início nos anos de 1960, quando ainda principiava o uso das TICs. Tal movimento se iniciou na área da educação, com o Educational Resources Information Center (ERIC).

Nessa linha do tempo de Suber, que chega até 2009, constam entidades coletivas e pessoas a favor da disponibilização de publicações científicas sem impedimentos, como a precificação e as licenças de uso do material, principalmente quando se trata de pesquisas financiadas com recursos públicos.

Com base em Suber, pode-se inferir que houve uma extensão de uma iniciativa para outra, isto é, de acesso aberto para a dados abertos. Outra influência a ser apontada, também, é a do movimento *software* livre que tem em sua base “quatro liberdades”: executar, estudar, redistribuir e aperfeiçoar um programa para qualquer propósito de forma a colaborar com outros usuários. (FREE SOFTWARE FOUNDATION, 2015).

Cabe destacar que a adesão ao movimento aberto dedicado às obras de referência demanda procedimentos, ou mesmo exigências legais, diferenciados dos aplicados a dados produzidos em âmbitos privados e públicos, sobretudo se são dados pessoais.

O estudo sobre dados abertos requer certa atenção, já que há uma

confusão que se cria em torno da corrente “open” e de maneira ainda mais acentuada em torno da corrente dados abertos. Confunde-se aquilo que é “acessível” no sentido de que está em rede, com o que por sua vez é reutilizável de maneira mais ou menos plena, e o caso dos dados abertos é emblemático. Um catálogo em rede, um OPAC⁴⁰, é obviamente acessível a qualquer pessoa, porém não significa que os seus dados possam ser distribuídos, em bloco ou em parte, e reutilizado para construir um novo catálogo ou reelaborá-lo para fins comerciais⁴¹. (DE ROBBIO, 2013, p. 46, tradução nossa, grifos do autor).

Para dado aberto, a principal definição é a difundida pela OKF, no manual Open Data Handbook: “dado que pode ser livremente usado, reutilizado e redistribuído por qualquer pessoa – sujeito apenas à atribuição de crédito àquele que publicou o dado e à condição de

40 OPAC é o acrônimo de Open Public Access Catalog. <https://www.dur.ac.uk/library/faq/?faqno=64>

41 confusione che si genera attorno alle correnti “open” e in maniera ancor più accentuata attorno alle correnti open data. Si confonde quello che è “accessibile” nel senso che sta in rete, con ciò che invece è riutilizzabile in modo più o meno pieno, e l’esempio degli open data è emblematico. Un catalogo in rete, un OPAC, è ovviamente accessibile a chiunque, ma ciò non significa che i suoi dati possano essere scaricati, in blocco o solo in parte, e riutilizzati per costruire nuovi cataloghi o rielaborati per esempio per scopi commerciali.

publicar também como dado aberto o resultado da combinação entre dados”⁴² (OPEN KNOWLEDGE FOUNDATION, 2012, tradução nossa; OPEN DATA INSTITUTE, 2013).

Para Hall et al. (2012, p. 10, tradução nossa), o dado aberto é definido por critérios e, assim, um *dataset* é um dado aberto se: “1. disponibilizado online; 2. publicado em formato aberto legível por máquina; e 3. licenciado com permissão de reuso por outras pessoas”⁴³.

O Open Data Institute sustenta que um bom dado aberto (*good open data*)

- é possível de ser conectado de um ponto a outro, podendo facilmente ser compartilhado e referenciado.
- é acessível em um padrão, formato estruturado, podendo facilmente ser processado [por máquina].
- tem acessibilidade e consistência garantidas continuamente, então outras pessoas podem confiar neles.
- é rastreável, através de qualquer processamento, direto na fonte, então outros podem descobrir se ele é verdadeiro. (OPEN DATA INSTITUTE, 2013, Guides, tradução nossa).

Há autores, porém, que compreendem dados abertos como uma filosofia (DE ROBBIO, 2013; CALDERÓN, LORENZO, 2010), um conjunto de ideias sobre a coleta de dados sem valor agregado, os quais serão dotados de significado e, por isso, fontes para outros usuários gerarem conhecimento (CALDERÓN; LORENZO, 2010).

Com relação ao uso e reuso de dados disponibilizados, De Robbio (2013) alerta para a questão do direito autoral, levando à reflexão sobre a difusão dos acervos literários e pessoais, já que

Um aspecto que não é secundário e nem sempre percebido pela maior parte dos usuários da rede é que, na ausência de uma licença para pesquisa em rede, impõem-se as normas de direitos autorais, e até mesmo uma obra presente na rede não significa que ela seja aberta no sentido de reutilizável, usável para fins comerciais ou possa ser reelaborada ou manipulada ou considerada “open” no sentido pleno do termo.⁴⁴ (DE ROBBIO, 2013, p. 10, tradução nossa, grifo do autor).

É válido lembrar que a iniciativa para dados abertos tem como objeto a ser trabalhado o conjunto de dados formado sem agregação de valor, em estado bruto. Isso

42 Data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.

43 1. it is made accessible online, 2. it is published in an open machine readable format, 3. it is licensed to allow others to re-use it.

44 Un aspetto non secondario e non sempre noto alla maggior parte degli utenti della rete è che, in mancanza di licenze apposite sulla risorsa di rete, valgono le norme sul diritto d’autore e quindi anche in presenza di un’opera presente in rete ciò non significa che questa sia fruibile nel senso di riutilizzabile, usabile per fini commerciali, o che si possa rielaborare o manipolare o comunque considerare “open” nel senso pieno del termine.

significa que, nesse contexto, o *dataset* não é um documento, tampouco um documento arquivístico, conforme apresentado na seção 2.1. No entanto, se produzido em um outro contexto (de registro das atividades desenvolvidas por um órgão ou entidade), faz-se necessário seu respectivo reconhecimento.

4.1 DADOS GOVERNAMENTAIS ABERTOS

Governança, transparência e *accountability* são termos adotados para tratar do desempenho administrativo-corporativo, seja público ou privado. De maneira sucinta, a palavra “governança” é compreendida como a administração de recursos humanos e financeiros para consecução dos objetivos pretendidos por um órgão. “Transparência”, por sua vez, é tornar de conhecimento público as ações governamentais e *accountability* – cuja tradução incorreria em perda do sentido – em português, assume o significado de que toda e qualquer ação realizada por um órgão tem que ser responsabilizada; o que significa que quem está à frente das tomadas de decisões deve reconhecer as consequências dessa decisão. É com esse enfoque que os governos de diferentes países têm adotado iniciativas de dados governamentais abertos (*open government data*).

Nas referências nacionais, encontram-se duas traduções para *open government data*: 1) dados governamentais abertos; e 2) dados abertos governamentais. É oportuno lançar três questionamentos para reflexão: há diferença nas expressões “dados abertos governamentais” e “dados governamentais abertos”? A expressão em inglês *open government data* permite essa tradução de duas diferentes maneiras? Qual seria o prejuízo entre uma escolha e outra?

Ao se considerar cada termo isoladamente – dados + abertos + governamentais –, a ordem das palavras na expressão não causa distorção de sentido. Para proceder a essa averiguação, o substantivo “dados” recebe, aqui, os adjetivos “abertos” e “governamentais”, entendendo-se que:

- a) dados = conjunto de símbolos; estado bruto; sem valor agregado;
- b) abertos = permitido reuso e redistribuição por qualquer pessoa; e
- c) governamentais = proveniente da coisa pública.

Na expressão “dados governamentais abertos”, tem-se: conjunto de símbolos provenientes da coisa pública dos quais é permitido reuso e redistribuição por qualquer

pessoa. Por outro lado, “dados abertos governamentais” é entendido como conjunto de símbolos dos quais é permitido reuso e redistribuição por qualquer pessoa, sendo proveniente da coisa pública.

O que se percebe é que não há prejuízo conceitual quanto à adoção de uma expressão ou outra, porém, considerou-se “dados governamentais abertos” a tradução mais adequada para fins desta exposição.

A discussão feita na seção anterior sobre dados abertos apresenta informações para a compreensão da noção⁴⁵ de dados governamentais abertos. De acordo com material de divulgação do *World Wide Web Consortium (W3C)*,

dados abertos governamentais são a publicação e disseminação das *informações* do setor público na Web, compartilhadas *em formato bruto e aberto*, compreensíveis logicamente, de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade. (WORLD WIDE WEB CONSORTIUM, 2012, grifo nosso).

No ano de 2007, na cidade de Sebastopol, na Califórnia, Estados Unidos, um grupo internacional de ativistas para governo aberto elencou os oito princípios a serem adotados para dados governamentais abertos, a saber: completos, primários, atuais, acessíveis, compreensíveis por máquina, não discriminatórios, não proprietários e licenças livres⁴⁶. De acordo com o *Manual dos dados abertos: governo*, em sua tradução para o português, cada princípio tem a seguinte definição:

Completos: Todos os dados públicos devem ser disponibilizados. Dado público é aquele que não está sujeito a restrições de privacidade, segurança ou outros privilégios.

Primários: São apresentados tal como colhidos na fonte, com o maior nível possível de granularidade, sem agregação ou modificação (por exemplo, um gráfico não é fornecido aberto, mas os dados utilizados para construir a planilha que deu origem a ele podem ser abertos).

Atuais: Devem ser publicados o mais rápido possível para preservar seu valor. Em geral, têm periodicidade: quanto mais recentes e atuais, mais úteis para seus usuários.

Acessíveis: São disponibilizados para a maior quantidade possível de pessoas, atendendo, assim, aos mais diferentes propósitos.

Compreensíveis por máquina: Devem ser estruturados de modo razoável, possibilitando que sejam processados automaticamente (por exemplo, uma tabela em PDF é muito bem compreendida por pessoas, mas para um computador é apenas uma imagem; uma tabela em formato estruturado, como CSV ou XML, é processada mais facilmente por *softwares* e sistemas).

Não discriminatórios: Devem estar disponíveis para qualquer pessoa, sem necessidade de cadastro ou qualquer outro procedimento que impeça o acesso.

45 Neste trabalho compreende-se “dados governamentais abertos” como uma noção (cf. nota 20).

46 Disponível em: <https://public.resource.org/8_principles.html>.

Não proprietários: Nenhuma entidade ou organização deve ter controle exclusivo sobre dados disponibilizados.

Livres de licenças: Não devem estar submetidos a *copyrights*, patentes, marcas registradas ou regulações de segredo industrial. (WORLD WIDE WEB CONSORTIUM, 2011, grifos do autor).

A partir da definição desses princípios para dados governamentais abertos, entende-se que ao cidadão seria facilitado o acesso à informação pública.

O desejo dos membros de uma sociedade em ter conhecimento de e participação em toda e qualquer ação promovida por seus governantes baseia-se na satisfação de que suas contribuições tributárias estão sendo bem empregadas. Por outro lado, os governantes – em sua retórica – prezam pela participação social nas discussões e decisões governamentais e, por sua vez, utilizam-se de instrumentos para a participação social, tais como as consultas públicas realizadas de tempos em tempos pelos diferentes órgãos públicos e a publicação de conjuntos de dados governamentais.

4.2 A PARTICIPAÇÃO CIDADÃ NAS DECISÕES GOVERNAMENTAIS

A participação cidadã nas tomadas de decisão a respeito de políticas públicas para o bem-estar social é relevante, pois as deliberações serão feitas em conjunto. Com o uso de recursos computacionais, a interação cidadão-governo torna-se dinâmica e efetiva. Segundo Calderón e Lorenzo,

as práxis do *Open Data* pretendem aumentar [o volume de] a informação que colocamos em âmbito público para incrementar o empoderamento, ou seja, a capacidade de influência das pessoas e coletivos organizados baseados em interesses legítimos nas decisões de governo⁴⁷. (CALDERÓN; LORENZO, 2010, p. 110-111, tradução nossa).

A decisão dos governos de disponibilizarem os dados armazenados em seus bancos de dados vem ao encontro da democracia a ser exercida num país. No entanto, quem estaria fazendo uso desses dados governamentais?

Existem algumas iniciativas, nacionais e internacionais, visando ao uso dos dados governamentais abertos, de forma a contribuir para a melhoria das sociedades nas quais estão atuando. No que tange aos governos, são lançados concursos para desenvolvimento de

⁴⁷ Las praxis de *Open Data* pretenden [...] aumentar la información que ponemos en el ámbito de lo público, para incrementar el empoderamiento, o lo que es lo mismo, la capacidad de influencia de las personas y los colectivos organizados en base a legítimos intereses, en las decisiones del gobierno.

aplicativos; à sociedade civil, há organizações não governamentais, como a Transparência Hacker; às universidades, o estímulo à produção de conhecimento para reuso de dados. Outra iniciativa é a realização de maratonas para *hackers* (*hackaton*), das quais participam representantes de governos, universidades e sociedade civil, para desenvolvimento de aplicativos.

É com esse cenário que os setores públicos – e também privados – vêm participando das iniciativas de dados abertos e do seu reuso por parte da sociedade, com o objetivo de oferecer aos cidadãos participação nas ações governamentais, assim como transparência e *accountability* em suas atividades. Dessa maneira, busca-se um estado democrático de direito, como lembra Alberch y Fulgueras:

Até há bem pouco tempo, associar as palavras democracia e arquivo poderia parecer uma verdadeira heresia. Na atualidade, a demonstrada utilidade dos documentos de arquivo no fortalecimento do direito do cidadão à informação ou o inevitável recurso aos documentos em processos relacionados com direitos humanos e com as restituições morais e econômicas a pessoas ou instituições submetidos a diferentes níveis de repressão sob regimes ditatoriais⁴⁸. (ALBERCH Y FULGUERAS, [19--], tradução nossa).

De acordo com Brochu (2006), a conquista de vantagens e garantias é alcançada com a atuação de arquivistas à frente da gestão de arquivos tidos como fontes de informações estratégicas. O autor destaca também o trabalho do arquivista com os documentos digitais e o quanto esse profissional é confrontado com as necessidades dos usuários que, por sua vez, estão cada vez mais adeptos de novas tecnologias computacionais. O arquivista precisa estar preparado para definir e implantar novos métodos, processos e ferramentas para atendimento a esses usuários.

O uso dos arquivos – corrente, intermediário ou permanente – por uma sociedade possibilita que ela tenha uma fonte primária a respeito dos seus direitos individuais e/ou coletivos sobre saúde, educação, transporte, meio ambiente, desenvolvimento urbano, contas públicas, dentre outros temas de interesse público. Por isso, os

[...] arquivistas devem transformá-lo [o arquivo], ou encontrar novos significados ou propósitos que sejam relevantes ao máximo para as demandas atuais. Este crescimento na utilidade dos arquivos, [...] fará com que se consiga o apoio que assegurará que o significado original e os documentos sobrevivam (ou sejam

48 Hasta hace unos pocos años asociar las palabras democracia y archivo hubiera podido parecer una auténtica herejía. En la actualidad, la demostrada utilidad de los documentos de archivo en la potenciación del derecho ciudadano a la información o el inevitable recurso a los documentos en contenciosos relacionados con los derechos humanos y con las restituciones morales y económicas a personas o instituciones sometidas a distintos grados de represión bajo regímenes dictatoriales.

preservados). Se os arquivos devem permanecer viáveis e relevantes, precisam constantemente ser abertos e reabertos, renovados e redescobertos para e por novos significados, usos e usuários. (NESMITH, 2010, p. 20).

A reabertura, a redescoberta, os novos usos e usuários de um arquivo, defendidos por Nesmith, vem ao encontro do objetivo da iniciativa sobre dados governamentais abertos; com a disponibilização dos dados governamentais, tem-se “*accountability*, democracia, boa governança e cidadania”. (NESMITH, 2010, p. 19).

O fator humano também pode colaborar com a obtenção do objetivo pretendido com os dados governamentais abertos. São as estruturas mentais (VAZ, 2006) envolvidas na definição de políticas arquivísticas, de segurança e acesso à informação, de governo aberto, que definirão “como” e “onde” o cidadão conseguirá munir-se daquilo que lhe é de interesse. Mas, para isso, Vaz alerta que

Em relação à pressão por participação cidadã, transparência e controle social, é flagrante a necessidade de avançar nas iniciativas de capacitação que fortaleçam a abertura dos agentes públicos para uma real valorização dessa dimensão como afirmação da cidadania e como parte essencial de nossa vida democrática. (VAZ, 2006, p. 7).

O reúso dos dados governamentais abertos traz o empoderamento dos indivíduos de uma sociedade no exercício de sua cidadania, assim como o uso dos arquivos, sejam eles digitais ou não digitais. Para tanto, instrumentos legais são criados por legisladores nas diferentes esferas governamentais, como o projeto de lei PL 7804/2014⁴⁹, por exemplo, que institui a Lei de Dados Abertos, visando “à disponibilização pública de todos os dados primários produzidos, coletados ou armazenados, da forma mais ampla possível, através da rede mundial de computadores (Internet), ressalvados os casos de dados sigilosos ou informações pessoais”. (BRASIL, 2014a).

O debate e reflexões relacionados ao “movimento aberto” – o qual expande-se para acesso aberto, governo aberto, arquivo aberto, repositório aberto, dado aberto, dado governamental aberto, conhecimento aberto e outras iniciativas vindouras – permeiam diferentes áreas do conhecimento. Buscou-se, assim, apresentar elementos para fomentar reflexões no âmbito da Arquivologia, sem a pretensão de esgotar o assunto.

49 O projeto está aguardando parecer do relator na Comissão de Ciência e Tecnologia, Comunicação e Informática (CCTCI). Cf. “Projetos de lei e proposições”, em: <<http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=620193>>.

5 O PROJETO “DADOS ABERTOS GOVERNAMENTAIS”, DO PODER EXECUTIVO FEDERAL BRASILEIRO

O projeto “Dados Abertos Governamentais” é um dos projetos que integram o Programa de Governo Eletrônico brasileiro (GOV.BR) e consiste no estabelecimento de planos para publicação, em sítios na Internet, dos dados produzidos pelos órgãos do Poder Executivo federal, para serem reutilizados, principalmente, pela sociedade.

O objetivo é dar transparência às ações governamentais, especialmente quanto aos gastos públicos e serviços prestados aos cidadãos, possibilitando ampliar a participação social na administração pública do país. Para compreensão do projeto, são necessários alguns esclarecimentos a respeito de como o Brasil, representado pelo Poder Executivo federal, ingressou nas temáticas “Governo Aberto” e “Dados governamentais abertos”, o que é a Infraestrutura Nacional de Dados Abertos (INDA), e quem são os responsáveis pelo Projeto “Dados Abertos Governamentais”.

A Lei Complementar nº 101, de 4 de maio de 2000, “estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências” (BRASIL, 2000). No capítulo IX, seção I, orienta quanto à transparência da gestão fiscal da União, dos estados, do distrito federal e dos municípios, sendo compreendidos os Poderes Executivo, Legislativo, Judiciário e Ministério Público. De acordo com Vieira (2012, p.137) essa lei complementar é “um dos grandes marcos” para o processo de transparência ativa e acesso à informação pública.

Nesse mote da transparência das ações governamentais e acesso à informação pública, o Poder Executivo federal tomou outras iniciativas, dentre elas a disponibilização do Portal da Transparência, destacado nacional e internacionalmente (VIEIRA, 2012, p. 146). Devido a isso, o Brasil foi convidado a copresidir – 2011-2012 – a iniciativa *Open Government Partnership* (OGP); em português, Parceria para Governo Aberto (Figura 8).

Figura 8 – Linha do tempo de algumas ações do GOV.BR.



Fonte: elaboração própria.

A Parceria para Governo Aberto é uma iniciativa internacional, lançada em 2011, de adesão voluntária, que busca o compromisso dos governos quanto à transparência, combate à corrupção e prestação de contas aos cidadãos. Para o seu comitê deliberativo, conta com a participação de representantes de governos e de organizações da sociedade civil. Os governos integrantes da Parceria devem estar de acordo com os princípios por ela definidos para obtenção do seu objetivo, que é

assegurar compromissos concretos dos governos para promover a transparência, dar poder aos cidadãos, combater corrupção e adquirir novas tecnologias para fortalecer a governança. Com o espírito de colaboração múltipla, a OGP é supervisionada por um Comitê Gestor que inclui representantes de governos e de organizações da sociedade civil⁵⁰. (OPEN GOVERNMENT PARTNERSHIP, 2013, tradução nossa).

Um dos comprometimentos que os países-membros da Parceria assumem é a definição de um plano de ação de dois anos. Assim, o Decreto s/nº, de 15 de setembro de 2011, instituiu o Plano de Ação Nacional sobre Governo Aberto, no âmbito federal brasileiro. É pertinente apontar que esse foi o 1º plano 2011-2013 e o 2º plano 2013-2015 está em andamento.

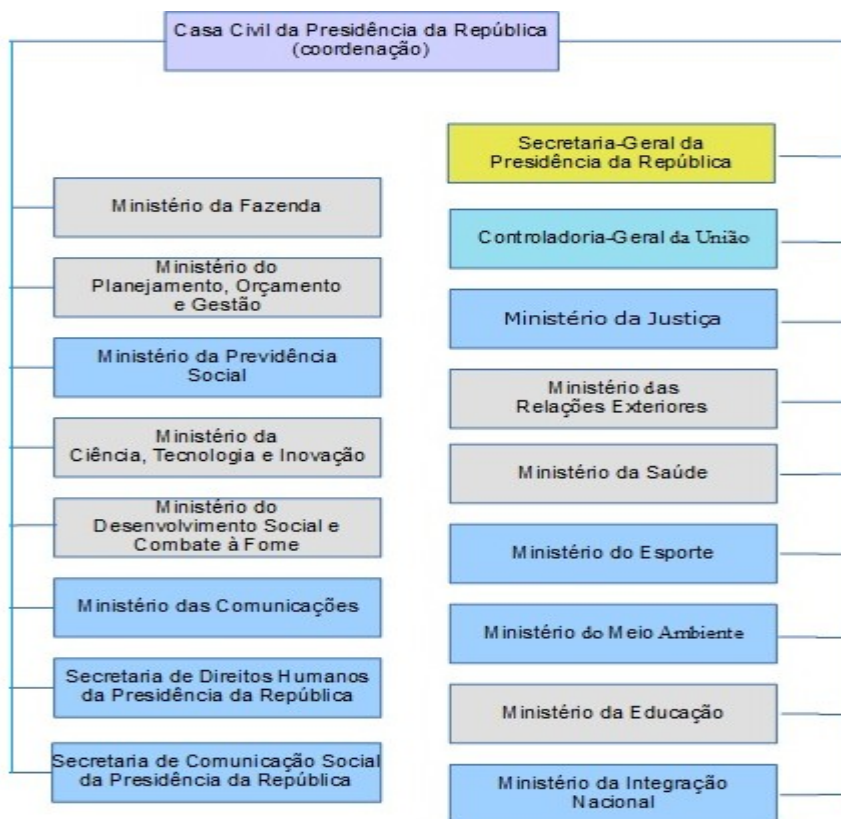
⁵⁰ to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. In the spirit of multi-stakeholder collaboration, OGP is overseen by a Steering Committee including representatives of governments and civil society organizations.

O art. 1º desse Decreto define que o Plano de Ação tem o objetivo de “promover ações e medidas que visem ao incremento da transparência e do acesso à informação pública, à melhoria na prestação de serviços públicos e ao fortalecimento da integridade pública” (BRASIL, 2011a). Cabe destacar, portanto, algumas das diretrizes para a sua realização:

- I – aumento da disponibilidade de informações acerca de atividades governamentais, incluindo dados sobre gastos e desempenho das ações e programas;
- II – fomento à participação social nos processos decisórios;
- III – estímulo ao uso de novas tecnologias na gestão e prestação de serviços públicos, que devem fomentar a inovação, fortalecer a governança pública e aumentar a transparência e a participação social; e
- IV – incremento dos processos de transparência e de acesso a informações públicas, e da utilização de tecnologias que apoiem esses processos. (BRASIL, 2011a).

No art. 3º do Decreto s/nº, com a finalidade de coordenar a elaboração, a implantação, o monitoramento e a avaliação do Plano de Ação, foi instituído o Comitê Interministerial Governo Aberto (CIGA) e o Grupo Executivo (GE), o qual é responsável por elaborar a proposta, coordenar a implementação e a execução desse Plano. Os órgãos integrantes do CIGA são indicados na figura 9.

Figura 9 – Membros do CIGA, da Parceria para Governo Aberto



Fonte: elaboração própria, com base no Decreto s/nº, de 15 de setembro de 2011.

Os membros que compõem o GE são apresentados na figura 10. Cabe ressaltar que, no sítio da Controladoria-Geral da União (CGU), na seção “Governos Abertos”⁵¹, consta o Ministério da Justiça como integrante do GE do CIGA, diferentemente do art. 7º, do Decreto s/nº, de 15 de setembro de 2011.

Figura 10 – Membros do GE, do CIGA.



Fonte: elaboração própria com base no Decreto s/nº, de 15 de setembro de 2011.

Pelo Decreto s/nº, de 15 de setembro de 2011, é atribuído ao GE/CIGA a realização de convites aos demais órgãos do CIGA e a constituição de grupos de trabalho temáticos.

Nas discussões sobre a elaboração do Plano de Ação nacional sobre Governos Abertos, estão previstas consultas para a manifestação de movimentos sociais, instituições científicas, entidades e agentes interessados, assim como a participação de especialistas nos trabalhos desenvolvidos pelo CIGA e pelo GE/CIGA. (BRASIL, 2011).

A mobilização de 18 ministérios dos 37 existentes na estrutura do Poder Executivo federal (INDOLFO, 2013) – o que representa 48,64% de órgãos subordinados à Presidência da República – é um indício da relevância que o governo federal tem atribuído ao

51 Disponível em: <<http://www.cgu.gov.br/governoaberto/ciga.asp>>.

cumprimento da LAI e à iniciativa de governo aberto.

5.1 LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011, A “LEI DE ACESSO À INFORMAÇÃO PÚBLICA”

A Lei nº 12.527, de 18 de novembro de 2011, tem sido comumente chamada de Lei de Acesso à Informações públicas (LAI). Em seu art. 1º, consta que ela “dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal” (BRASIL, 2011b). A LAI foi regulamentada pelo Decreto nº 7.724, de 16 de maio de 2012.

Ao entrar em vigor, a LAI efetivou o desafio do governo federal de promover a cidadania dentre os membros da sociedade brasileira por meio da disponibilização das informações públicas. Com isso, é possível um maior controle sobre as ações governamentais e facilidade na obtenção daquilo que é de direito do cidadão: serviço público de qualidade. Caberá também ao próprio cidadão estar atento a mais esse instrumento para a garantia de seus direitos.

O direito à prestação de um serviço público de qualidade também é conquistado através da gestão arquivística de documentos. Porém, a LAI não faz menção a termos conceitualmente trabalhados pela Arquivologia, conforme aponta Bittencourt (2014):

Nos termos que compõem a base conceitual da lei que regula o acesso à informação não constam os termos arquivo e gestão de documentos [...] O termo documento aparece treze vezes no texto da lei nº 12.527/2011. Porém, em nenhum trecho aparece relacionado à gestão ou arquivo. (BITTENCOURT, 2014, p. 63-64).

Essa ausência identificada por Bittencourt indica uma lacuna na cooperação e interação entre órgãos do Poder Executivo federal com a instituição arquivística federal, especialmente frente aos desafios que as TICs proporcionam à gestão de documentos arquivísticos.

Uma instituição arquivística tem como missão preservar e dar acesso ao patrimônio documental da administração pública, qualquer que seja sua esfera de atuação: federal, estadual, municipal ou do distrito federal. Isso somente é possível porque é o “órgão responsável pelo recolhimento, preservação e acesso dos documentos gerados” por tais

administrações públicas (FONSECA, 2007, p. 40). Cabe ressaltar que essa visão de instituição arquivística destinada apenas à preservação de um patrimônio, de um arquivo histórico, foi ultrapassada após a 2ª Guerra Mundial e o crescimento exponencial da produção de documentos públicos, exigindo a racionalização dessa produção.

De acordo com Jardim (2012),

as instituições arquivísticas passam a ter como objeto de intervenção todo o ciclo documental e não mais apenas os arquivos históricos. Devem atuar, também, mediante políticas arquivísticas, na gestão dos documentos correntes e intermediários produzidos pela administração pública. (JARDIM, 2012, p. 397).

No Brasil, o Arquivo Nacional (AN) é a instituição arquivística responsável pela gestão e preservação dos arquivos do Poder Executivo federal, por dar acesso aos conjuntos documentais sob sua guarda, assim como pelo acompanhamento e implementação da política nacional de arquivos (BRASIL, 1991). É também o órgão central do Sistema de Gestão de Documentos de Arquivos (SIGA), da administração pública federal, conforme item I, do art. 3º, do Decreto nº 4.915, de 12 de dezembro de 2003.

Como órgão central do SIGA, o Arquivo Nacional deve, dentre suas responsabilidades, “promover e manter intercâmbio de cooperação técnica com instituições e sistemas afins, nacionais e internacionais” (BRASIL, 2003, item VI). Com isso, pode-se depreender a importância de o AN integrar determinados comitês e grupos executivos ou de trabalho cujas ações repercutam na gestão, preservação e acesso a documentos arquivísticos. No entanto, cabe questionar se o Arquivo Nacional se fez (ou faz) presente nas discussões sobre governo e dados abertos.

A expectativa é de um “sim” como resposta à questão acima, porém, no sítio do AN na Internet não há informações ou documentos disponibilizados que indiquem a sua participação nas discussões sobre as iniciativas de governo aberto ou de dados abertos. Isso causa estranheza, já que é um órgão subordinado ao Ministério da Justiça que, por sua vez, integra o Comitê Interministerial Governo Aberto (CIGA).

Tais iniciativas, de governo aberto ou de dados abertos, têm tomado corpo devido ao uso de TICs pelos governos no desenvolvimento de suas funções e atividades. Essas tecnologias facilitam o acompanhamento das ações governamentais – as quais são registradas em documentos arquivísticos digitais – pelos cidadãos, qualquer que seja seu país de origem.

Com relação aos documentos arquivísticos digitais produzidos e recebidos pelos órgãos do Poder Executivo federal, eles têm sido objeto de estudo do AN quanto à gestão e preservação. Os produtos que abarcam essas duas funções arquivísticas são a apostila do curso de capacitação para os integrantes do Sistema de Gestão de Documentos de Arquivo (SIGA)⁵² e a política de preservação digital (AN Digital)⁵³. A primeira dedica um capítulo aos documentos arquivísticos digitais, enquanto o AN Digital contempla as deliberações e os projetos para a implantação de um repositório digital.

5.2 MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO (MP)

No Brasil, é o Poder Executivo federal que atua na implantação dos programas governamentais e na prestação de serviços aos cidadãos brasileiros. Dentre os programas federais elaborados a partir das atividades dos órgãos públicos federais, está o Programa de Governo Eletrônico brasileiro (GOV.BR), cujo responsável é o Ministério do Planejamento, Orçamento e Gestão (MP).

O MP é um órgão da administração pública federal direta, cuja estrutura regimental está aprovada pelo Decreto nº 8.189, de 21 de janeiro de 2014. Em seu Anexo 1, art. 1º, estão previstas 11 competências para esse Ministério:

- I- participação na formulação do planejamento estratégico nacional;
- II- avaliação dos impactos socioeconômicos das políticas e programas do Governo federal e elaboração de estudos para a reformulação de políticas;
- III- realização de estudos e pesquisas para acompanhamento da conjuntura socioeconômica e gestão dos sistemas cartográficos e estatísticos nacionais;
- IV- elaboração, acompanhamento e avaliação das leis de iniciativa do Poder Executivo federal previstas no art. 165 da Constituição;
- V- viabilização de novas fontes de recursos para os planos de governo;
- VI- coordenação da gestão de parcerias público-privadas;
- VII- formulação de diretrizes, coordenação das negociações, acompanhamento e avaliação dos financiamentos externos de projetos públicos com organismos multilaterais e agências governamentais;
- VIII- coordenação e gestão dos sistemas de planejamento e orçamento federal, de pessoal civil, de administração de recursos da informação e informática e de serviços gerais, bem como das ações de organização e modernização administrativa do Governo federal;

52 ARQUIVO NACIONAL (BRASIL). Gestão de documentos: curso de capacitação para os integrantes do Sistema de Gestão de Documentos de Arquivo – SIGA, da administração pública federal. Rio de Janeiro: Arquivo Nacional, 2011. (Publicações técnicas nº 55). CDD 025.171 4. Documento impresso.

53 Disponível em:

<http://www.siga.arquivonacional.gov.br/media/andigital/and_politica_preservao_digital_v1.pdf>.

IX- formulação de diretrizes, coordenação e definição de critérios de governança corporativa das empresas estatais federais;
 X- administração patrimonial; e
 XI- política e diretrizes para modernização da administração pública federal.
 (BRASIL, 2014b).

Dentre suas competências, a definição de “políticas e diretrizes para modernização da administração pública federal” e a “[...] administração de recursos da informação e informática [...]” (BRASIL, 2014b) centralizam o MP nas ações de administração e coordenação da infraestrutura de recursos da informação para dar acesso à informação pública federal. Por isso, além do GOV.BR, o Ministério assume ações quanto a “compras públicas, transferências voluntárias da União, *software* público, administração dos recursos de TI, infovia e dados abertos” por meio da Secretaria de Logística e Tecnologia da Informação (SLTI) (BRASIL, 2015c). Com referência aos dados abertos, no final do ano de 2013, o MP publicou o “Plano de Dados Abertos para o Ministério do Planejamento” (PDA), para o biênio 2014-2015⁵⁴.

Quando Tesoro (apud JARDIM, 2005) apresenta as diferentes expressões em espanhol para designar governo eletrônico e a que ele se destina, compreende-se que

as expressões “Gobierno electrónico”, “Gobierno digital”, “Gobierno en línea”, “Administración electrónica” e “Estado digital”, como sinônimos da anglo-saxônica “electronic government” (“e-government”, “e-gov”), pretendem aludir a um “novo” paradigma tecnológico associado à prestação telemática de serviços por parte dos poderes públicos para: a) prover mais e melhores prestações [de contas] aos cidadãos, b) prover pontos de acesso unificados e simples para satisfazer múltiplas necessidades informacionais e de serviços, c) adicionar produtividade, qualidade e valor aos serviços, d) prover atendimento às demandas de cada usuário, e) atender aos trâmites, às consultas, às reclamações e às sugestões “em rede”, f) aumentar a participação cidadã, g) aumentar a qualidade e reduzir o custo das transações no âmbito de governo, h) reduzir o custo (financeiro, tempo, aborrecimento e disposição) de deslocamento do público, i) aumentar a transparência da gestão pública⁵⁵. (TESORO, 2004 apud JARDIM, 2005).

54 Disponível em:

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/governo_aberto/plano_dados_abertos.pdf>.

55 las expresiones “Gobierno electrónico”, “Gobierno digital”, “Gobierno en línea”, “Administración electrónica” y “Estado digital”, como equivalentes a la anglosajona “electronic government” (“e-government”, “e-gov”), pretenden aludir a un “nuevo” paradigma tecnológico asociado a la prestación telemática de servicios por parte de los poderes públicos, para: a) proveer más y mejores prestaciones a los ciudadanos, b) proveer puntos de acceso unificados y sencillos para satisfacer múltiples necesidades informativas y de servicios, c) adicionar productividad, calidad y valor a los servicios, d) proveer atención “a la medida” de los requerimientos de cada usuario, e) resolver trámites, consultas, reclamos y sugerencias “en línea”, f) aumentar la participación ciudadana, g) aumentar la calidad y reducir el costo de las transacciones dentro del Estado, h) reducir el costo (en dinero, tiempo, molestia y carga anímica) de tramitaciones del público, i) aumentar la transparencia de la gestión pública.

No Brasil, a política⁵⁶ de governo eletrônico tem como pilares a “participação cidadã; a melhoria no gerenciamento interno do Estado e a integração com parceiros e fornecedores” (BRASIL, 2013a). A partir dessa base, visa-se à democratização do acesso à informação pública e à eficiência na prestação de serviços públicos à sociedade brasileira (BRASIL, 2013b).

Tais serviços prestados aos cidadãos geralmente resultam em documentos (para fins desta explanação, o digital) que, para a administração pública, precisam ser reconhecidos como arquivísticos, pois registram ou apoiam as ações desenvolvidas pelo órgão público, sendo, então, evidência de suas atividades.

Outros dois projetos desenvolvidos pelo MP que abrangem o documento arquivístico digital em sua produção e gerenciamento são o Processo Eletrônico Nacional (PEN) e o Sistema Protocolo Integrado (PI), sendo este último regulado pela Portaria Interministerial nº 2.320, de 30 de dezembro de 2014.

O projeto do PEN tem como objetivo construir uma “infraestrutura pública de processo administrativo eletrônico” que abarca diferentes órgãos e entidades das esferas governamentais: federal, estadual e municipal. Essa infraestrutura está composta por uma “solução de processo administrativo eletrônico (Serviço Eletrônico de Informação – SEI)” e “serviços centralizados de processo eletrônico (barramento)”. (BRASIL, 2015d).

Quanto ao Sistema de Protocolo Integrado – projeto em parceria com o Arquivo Nacional –, ele tem como público-alvo a sociedade, ou seja, consiste em “consolidar em uma única base de dados informações de protocolo oriundas dos diversos órgãos e entidades da Administração Pública Federal” (BRASIL, 2015e). Assim, é facultada a qualquer pessoa interessada a consulta sobre a situação do trâmite de documentos, avulsos ou processos, em um único endereço da Internet⁵⁷.

O Poder Executivo federal brasileiro tem-se mobilizado para implantar programas e projetos cuja finalidade é disponibilizar informações e dados sobre as ações governamentais – em cumprimento à Lei de Acesso a Informações públicas (LAI) –, os quais se encontram registrados em documentos arquivísticos digitais que, por sua vez, precisam ser essencialmente orgânicos, autênticos, confiáveis e acessíveis, assim como os repositórios

56 “A política é um conjunto de princípios e estratégias que orientam um curso de ação para alcançar um objetivo particular. As políticas podem ser desenvolvidas em nível orgânico ou institucional (micro) ou em nível nacional, regional ou internacional (macro) [...]”. (MONTVILOFF, 1990, [p. 7], tradução nossa).

57 Protocolo Integrado Beta, do Governo federal. Disponível em: <<https://protocolointegrado.gov.br>>.

digitais onde tais documentos são armazenados.

Dentre as diretrizes recomendadas aos órgãos do Poder Executivo federal para publicação de dados governamentais, em português, destacam-se aqui o “Manual dos dados abertos: governo” e o “Manual para a elaboração de plano de dados abertos”. Já os principais atos normativos são o Decreto s/nº, de 15 de setembro de 2011, que instituiu o Plano de Ação Nacional sobre Governo Aberto, e a Instrução Normativa nº 4, de 13 de abril de 2012, da SLTI que institui a Infraestrutura Nacional de Dados Abertos (INDA).

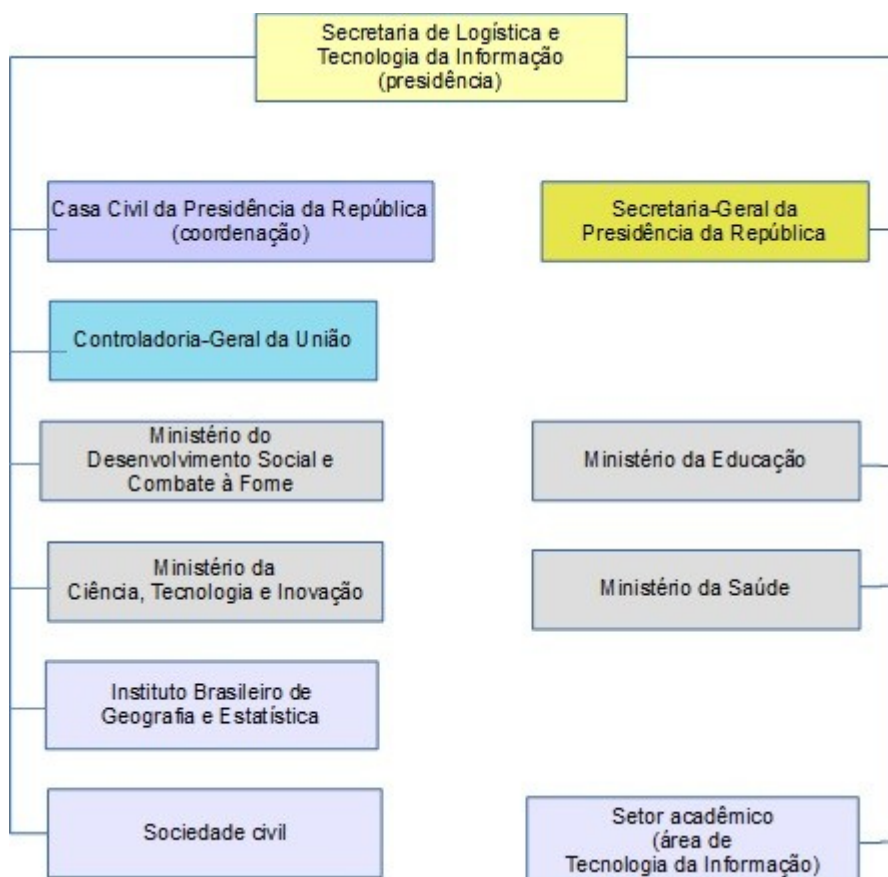
5.3 INFRAESTRUTURA NACIONAL DE DADOS ABERTOS (INDA)

O Plano de Ação Nacional sobre Governo Aberto, produto do comprometimento que os países-membros da Parceria para Governo Aberto assumem entre si, estabelece a implantação da Infraestrutura Nacional de Dados Abertos (INDA) pelo governo brasileiro.

A INDA foi instituída pela Instrução Normativa nº 4, de 13 de abril de 2012. No item I, do art. 1º, consta que um dos objetivos da INDA é *definir, estruturar e coordenar a política de dados abertos*. Além disso, em seu art. 4º, está definido que “A INDA disponibilizará o Portal Brasileiro de Dados Abertos, que será o sítio eletrônico de referência para a busca e o acesso aos dados públicos, seus metadados, informações, aplicativos e serviços relacionados”. (BRASIL, 2012).

A responsabilidade pela sua gestão está a cargo de representantes do Poder Executivo federal, do setor acadêmico – na especialidade de tecnologia da informação (TI) – e da sociedade civil organizada, que compõem o Comitê Gestor da INDA. Na figura 11, são nominados os membros do Comitê.

Figura 11 – Membros do Comitê Gestor da INDA.



Fonte: elaboração própria, com base na Instrução Normativa nº 4, da SLTI, de 13 de abril de 2012.

No que se refere ao Comitê, algumas observações podem ser pontuadas, com base na IN nº 4, da SLTI:

- a) O Ministério do Planejamento, Orçamento e Gestão é representado no Comitê Gestor pela Secretaria de Logística e Tecnologia da Informação (SLTI);
- b) A indicação do representante da sociedade civil é feita pela Secretaria Nacional de Articulação Social, da Secretaria-Geral da Presidência da República;
- c) A indicação do representante do setor acadêmico é feita pelo Ministério da Ciência, Tecnologia e Inovação; e
- d) Os órgãos e entidades serão formalmente convidados a indicar os seus respectivos representantes, titular e suplente, a serem nomeados pelo Secretário de Logística e Tecnologia da Informação, do Ministério do Planejamento, Orçamento e Gestão.

Uma vez pontuadas as informações acima, no quadro 9 estão representados os órgãos federais e entidades civis brasileiras a fim de se identificar em quais comitês ou grupos responsáveis pela definição, implantação e condução das ações destinadas às iniciativas de governo aberto e de dados abertos eles se encontram.

Quadro 9 – Membros do CIGA e do GE, do Governo Aberto, e do Comitê Gestor da INDA.

Representante	Comitê Interministerial do Governo Aberto (CIGA)	Grupo Executivo do Governo Aberto (GE)	Comitê Gestor da INDA
Casa Civil da Presidência da República*	X	X	X
Controladoria-Geral da União*	X	X	X
Instituto Brasileiro de Geografia e Estatística			X
Ministério da Ciência, Tecnologia e Inovação	X		X
Ministério da Educação	X		X
Ministério da Fazenda	X	X	
Ministério da Integração Nacional	X		
Ministério da Justiça	X		
Ministério da Previdência Social	X		
Ministério da Saúde	X		X
Ministério das Comunicações	X		
Ministério das Relações Exteriores	X	X	
Ministério do Desenvolvimento Social e Combate à Fome	X		X
Ministério do Esporte	X		

Ministério do Meio Ambiente	X		
Ministério do Planejamento, Orçamento e Gestão	X	X	
Secretaria de Comunicação Social da Presidência da República	X		
Secretaria de Direitos Humanos da Presidência da República	X		
Secretaria de Logística e Tecnologia da Informação*			X
Secretaria-Geral da Presidência da República	X	X	X
Setor acadêmico (TI)			X
Sociedade civil			X
TOTAL =	18	6	11
*Órgão que tem como função coordenação ou presidência.			

Fonte: elaboração própria.

A partir da observação do quadro 9, a Casa Civil da Presidência da República, a Controladoria-Geral da União e a Secretaria-Geral da Presidência da República são os órgãos que centralizam as ações quanto às iniciativas para governo aberto e dados abertos, no âmbito do Poder Executivo federal.

5.4 “PORTAL BRASILEIRO DE DADOS ABERTOS”

O “Portal Brasileiro de Dados Abertos” <<http://dados.gov.br/>> tem como objetivo ser a principal referência para publicação e reuso dos dados governamentais de órgãos do Poder Executivo federal. O Portal foi desenvolvido para cumprimento do compromisso assumido pelo Brasil no 1º Plano de Ação estipulado pela Parceria para Governo Aberto.

A Secretaria de Logística e Tecnologia da Informação (SLTI), vinculada ao Ministério do Planejamento, Orçamento e Gestão (MP), foi a responsável pelo

desenvolvimento do Portal em conjunto com outros órgãos públicos, instituições não governamentais e representantes da sociedade civil. Trata-se de um modelo de atuação condizente com as propostas das iniciativas de governo e de dados abertos.

Para a publicação de *datasets* no Portal, é necessário o preenchimento de um formulário por algum servidor lotado em órgão público, já que “cada órgão é responsável pela catalogação no portal dados.gov.br daqueles dados que publica na Internet”. (BRASIL, 2014e).

No próprio Portal, encontra-se a seguinte explicação:

O Portal Brasileiro de Dados Abertos é a *ferramenta* disponibilizada pelo governo para que todos possam encontrar e utilizar os dados e as informações públicas. O portal preza pela simplicidade e organização para que você possa encontrar facilmente os dados e informações que precisa. O portal também tem o objetivo de promover a interlocução entre atores da sociedade e com o governo para pensar a melhor utilização dos dados em prol de uma sociedade melhor. (BRASIL, 2014e, grifo nosso).

Na explicação acima, chama atenção o fato de que no Portal é possível “encontrar e utilizar os dados e as informações públicas”. A definição para dados abertos apresentada no Portal inclui ainda o termo *informações*:

[...] Dados Abertos é (sic) a publicação e disseminação dos dados e informações públicas na Internet, organizados de tal maneira que permita sua reutilização em aplicativos digitais desenvolvidos pela sociedade. (BRASIL, 2014e).

Há uma diferença conceitual entre dado e informação, o que conduz a uma reflexão sobre a abordagem do Portal Brasileiro de Dados Abertos: por que no Portal são publicadas *informações* – como o conjunto de símbolos com significado para um receptor – e não apenas *dados* – como o conjunto de símbolos sem significado, por isso apropriado para reuso com outras finalidades? Uma possível resposta encontra-se na seção “Perguntas frequentes”:

Os dados catalogados no portal estão organizados utilizando as estruturas de conjuntos de dados e recursos. Os conjuntos de dados são as entidades principais retornadas a partir das buscas. Cada conjunto de dados possui uma descrição, um ou mais recursos, e uma série de outros metadados, como periodicidade de atualização e órgão responsável. *Cada recurso compreende uma fonte de dados. Pode ser uma planilha, um método de webservice ou um documento. A princípio cada recurso poderia ser catalogado como um conjunto de dados separado, porém é preferível que os recursos sejam agrupados, sempre que forem oriundos da mesma base de dados, ou possuírem metadados em comum, para facilitar a busca e o entendimento de seu conteúdo.* (BRASIL, 2014e, grifo nosso).

A publicação de informações no Portal Brasileiro de Dados Abertos é condizente se considerada a definição de que os *datasets* são conjuntos de dados estruturados, contextualizados, com significado para o seu receptor, ou seja, o consumidor de dados. Por outro lado, a publicação de dados em seu estado bruto, como um conjunto de símbolos com pouco ou sem significado para seu receptor, efetivar-se-ia caso ela fosse feita por outro meio que não os *datasets*, como a ferramenta *big data*, por exemplo, que consiste em ter-se um volumoso *corpus* de informação que permita conexões, as quais não eram antes identificadas numa amostra pouco volumosa. (MAYER-SCHÖNBERGER; CUKIER, 2013).

5.5 MINISTÉRIO DA JUSTIÇA (MJ)

A missão do Ministério da Justiça (MJ) é “garantir e promover a cidadania, a justiça e a segurança pública, por meio de uma ação conjunta entre o Estado e a sociedade” (BRASIL, 2014c). Se “ação conjunta” significar a participação da sociedade nas ações do MJ, então, a disponibilização de dados governamentais na Internet favorece o cumprimento da missão do Ministério.

O MJ é um dos órgãos do Poder Executivo federal que publica dados governamentais no Portal Brasileiro de Dados Abertos e realiza concursos para o desenvolvimento de aplicativos (*apps*) que possibilitam o acompanhamento das atividades desempenhadas pelo órgão, com destaque para defesa do consumidor e segurança nas estradas federais. É pertinente dizer ainda que o MJ também publica *datasets* em seu sítio⁵⁸ na Internet, sendo alguns deles cópias dos que estão no Portal. Mas, por que o Ministério teve tal iniciativa? Quais diretrizes ou atos normativos são seguidos?

O Brasil é um dos países integrantes da Parceria para Governo Aberto (OGP, na sigla em inglês) e dentre os seus compromissos está a elaboração de um plano de ação para governo aberto, em que se incluem a disponibilidade de informações e dados governamentais. O MJ é um dos órgãos que integram o Comitê Interministerial Governo Aberto (CIGA), conforme apresentado no início da seção 5.

A Portaria nº 1.378, de 20 de agosto de 2014, que “institui, no âmbito do Ministério da Justiça, a Política de Dados Abertos e Espaciais e aprova o Plano Institucional de Dados

⁵⁸ No sítio <www.justica.gov.br/dados-abertos> estão listados os doze temas sobre os quais são disponibilizados *datasets*.

Abertos e Espaciais” (BRASIL, 2014d), contempla, em seu artigo 1º, a instituição da Política e aprovação do Plano; já o artigo 5º dedica-se à finalidade do Plano Institucional elaborado pela Secretaria-Executiva do MJ.

Neste ato normativo, porém, não é explicitada a participação do Arquivo Nacional na sua elaboração, tampouco da Política de Dados Abertos e Espaciais e do Plano Institucional de Dados Abertos e Espaciais, omissão esta que chama atenção, já que o AN está subordinado ao MJ⁵⁹ e é a principal instituição arquivística brasileira.

Há um descompasso (ou desarticulação política?) entre as ações do MJ e a instituição arquivística federal sob sua subordinação. Essa situação é preocupante, na medida em que os documentos arquivísticos digitais também são fontes de dados governamentais. Esses dados estão sendo publicados para uso, reúso e redistribuição pelos consumidores de dados, tais como ativistas *hackers*⁶⁰, universitários (da área de computação e informática) e organizações não governamentais (ONGs).

A responsabilidade dos órgãos públicos é resguardar a autenticidade, a confiabilidade e a acurácia dos documentos arquivísticos digitais em todo seu ciclo de vida, além de garantir o seu acesso. Ressalte-se ainda que esses documentos são produzidos e mantidos em sistemas informatizados armazenados em bases de dados as quais, por sua vez, também são reconhecidas como documentos arquivísticos por registrarem e apoiarem as atividades desenvolvidas por esses órgãos públicos; base de dados estas que são fontes de *datasets* para publicação na Internet sob as premissas da iniciativa de dados governamentais abertos.

59 O Decreto nº 7.430, de 17 de janeiro de 2011 dispõe sobre a transferência do Arquivo Nacional (AN) e do Conselho Nacional de Arquivos (Conarq) da Casa Civil da Presidência da República para o Ministério da Justiça (MJ).

60 *Hacker* é uma pessoa que sente prazer em ter uma compreensão profunda dos processos internos de um sistema, de um computador ou de uma rede informática em específico. O termo é frequentemente colocado em um contexto pejorativo, em que *cracker* seria o nome correto. (RFC 1.392, IETF, tradução nossa). Disponível em: <<http://www.ietf.org/rfc/rfc1392.txt>>.

6 METODOLOGIA

Este trabalho de natureza aplicada, com abordagem qualitativa e objetivo exploratório, adota o método científico de estudo de caso.

O estudo de caso como método científico possibilitou a análise de um exemplar de determinado universo a ser investigado e, a partir dele, analisar a problemática levantada na pesquisa.

Na primeira etapa, para a seleção do *dataset* tomado como objeto da pesquisa, consultou-se, aqueles disponíveis na seção “Dados abertos”, no sítio do Ministério da Justiça (MJ) <<http://www.justica.gov.br/dados-abertos>>, a saber (BRASIL, 2015f):

1) Lista de cartórios do Brasil: informações sobre Cartórios do Brasil mantidas pelo Conselho Nacional de Justiça – CNJ;

2) Comunidades terapêuticas: informações sobre Comunidades terapêuticas, entidades privadas, sem fins lucrativos, que realizam acolhimento de pessoas com transtornos decorrentes do uso, abuso ou dependência de substâncias psicoativas;

3) Postos de Campanha do desarmamento: informações sobre os postos de entrega voluntária de armas;

4) Unidades dos Procons: informações sobre Procons mantidas pela Secretaria Nacional do Consumidor - Senacon/MJ;

5) Atendimento de consumidores nos Procons (Sindec): o Boletim Sindec é a publicação periódica da Senacon que consolida todos os atendimentos (demandas) registrados pelos Procons integrados ao Sindec, para permitir uma análise sobre os assuntos, problemas e fornecedores que mais levam os consumidores aos Procons;

6) Cadastro Nacional de reclamações fundamentadas (PROCONS – Sindec): o Cadastro Nacional de Reclamações Fundamentadas é o cadastro formado pelas Reclamações finalizadas pelos Procons integrados ao Sindec, no período de 12 meses. Apresenta anualmente a consolidação dos Cadastros publicados por estados e municípios de todas as regiões brasileiras;

7) Sistema BR-Brasil – boletins de ocorrências em Rodovias Federais: o Sistema BR-Brasil cataloga todos os boletins de ocorrência registrados após um acidente em uma rodovia federal. São várias as informações cadastradas, incluindo diversos detalhes sobre os

veículos, as pessoas envolvidas e as condições do local, resguardando-se os dados pessoais;

8) Sistema Multa – autuações de trânsito em Rodovias Federais: informações coletadas na ocasião de uma autuação de trânsito em rodovias federais são catalogadas, resguardando-se os dados pessoais, que permitem a identificação do infrator;

9) Postos e delegacias do Departamento de Polícia Rodoviária Federal (DPRF): a base consolida a localização e contato de todos os postos e delegacias do Departamento de Polícia Rodoviária Federal distribuídos pelo país;

10) Postos e delegacias do Departamento de Polícia Federal (DPF): a base consolida a localização e contato de todos os postos e delegacias do Departamento de Polícia Federal distribuídos pelo país;

11) Reclamações do Consumidor.gov.br: informações anonimizadas sobre as reclamações cadastradas no Consumidor.gov.br, sítio criado para solução alternativa de conflitos de consumo por meio da Internet, que permite a interlocução direta entre consumidores e empresas;

12) Recall – campanhas de chamamento: a base de dados se refere às informações fornecidas pela Secretaria Nacional do Consumidor sobre os *recalls* (forma pela qual um fornecedor vem a público informar que seu produto ou serviço apresenta riscos aos consumidores podendo ser recolhido) realizados em território brasileiro;

13) Lista de terceirizados do Ministério da Justiça: informações sobre mão de obra terceirizada contratada para trabalhar nos diferentes órgãos do MJ; e

14) Lista de estagiários do Ministério da Justiça: informações sobre mão de obra estagiária atuando nos órgãos do MJ.

A seguir, foram selecionadas as planilhas de cálculo no formato *.xls (*software* Microsoft Excel⁶¹), item 13 – Lista de terceirizados do Ministério da Justiça/formato XLS⁶², do período de janeiro a dezembro de 2013. Cada uma das planilhas é um *dataset* formado pelo MJ e cuja finalidade é divulgar dados acerca dos funcionários terceirizados contratados para trabalhar no Ministério.

61 O Ministério da Justiça publica os mesmos *datasets* também no formato .csv (comma separate value). A opção de analisar *datasets* no formato .xls é por não seguir um dos princípios dos dados governamentais abertos.

62 Disponível em: <<http://www.justica.gov.br/dados-abertos>>. Acesso em 11 de maio de 2015.

Na segunda etapa, para identificar quais qualidades seriam inerentes a um dado governamental aberto, foi realizada análise bibliográfica sobre gestão de dados e dados abertos. A pesquisa documental, por sua vez, foi adotada para auxiliar no conhecimento dos trâmites de formação do Comitê Interministerial para Governo Aberto (CIGA) e do seu Grupo Executivo (GE), da Infraestrutura Nacional de Dados Abertos (INDA) e da Comissão Permanente de Avaliação de Documentos do Ministério da Justiça (CPAD/MJ). As principais fontes consultadas foram marcos legais (portarias, decretos, instrução normativa).

Na terceira etapa, para atingir o objetivo geral desta pesquisa, de investigar se o *dataset* (a planilha de cálculo) se trata de um dado bruto ou de um documento arquivístico, adotou-se o *Modelo de análise diplomática* (ANEXO A) aplicado em estudos de caso no âmbito do Projeto InterPARES 2 (DIPLOMATIC ANALYSIS TEMPLATE, 2008, p. 422). A escolha desse modelo foi porque ele orienta como identificar as características de um documento arquivístico: forma fixa e conteúdo estável, contexto da produção, pessoas envolvidas, relação orgânica e participar de uma ação.

Para a coleta de dados, os instrumentos adotados foram a observação individual e assistemática dos *datasets* visando a identificar suas propriedades e campos da tabela, e o sistema eletrônico do Serviço de Informações ao Cidadão (e-SIC), do Poder Executivo federal.

Na observação individual e assistemática dos *datasets*, examinaram-se as propriedades de identidade e de integridade – elementos que compõem a autenticidade de documentos arquivísticos.

O e-SIC, do Poder Executivo federal, é a ferramenta que o cidadão utiliza para obter informações públicas. Dessa forma, as questões direcionadas à Coordenação Geral de Recursos Humanos (CGRH) do MJ foram as seguintes: quem é o responsável por publicar os *datasets*?; qual é o critério de seleção dos dados a serem publicados?; e quais os documentos arquivísticos usados como fonte dos dados a serem publicados?

A compreensão sobre a relevância da Comissão Permanente de Avaliação de Documentos (CPAD) nas questões referentes aos instrumentos de gestão de documentos arquivísticos, o e-SIC também foi adotado para coletar dados acerca de sua participação nas ações da iniciativa para dados abertos do MJ. E os seguintes questionamentos foram formulados: se a CPAD foi convidada para participar do projeto; se foi estabelecido contato

com o Arquivo Nacional (AN) para apoio a essa iniciativa e, em caso positivo, quais documentos arquivísticos registram essa participação; se a CPAD instituiu a política arquivística para o MJ; e se há um programa de gestão de documentos formalizado.

A compilação dos dados coletados foi realizada a partir da leitura das respostas aos questionamentos feitos ao Ministério da Justiça, cuja análise será apresentada na seção 8.

A proposta de um instrumento de *benchmark* para a publicação de dados governamentais abertos, tendo como apoio preceitos da gestão arquivística de documentos, foi elaborada para este estudo de caso a partir da adaptação de dois outros instrumentos de *benchmark* internacionais: o documento *Benchmark requirements supporting the presumption of authenticity of electronic records*, do Projeto InterPARES 2 (ANEXO B) e o documento *Governo Aberto e documentos arquivísticos confiáveis: instrumento de referência para marcos regulatórios/institucionais e capacidades*, do International Records Management Trust (ANEXO C).

7 ANÁLISE DIPLOMÁTICA DO *DATASET* “LISTA DE TERCEIRIZADOS DO MINISTÉRIO DA JUSTIÇA”

Introdução

A realização da análise diplomática de um documento visa atestar sua veracidade e também identificar se ele tem as características necessárias para ser reconhecido como um documento arquivístico. Esse reconhecimento faz-se relevante, pois os procedimentos de gestão e preservação de documentos arquivísticos, especialmente em ambientes digitais, diferem-se de outros objetos digitais, como os documentos bibliográficos ou um *dataset*, por exemplo.

A Diplomática é a ciência cujo arcabouço teórico apresenta terminologia e princípios para a análise de um documento a fim de verificar sua autenticidade ou falsidade. Segundo Tognoli (2010, p. 18) “[...] devido ao crescente número de falsificações, no século VI, foram introduzidas no Código Civil de Justiniano, regras práticas para distinguir os documentos autênticos dos falsos. Essas regras, porém, eram aplicáveis apenas aos documentos contemporâneos a ela, sendo impossível aplicá-las aos de épocas anteriores”.

No século XX, a adoção dos elementos para a crítica diplomática cunhados a partir de um contexto histórico-social medieval fica obsoleta e, por isso, estudiosos da Arquivologia testam a aplicação da Diplomática “clássica” nos conjuntos de documentos arquivísticos pós século XX, dentre eles Robert-Henri Bautier, Paola Carucci, Bruno Delmas e Luciana Duranti. Para Bautier (1961 apud TOGNOLI, 2010, p. 87), a Arquivologia e a Diplomática são duas ciências que têm em comum o “documento de arquivo” como objeto de estudo.

Nos estudos de Luciana Duranti foram apontados novos usos para a Diplomática, isto é, a aplicação de seus conceitos e métodos ao documento arquivístico contemporâneo. A metodologia de pesquisa adotada pelo Projeto InterPARES 2 abarca a análise diplomática como uma ferramenta para analisar os estudos de caso desenvolvidos pelo Projeto pois, para suas finalidades,

[...] a análise diplomática restringiu-se ao teste de cada tipo de objeto digital com base nas cinco características necessárias a um documento arquivístico visando determinar se cada objeto poderia ser reconhecido como um documento arquivístico ou se era identificado mais apropriadamente como, por exemplo, dado, documento ou publicação. Documentos não arquivísticos geralmente exigem um modelo de

preservação mais simplificado porque eles existem automaticamente a partir de outros documentos e com o propósito de disseminação de informação⁶³. (HACKETT; UNDERWOOD; EPPARD, 2008, p. 13, tradução nossa).

Esse mesmo teste foi aplicado ao *dataset* sobre funcionários terceirizados contratados pelo Ministério da Justiça e publicado em seu sítio na Internet, na seção dedicada aos dados abertos.

Identificação do objeto digital

Uma planilha de cálculo, neste caso em formato .xls, é um objeto de informação formado por uma cadeia de dígitos binários que é lida por sistemas computacionais. De acordo com Thibodeau (2002, tradução nossa) “um objeto digital é um objeto de informação, qualquer tipo de informação ou formato, expressado de forma digital”⁶⁴.

Para esse mesmo autor, o objeto digital é constituído por três níveis, a saber:

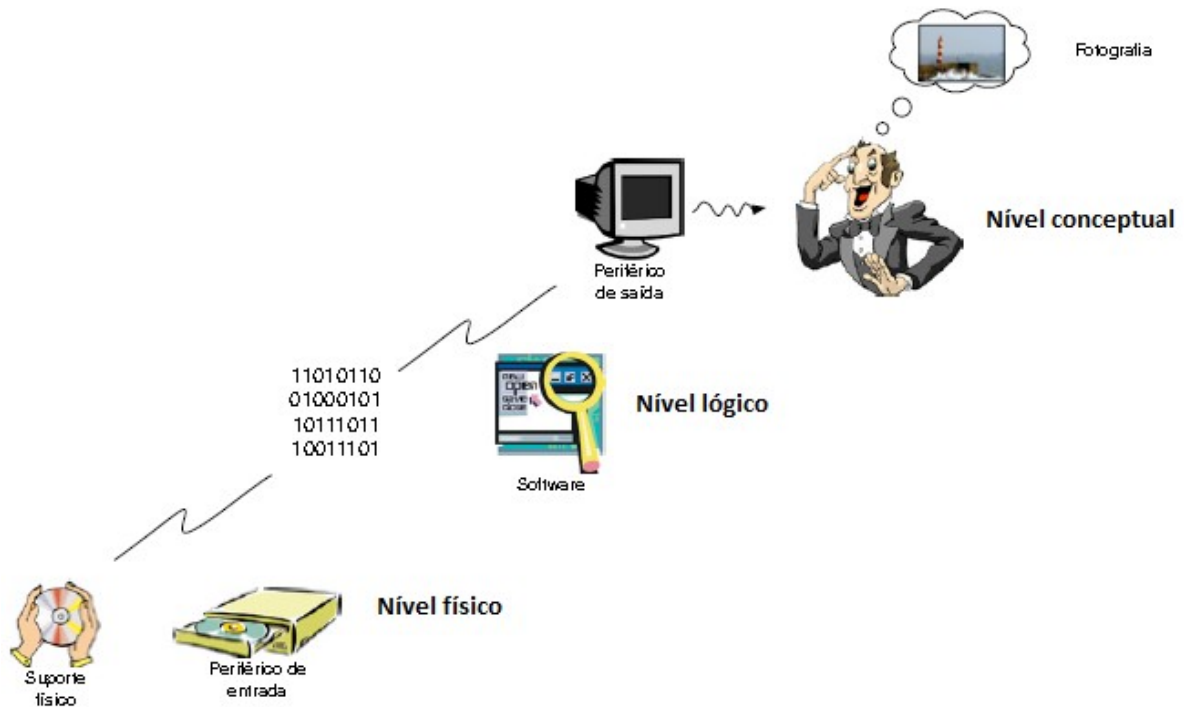
- a) nível físico: aquele em que o objeto é registrado; o suporte.
- b) nível lógico: aquele em que estão as cadeias de bits; os códigos binários.
- c) nível conceitual: aquele em que é compreendido pelo ser humano ou outro sistema computacional.

A representação desses níveis foi elaborada pelo autor português Miguel Ferreira (Figura A-1) a fim de demonstrar a “anatomia de um objeto digital”.

63 The diplomatic analysis was largely restricted to the testing of each type of digital object against the five necessary characteristics of a record to determine if each object could be considered a record, or whether an object was more appropriately identified as, for example, data, documents or publications. Non-records generally require a simpler preservation model because they exist autonomously from other documents and their purpose is, typically, limited to dissemination of information.

64 [...] a digital object is an information object, of any type of information or any format, that is expressed in digital form.

Figura A-1 – Níveis de um objeto digital.



Fonte: FERREIRA, Miguel. Introdução à preservação digital: conceitos, estratégias e actuais consensos, 2006. Disponível em: <<https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>.

Questões para a análise diplomática

1. Para ser identificado como um documento arquivístico, a entidade digital tem que possuir conteúdo estável e forma fixa e estar fixado em mídia estável.

A análise do objeto digital foi realizada no nível conceitual, isto é, na forma que é compreensível pelo ser humano. O conteúdo é estável, tem forma fixa e está fixado numa mídia estável.

2. Um documento arquivístico tem que participar de uma ação, definida como o exercício consciente de uma vontade praticada por pessoa física ou jurídica, com o objetivo de criar, manter, modificar ou extinguir situações. Um documento arquivístico é o resultado natural da ação.

A terceirização de serviços na administração pública federal está prevista na Lei nº 8.666, de 1993. Para tal finalidade, empresas privadas participam de um processo de licitação que, se cumprirem com as exigências regulamentadas por essa lei, vencem e, então, é

efetivado o contrato com o órgão contratante. O registro dessa ação é feito por meio de edital de licitação, propostas de serviço das empresas candidatas, contratos efetuados e relatórios periódicos de contratos de terceirização, portanto, documentos arquivísticos.

O *dataset* analisado, por sua vez, não tem o “objetivo de criar, manter, modificar ou extinguir situações” e, sim, publicizar dados sobre tipo de serviço, número do contrato, razão social, fiscal do contrato/órgão, objeto do contrato, período de vigência, cadastro de pessoa física (CPF), nome do terceirizado, função e local da prestação do serviço. Esses dados são extraídos dos “Relatórios Quadrimestrais de Contratos de Terceirização de Mão-de-Obra” elaborados conforme previsto no §1º, do art. 125 da Lei nº 13.080, de 02 de janeiro de 2015 (LDO 2015) e em ofícios-circulares da Controladoria-Geral da União (CGU).

Por isso, a finalidade do *dataset* é dar transparência à ação de contratações de empresas privadas para prestação de mão de obra terceirizada e ser reutilizado por qualquer pessoa interessada no assunto. Ele não participa da atividade de contratação de terceiros.

3. Um documento arquivístico tem que possuir relação orgânica com os demais documentos, estejam eles dentro ou fora do sistema informatizado. A relação orgânica é definida como o elo de cada documento ao anterior e ao subsequente da mesma ação e, de forma incremental, a todos os documentos arquivísticos que participam de uma mesma ação.

Conforme apontado no item 2, o *dataset* não participa da ação de contratação de prestadores de mão de obra terceirizada, portanto, não tem relação orgânica com documentos arquivísticos que registram tal atividade.

4. A criação de um documento arquivístico tem que envolver ao menos três pessoas, mesmo que não apareçam explicitamente no documento. Estas pessoas são o autor, o destinatário e o redator; no ambiente eletrônico uma delas deve necessariamente envolver duas outras pessoas: o produtor e o originador.

a) Autor: pessoa física ou jurídica com autoridade e capacidade de produzir o documento arquivístico ou em nome de outra pessoa ou por ordem de outrem o documento arquivístico é produzido.

O autor do *dataset* é o Ministério da Justiça.

b) Redator: pessoa física ou jurídica com autoridade e capacidade de articular o conteúdo

do documento arquivístico.

O redator do *dataset* é um servidor público, designado como “fiscal de contrato”, lotado na Coordenação de Contratos, da Coordenação-Geral de Logística, do Ministério da Justiça.

c) Destinatário: pessoa física ou jurídica a quem o documento arquivístico está endereçado ou direcionado.

O destinatário do *dataset* é o cidadão brasileiro ou estrangeiro.

d) Produtor: fundo ao qual o documento arquivístico integra.

Não foi possível identificar a partir do objeto conceitual que foi analisado.

e) Originador: pessoa à qual uma conta de Internet é disponibilizada ou o servidor de arquivo ao qual o documento arquivístico pertence.

O *dataset* está disponível para *download* do sítio do Ministério da Justiça, na seção “Dados abertos”.

5. Um documento arquivístico tem que possuir contexto identificável, definido como o ambiente que envolve a ação na qual o documento participa. Os tipos de contexto são: jurídico-administrativo, de proveniência, de procedimentos, documental e tecnológico.

a) O contexto jurídico-administrativo é o sistema legal e organizacional ao qual a instituição produtora pertence.

O Ministério da Justiça (MJ) é um dos órgãos que integram o Comitê Interministerial Governo Aberto (CIGA), conforme Decreto s/nº de 15 de setembro de 2011. A Portaria nº 1.378, de 20 de agosto de 2014, que “institui, no âmbito do Ministério da Justiça, a Política de Dados Abertos e Espaciais e aprova o Plano Institucional de Dados Abertos e Espaciais” (BRASIL, 2014d), contempla, em seu artigo 1º, a instituição da Política e aprovação do Plano; já o artigo 5º dedica-se à finalidade do Plano Institucional elaborado pela Secretaria-Executiva do MJ.

O MJ é um dos órgãos do Poder Executivo federal que publica dados governamentais no Portal Brasileiro de Dados Abertos e realiza concursos para o desenvolvimento de aplicativos (*apps*) que possibilitam o acompanhamento das atividades desempenhadas pelo órgão, com destaque para defesa do consumidor e segurança nas estradas federais. É pertinente dizer ainda que o MJ também publica *datasets* em seu sítio na Internet, sendo

alguns deles cópias dos que estão no Portal.

b) O contexto de proveniência se refere à entidade produtora, seu mandato, estrutura e funções.

Não foi possível identificar o produtor a partir do objeto conceitual analisado. Porém, de acordo com o “Plano Institucional de Dados Abertos e Espaciais”, do Ministério da Justiça, a responsabilidade pelo *dataset* é da Coordenação-Geral de Recursos Humanos (CGRH), subordinada à Secretaria-Executiva (SE).

c) O contexto de procedimentos compreende os procedimentos relativos às atividades no curso das quais o documento é produzido.

Não se aplica porque o *dataset* não é um documento arquivístico.

d) O contexto documental é definido como o fundo arquivístico ao qual o documento pertence e sua estrutura interna.

Não se aplica porque o *dataset* não é um documento arquivístico.

e) O contexto tecnológico é definido como as características dos componentes tecnológicos de um sistema informatizado no qual os documentos são criados.

Os *datasets* são planilhas de cálculo, no formato proprietário .xls, Microsoft Excell.

Conclusões da análise diplomática

O *dataset* analisado neste estudo de caso é um dado bruto com objetivo de ser publicado no sítio do Ministério da Justiça, na seção “Dados abertos”.

A metodologia do InterPARES não define qual a porcentagem de elementos precisa ser atingida por um documento para que ele seja reconhecido como arquivístico. Porém, a principal característica do documento arquivístico é apresentar relação orgânica e participar de uma ação. Assim, o *dataset* não é um documento arquivístico por não ter essas duas características arquivísticas.

8 ANÁLISE E DISCUSSÃO DOS RESULTADOS

O arcabouço teórico construído e desenvolvido aqui visou fundamentar a compreensão do universo de pesquisa escolhido e, a partir daí, subsidiar a coleta de dados, objetivando investigar se o *dataset* publicado na seção "Dados abertos" do sítio do MJ é um documento arquivístico ou um dado bruto. Nessa seção do sítio, foram publicados *datasets* de 14 temas⁶⁵ diferentes, a saber (BRASIL, 2015f):

1) Lista de cartórios do Brasil: informações sobre Cartórios do Brasil mantidas pelo Conselho Nacional de Justiça – CNJ;

2) Comunidades terapêuticas: informações sobre Comunidades terapêuticas, entidades privadas, sem fins lucrativos, que realizam acolhimento de pessoas com transtornos decorrentes do uso, abuso ou dependência de substâncias psicoativas;

3) Postos de Campanha do desarmamento: informações sobre os postos de entrega voluntária de armas;

4) Unidades dos Procons: informações sobre Procons mantidas pela Secretaria Nacional do Consumidor - Senacon/MJ;

5) Atendimento de consumidores nos Procons (Sindec): o Boletim Sindec é a publicação periódica da Senacon que consolida todos os atendimentos (demandas) registrados pelos Procons integrados ao Sindec, para permitir uma análise sobre os assuntos, problemas e fornecedores que mais levam os consumidores aos Procons;

6) Cadastro Nacional de reclamações fundamentadas (PROCONS – Sindec): o Cadastro Nacional de Reclamações Fundamentadas é o cadastro formado pelas Reclamações finalizadas pelos Procons integrados ao Sindec, no período de 12 meses. Apresenta anualmente a consolidação dos Cadastros publicados por estados e municípios de todas as regiões brasileiras;

7) Sistema BR-Brasil – boletins de ocorrências em Rodovias Federais: o Sistema BR-Brasil cataloga todos os boletins de ocorrência registrados após um acidente em uma rodovia federal. São várias as informações cadastradas, incluindo diversos detalhes sobre os veículos, as pessoas envolvidas e as condições do local, resguardando-se os dados pessoais;

65 Disponível em: <<http://www.justica.gov.br/dados-abertos>>. Acesso em 12 de maio de 2015.

8) Sistema Multa – autuações de trânsito em Rodovias Federais: informações coletadas na ocasião de uma autuação de trânsito em rodovias federais são catalogadas, resguardando-se os dados pessoais, que permitem a identificação do infrator;

9) Postos e delegacias do Departamento de Polícia Rodoviária Federal (DPRF): a base consolida a localização e contato de todos os postos e delegacias do Departamento de Polícia Rodoviária Federal distribuídos pelo país;

10) Postos e delegacias do Departamento de Polícia Federal (DPF): a base consolida a localização e contato de todos os postos e delegacias do Departamento de Polícia Federal distribuídos pelo país;

11) Reclamações do Consumidor.gov.br: informações anonimizadas sobre as reclamações cadastradas no Consumidor.gov.br, sítio criado para solução alternativa de conflitos de consumo por meio da Internet, que permite a interlocução direta entre consumidores e empresas;

12) Recall – campanhas de chamamento: a base de dados se refere às informações fornecidas pela Secretaria Nacional do Consumidor sobre os *recalls* (forma pela qual um fornecedor vem a público informar que seu produto ou serviço apresenta riscos aos consumidores podendo ser recolhido) realizados em território brasileiro;

13) Lista de terceirizados do Ministério da Justiça: informações sobre mão de obra terceirizada contratada para trabalhar nos diferentes órgãos do MJ; e

14) Lista de estagiários do Ministério da Justiça: informações sobre mão de obra estagiária atuando nos órgãos do MJ.

Dentre os temas listados, o objeto de investigação selecionado foi o *dataset*, identificado como “Lista de terceirizados do Ministério da Justiça” em formato *.xls. Isso porque:

1) é um *dataset* publicado em formato proprietário. De acordo com os princípios definidos para dados governamentais abertos, os formatos abertos e não proprietários são adequados para a publicação de *datasets*, pois possibilita o reúso de dados sem restrições tanto no âmbito tecnológico quanto no financeiro. Por isso, o formato *.xls (planilhas de cálculo do Microsoft Office) não é condizente com os princípios dos dados governamentais abertos;

2) assemelha-se a um documento arquivístico digital em sua forma manifestada. Um documento digital em sua forma manifestada é aquele compreensível por seres humanos ou sistemas informatizados, e um documento arquivístico é o registro das atividades de uma pessoa física ou jurídica visando a uma ação ou referência. Por meio da observação direta, pôde-se inferir que as planilhas de cálculo publicadas pela Coordenação de Contratos, da Coordenação Geral de Logística do Ministério da Justiça (CCONT/CGL/MJ) apresentam essa semelhança.

Nesta exposição, destaca-se a análise qualitativa do *dataset*, em sua forma compreensível ao ser humano (de exibição na tela do computador), como um documento arquivístico.

Para um documento digital ser reconhecido como arquivístico, é necessário apresentar as seguintes características: assumir forma fixa e conteúdo estável; estar fixado em mídia estável; registrar ou apoiar uma ação; ter relação orgânica; envolver cinco indivíduos ou, no mínimo, três deles na ação; e apresentar contexto identificável.

A planilha de cálculo publicada como *dataset* contém os seguintes campos: tipo de serviço, número do contrato, razão social, fiscal do contrato/órgão, objeto do contrato, período de vigência, cadastro de pessoa física (CPF), nome do terceirizado, função e local da prestação do serviço (Quadro 10).

Quadro 10 – Campos do *dataset* “Lista de terceirizados do Ministério da Justiça”.

Tipo de Serviço	Tecnologia da Informação		
Nº Contrato	38/2009		
Razão Social	Solução Serviços Especializados Ltda		
Fiscal do Contrato/Órgão	Paulo César Sousa		
Objeto do contrato	Prestação de serviços especializados na área de tecnologia da informação para a execução de suporte técnico remoto e presencial a usuários do Ministério da Justiça.		
Período de Vigência	19/05/2012 - 18/05/2013		
CPF (***.xxx.xxx-**) 	Nome do terceirizado	Função	Local da prestação do serviço
***.096.061-**	Alessandro da S F Cavalcanti	Analista de Prospecção Técnica	Ministério da Justiça

Fonte: extrato da tabela “Lista de servidores terceirizados do Ministério da Justiça”, fevereiro de 2013, aba TI-Contr 38-2009.

Na Administração Pública brasileira, a contratação de mão de obra terceirizada está prevista na Lei nº 8.666, de 21 de junho de 1993 e tal atividade é contemplada pela área meio

dos órgãos e entidades públicos registrada em relatórios compilados quadrimestralmente pelo seu produtor.

Ao se averiguarem quais eram os campos para inserção de dados nas planilhas do período de janeiro a dezembro de 2013, constatou-se que eles se repetiam em meses e anos posteriores, porém, sem uma motivação explícita.

Em resposta ao pedido de informação SIC/MJ nº 08850.000375/2015-65, foi informado que a CGU, por meio de ofícios-circulares, comunica aos órgãos e entidades o cronograma e como têm que ser enviadas as informações de terceirizados a esse órgão de controle externo. E que os dados a serem preenchidos nas planilhas de cálculo foram pré-definidos nas Leis de Diretrizes Orçamentárias, publicadas anualmente, como no art. 125 da LDO de 2015, por exemplo:

Art. 125. Os instrumentos de contratação de serviços de terceiros deverão prever o fornecimento pela empresa contratada de informações contendo nome completo, CPF, cargo ou atividade exercida, lotação e local de exercício dos empregados na contratante, para fins de divulgação na *internet*.

§ 1º Os órgãos e entidades federais deverão divulgar e atualizar quadrimestralmente as informações previstas no caput. (BRASIL, 2015a).

A predefinição de quais dados governamentais serão publicizados pelos seus produtores para reuso dos cidadãos suscita a reflexão sobre um governo “semiaberto”, cujos conjuntos documentais ostensivos não são disponibilizados em sua íntegra e, sim, alguns itens pinçados, os *datasets*, sob a alegação de que são os temas com maior demanda pelo cidadão brasileiro.

A partir da análise das planilhas de cálculo publicadas pelo MJ, identificou-se que se trata de um documento digital em sua forma compreensível ao ser humano e de um documento digital com conteúdo informacional para atender ao interesse de um receptor, no caso, o consumidor de dados.

Em ambos os pontos, o documento digital não pôde ser caracterizado como “arquivístico” por não apresentar, sobretudo, relação orgânica com outros documentos, além de não participar de uma ação determinada. Isso porque o conjunto de dados teria sido construído fora do contexto de uma transação em que o MJ estivesse envolvido, devido às suas atividades-meio ou fim, mais especificamente, a atividade de contratação de mão de obra terceirizada.

Desde a observação direta da forma manifestada do *dataset* e o uso do recurso “Propriedades” do arquivo digital, foi possível identificar os seguintes elementos: autor, redator, data cronológica, assunto, forma fixa, conteúdo estável e contexto identificável, induzindo ao equívoco de reconhecê-lo como um documento arquivístico e confirmando o alerta feito por Duranti (1994) de que há uma maior complexidade na distinção entre um documento comum e um documento arquivístico e este último de um conjunto de dados num ambiente digital: “[...] o que distingue um documento arquivístico de um documento que não é arquivístico é a natureza de seu relacionamento com a atividade do produtor em vez de suas características de forma ou conteúdo”⁶⁶. (DURANTI; THIBODEAU, 2008, p. 348).

Assim, o *dataset* publicado na seção destinada aos dados abertos no sítio do MJ é um dado bruto, um recorte do relatório quadrimestral que, por sua vez, constitui-se num documento que participa da atividade de contratação de mão de obra terceirizada, ou seja, arquivístico.

No entanto, a reflexão feita sobre um *dataset* publicado na Internet se tratar de um documento arquivístico continua latente. A depender do seu contexto e finalidade de produção, as características apresentadas na seção 2.1, quadro 2, estarão presentes em tal *dataset*. Faz-se pertinente, porém, a observação de que, se um documento arquivístico não apresentar todas as suas características, ele pode ser qualificado como um mau documento arquivístico (*bad record*).

Conforme descrito, o *dataset* “Lista de terceirizados do Ministério da Justiça” é um dado bruto publicado com a finalidade de reuso pelo consumidor de dados e, por isso, precisa ter qualidades para que novas informações sejam produzidas corretamente.

A identificação de quais qualidades são inerentes a um dado baseou-se na revisão teórica sobre qualidade de dados na seção 3. Dessa maneira, foram identificadas acurácia, completeza e relevância do dado/informação como as qualidades defendidas por dois diferentes grupos de pesquisadores da área de Computação.

Acurácia e completeza foram identificadas por estudiosos da área de Arquivologia e Diplomática como qualidades próprias de documentos arquivísticos. Assim, ao se cotejarem as suas qualidades com as de um dado, chegou-se ao resultado de que são comuns, além das já citadas, acessibilidade e confiabilidade.

66 [...] what distinguish a record from a document that is not a record is the nature of its relationship with the activity of the creator rather than its formal or content characteristics.

No momento em que foram correlacionadas as qualidades de dados e de documentos arquivísticos, observa-se a pertinência do questionamento sobre a participação da Comissão Permanente de Avaliação de Documentos/MJ (CPAD/MJ) na iniciativa de dados abertos, no âmbito do MJ. A responsabilidade de uma CPAD pela organização e tratamento de documentos arquivísticos está estabelecida no Decreto nº 4.073, de 3 de janeiro de 2002, que regulamenta a Lei nº 8.159, de 8 de janeiro de 1991.

Em resposta ao pedido de informação SIC/MJ nº 08850.001083/2015-40, apurou-se que o MJ constituiu a sua CPAD/MJ através da Portaria SE nº 825, de 22 de julho de 2013. Nessa mesma portaria, constam as suas atribuições, os servidores que a integram e o apoio para seu funcionamento. Cabe ressaltar que nessa Portaria não há menção à responsabilidade da CPAD para com as ações de governo aberto ou dados governamentais abertos.

A participação da CPAD na elaboração de uma política arquivística é necessária, já que, em sua composição, há representantes de diferentes unidades administrativas de um órgão, o que possibilita uma visão macro das ações a serem implementadas. Com isso, observa-se que a instituição de uma política arquivística num órgão ou entidade intensifica o desenvolvimento das ações de gestão e preservação de documentos arquivísticos, independentemente do suporte.

No MJ, porém, ainda não há tal política devido ao alto número de membros que compõem a CPAD e, conseqüentemente, à impossibilidade de conciliação de agendas de reuniões e atualizações do quadro de membros, segundo o coordenador da CDI/MJ:

[...] a CPAD do Ministério da Justiça é composta por dois servidores de cada Secretaria, Departamento, Subsecretaria e algumas Coordenações-Gerais. Por causa do tamanho da estrutura do MJ, a composição da CPAD tem 27 integrantes. Temos muita dificuldade em manter estes 27 integrantes atualizados assim como realizar reuniões técnicas. A CPAD do MJ está trabalhando atualmente na elaboração do Código de Classificação de Documentos das Atividades Finalísticas do MJ. (Coordenador da CDI/MJ).

No que se refere à existência ou não de um programa de gestão de documentos formalizado no MJ, o coordenador da CDI/MJ declarou que

A política de gestão documental do MJ é composta de portarias normativas elaboradas pela equipe técnica da Coordenação de Documentação e Informação e submetida à apreciação e aprovação da Consultoria Jurídica e Secretaria-Executiva. (Coordenador da CDI/MJ, grifos nossos).

A resposta dá margem a um entendimento equivocado por parte do coordenador da CDI/MJ acerca do que consistem uma política arquivística e um programa de gestão de documentos. A primeira está pautada em agenda definida, em conjunto, por responsáveis da alta administração (neste caso, do MJ) e outros servidores envolvidos na gestão e preservação de documentos arquivísticos, a fim de estabelecer as diretrizes para a disponibilização dos documentos públicos arquivados no MJ.

Um programa de gestão de documentos está contemplado numa política arquivística. Ele tem por finalidade elencar as ações necessárias à implantação de procedimentos quanto à produção, uso/manutenção e destinação dos documentos arquivísticos, abrangendo a elaboração de instrumentos de gestão e a capacitação do corpo funcional. Por meio da publicação de portarias da Secretaria-Executiva (SE) e do Gabinete do Ministro (GM) – ambos do MJ –, o Ministério, de acordo com o coordenador da CDI/MJ, tem

normativos referentes à utilização dos serviços postais, procedimentos de protocolo, transferência de arquivos, empréstimo, consulta e desarquivamento de arquivos, utilização do sistema informatizado de gestão documental, eliminação e recolhimento de arquivos. (Coordenador da CDI/MJ).

O MJ, por meio da Coordenação de Documentação e Informação (CDI), tem institucionalizado as atividades de gestão de documentos arquivísticos, porém, ações para a sua preservação ainda não foram evidenciadas em atos normativos, salvo a Portaria nº 1.246, de 31 de dezembro de 2014, que “Institui o Comitê Gestor do Sistema Eletrônico de Informações - SEI e estabelece os procedimentos de gestão de documentos, processos e arquivos pelo SEI, no âmbito do Ministério da Justiça”. No §3º, do art. 31, consta que “A CGMA e a CGTI deverão desenvolver um plano de preservação de documentos digitais, a ser submetido e aprovado pelo Comitê Gestor do SEI”. (BRASIL, 2015b).

A partir do exposto, verifica-se que há deficiências na atuação da CPAD/MJ, e o fato de isso não estar sendo considerado na implementação de ações dedicadas à iniciativa de dados governamentais abertos compromete a publicação de *datasets* de qualidade para reuso dos consumidores de dados, pois os documentos arquivísticos também são fontes de dados governamentais e exigem ações de gestão e preservação adequadas, especialmente os digitais.

Quando o coordenador da CDI/MJ declara que “A CPAD do MJ não participou da elaboração do Plano Institucional de Dados Abertos e Espaciais”, isso pode ser visto como um indício, em sentido amplo, de que o tema dados abertos está afeto apenas às áreas de

Computação e Tecnologia da Informação.

Ainda que a CPAD/MJ não tenha sido participante da elaboração do plano institucional de dados abertos e espaciais (PDAE/MJ), mesmo porque não é uma atribuição direta, o Arquivo Nacional – como a principal instituição arquivística do país e vinculada ao MJ – seria o órgão apropriado para compor o grupo de elaboração do PDAE/MJ. A governança desse Plano está no bojo do “Programa de Transparência e Acesso a Informações”, da Secretaria-Executiva/MJ, e o respondente da SE informou que “[...] o Arquivo Nacional não foi consultado na elaboração do Plano Institucional de Dados Abertos e Espaciais do Ministério da Justiça”.

No documento norteador do PDAE/MJ foram apresentadas as ações do MJ diante do compromisso de abertura de dados governamentais. Diferentes bases de dados com informações acerca de unidades do Procon, postos e delegacias dos Departamentos de Polícia Rodoviária Federal e da Polícia Federal, por exemplo, encontram-se publicizadas no Portal Brasileiro de Dados Abertos. Entretanto, a planilha de cálculo denominada “Lista de terceirizados do Ministério da Justiça” está publicada apenas na seção “Dados abertos” do sítio do MJ.

Com base na análise dos dados coletados para esta pesquisa, julgou-se relevante a elaboração de uma proposta de um instrumento de referência (*benchmark tool*) para assegurar identidade e integridade de documentos arquivísticos fontes de dados governamentais, presumindo, assim, a autenticidade desses documentos.

9 PROPOSTA DE INSTRUMENTO DE *BENCHMARK*

A finalidade de um instrumento de *benchmark* é estabelecer padrões ou linhas de base para aprimorar o desempenho de uma organização, pública ou privada, em determinada área ou atividade específica, como no caso da publicação de *datasets* com qualidade para uso e reúso do consumidor de dados.

Ao tomar-se como premissa que documentos arquivísticos também são fontes de dados, então, a gestão arquivística de documentos apoia a atividade do Ministério da Justiça de publicar *datasets* na seção específica para dados abertos em seu sítio na Internet, pois ela oferece o arcabouço teórico para produção, uso, manutenção, destinação e preservação de documentos como evidência das atividades desenvolvidas pelo órgão.

Dentre as qualidades de um dado governamental aberto está a atualidade de tal dado. Considera-se, aqui, que dados atuais são aqueles coletados em arquivo corrente, porém, antes de um documento arquivístico atingir a primeira idade do arquivo, o foco é na produção e uso desse documento.

Por isso, esta proposta de instrumento de *benchmark* foi elaborada para aplicação no nível básico da gestão de documentos, isto é, na produção do documento arquivístico digital, sendo aplicável em unidades produtoras de documentos arquivísticos que possuem sistema informatizado de gestão arquivística (SIGAD), gerenciamento eletrônico de documentos (GED) ou gerenciem “manualmente” aqueles natos digitais, resguardados os graus de dificuldade para cada recurso.

O documento arquivístico tem como uma de suas qualidades a autenticidade, isto é, a capacidade de ser o que realmente diz ser, estando livre de adulterações ou corrupções. No ambiente digital, os documentos arquivísticos estão suscetíveis à corrupção de suas cadeias de bits no momento de sua transmissão e à obsolescência tecnológica:

[...] a presunção da sua autenticidade deve se apoiar na evidência de que eles [os documentos arquivísticos] foram mantidos com uso de tecnologias e procedimentos administrativos que garantiram a sua identidade e integridade (componentes da autenticidade); ou que, pelo menos, minimizaram os riscos de modificações dos documentos a partir do momento em que foram salvos pela primeira vez e em todos os acessos subsequentes. (CONSELHO NACIONAL DE ARQUIVOS, 2012, p. 1).

Nesta proposta (Quadro 11), cuja motivação foi um modelo teórico e não se fez a sua aplicação prática, as linhas de base são para presunção de autenticidade de documentos arquivísticos digitais, sendo composta pela identidade e integridade desses documentos. A

produção e manutenção de documentos arquivísticos autênticos faz-se relevante também para a publicação de dados governamentais abertos e, sobretudo, para confiança na administração pública.

Quadro 11 – Proposta de instrumento de *benchmark*.

INSTITUIÇÃO DA POLÍTICA ARQUIVÍSTICA DO ÓRGÃO	
	Estabelecido? S (sim) N (não)
Proposta definindo os eixos de atuação (funções arquivísticas) e as diretrizes a serem executadas para a estruturação e organização do arquivo.	
Proposta de diálogo contínuo entre o serviço arquivístico e as autoridades de Governo Aberto/ governo eletrônico/ acesso à informação e Tecnologias de Informação e Comunicação (TIC), no tocante a iniciativas acerca das informações governamentais, pois há uma interdependência da gestão de documentos/ informações e a abertura de dados e a lei de acesso à informação.	
Plano de conscientização do corpo funcional a fim de estimular uma cultura de acurácia de dados e de acesso do cidadão à informação, com base em fontes documentais confiáveis e íntegras, em conformidade com leis, regulamentos e normas.	
IMPLANTAÇÃO DOS PROGRAMAS DE GESTÃO E PRESERVAÇÃO DE DOCUMENTOS	
	Estabelecido? S (sim) N (não)
Cronograma para a produção e atualização de normas e orientações voltadas para a gestão e preservação dos principais tipos de documentos públicos, digitais ou não digitais.	
Programa de inspeção de gestão de documentos nas unidades organizacionais. O serviço arquivístico efetua inspeções regulares e possui autoridade para exigir a observância de normas e políticas de gestão e preservação de documentos (p. ex. regras para a destinação de documentos).	
Projeto de cooperação com autoridades em dados abertos/ acesso à informação, orientando as unidades organizacionais sobre aprimoramento do acesso aos documentos e sobre as condições em que controles legítimos de acesso podem ser implementados, sem comprometer os objetivos gerais do Governo Aberto e do acesso à informação. O serviço arquivístico assume um papel de liderança na garantia de que os documentos arquivísticos estarão disponíveis para fornecer informações íntegras e confiáveis ao público e no desenvolvimento de normas de gestão de documentos.	
Programa de treinamento em boas práticas de gestão e preservação de documentos para as equipes das unidades organizacionais.	

EXPRESSÃO DOS ATRIBUTOS DO DOCUMENTO ARQUIVÍSTICO E CONEXÃO COM DOCUMENTO ARQUIVÍSTICO	
	Estabelecido? S (sim) N (não)
A – IDENTIDADE DO DOCUMENTO ARQUIVÍSTICO	
A.i Nomes das pessoas envolvidas na elaboração do documento arquivístico, sendo:	
Nome do autor	
Nome do redator (se diferente do autor)	
Nome do originador (se diferente do nome do autor ou do redator)	
Nome do destinatário	
A.ii Nome da ação ou assunto	
A.iii Data(s) da criação e transmissão, sendo:	
Data cronológica	
Data de recebimento	
Data de arquivamento	
Data(s) de transmissão	
A.iv Expressão da relação orgânica (p.ex. código de classificação, identificador do arquivo)	
A.v Indicação de anexos	
B – INTEGRIDADE DO DOCUMENTO ARQUIVÍSTICO	
B.i Nome do espaço de trabalho	
B.ii Nome da área de primeira responsabilidade (se diferente de B.i)	
B.iii Indicação dos tipos de anotações acrescentadas no documento arquivístico	
B.iv Indicação de modificações técnicas	
PRIVILÉGIOS/ PERMISSÕES DE ACESSO	
O produtor definiu e implementou efetivamente permissões de acesso concernentes à criação, modificação, anotação, mudança e destruição de documentos arquivísticos.	

PROCEDIMENTOS DE PROTEÇÃO: PERDA E CORRUPÇÃO DE DOCUMENTOS ARQUIVÍSTICOS	
O produtor estabeleceu e efetivamente implementou procedimentos para prevenir, pesquisar e corrigir perda ou corrupção dos documentos arquivísticos.	
PROCEDIMENTOS DE PROTEÇÃO: MÍDIA E TECNOLOGIA	
O produtor estabeleceu e efetivamente implementou procedimento para garantir a identidade e integridade contínuas dos documentos arquivísticos contra deterioração de mídia e mudanças tecnológicas.	
ESTABELECIMENTO DA FORMA DOCUMENTAL	
O produtor estabeleceu a forma documental dos documentos arquivísticos associado ao procedimento tanto de acordo com os requisitos do sistema jurídico ou aqueles do produtor.	
AUTENTICAÇÃO DOS DOCUMENTOS ARQUIVÍSTICOS	
Se a autenticação é requerida pelo sistema jurídico ou por necessidades da organização, o produtor estabeleceu regras específicas para aqueles documentos arquivísticos que devem ser autenticados, por quem e a validação da autenticação.	
IDENTIFICAÇÃO DO DOCUMENTO ARQUIVÍSTICO AUTORITÁRIO	
Se existirem múltiplas cópias de um mesmo documento arquivístico, o produtor estabeleceu procedimentos para identificar qual documento arquivístico é o autoritário.	
REMOÇÃO E TRANSFERÊNCIA DE DOCUMENTAÇÃO RELEVANTE	
Se há a transição de documentos arquivísticos das idades corrente para intermediária e permanente, quanto à remoção de documentos arquivísticos do sistema eletrônico, o produtor estabeleceu e efetivamente implementou procedimentos determinando que documentação tem que ser removida e transferida para o preservador conjuntamente aos documentos arquivísticos.	

Fonte: DIPLOMATIC ANALYSIS TEMPLATE (2008, apêndice 21a); INTERNATIONAL RECORDS MANAGEMENT TRUST ([2013]). Adaptado; tradução nossa.

10 CONSIDERAÇÕES FINAIS

Na administração pública, seja em cenário nacional ou internacional, itens como transparência, governança, *accountability* e prestação de contas têm sido divulgados como ações centrais dos governos, em suas diferentes esferas de atuação, com o objetivo de combater a corrupção e ampliar a participação cidadã nas ações governamentais.

Nos últimos cinco anos, essa participação dos cidadãos vem sendo impulsionada pelos órgãos públicos através da iniciativa de dados governamentais abertos. Com o reúso de dados governamentais brutos, qualquer pessoa é capaz de desenvolver ferramentas (aplicativos) para apontar irregularidades em contas públicas ou indicar em quais bairros da cidade há iluminação precária, por exemplo.

Os dados governamentais abertos, em estado bruto, sem interpretações, precisam seguir os princípios estabelecidos pela *Open Government Data*. Numa concepção ideal, os oito pontos a serem contemplados por um *dataset* são: completos, primários, atuais, acessíveis, compreensíveis por máquina, não discriminatórios, não proprietários e livres de licenças. Assim, o risco de um consumidor de dados produzir informações inadequadas ou incorretas é minimizado.

Ao assumir o ponto de vista de um consumidor de dados que se questiona sobre a origem desses dados governamentais e partir da premissa de que uma das fontes de tais dados são os documentos arquivísticos digitais, buscou-se, neste trabalho, investigar se o *dataset* denominado “Lista de terceirizados do Ministério da Justiça”, publicado na seção destinada aos dados governamentais abertos no sítio do Ministério da Justiça, é um dado bruto ou um documento arquivístico.

O MJ publica nessa seção do seu sítio catorze temas tidos como os de maior interesse para o cidadão brasileiro, como a contratação de mão de obra terceirizada para trabalhar no Ministério – atividade que envolve recursos financeiros públicos. As planilhas de cálculo com os dados referentes aos contratos efetuados pelo MJ, de janeiro a dezembro de 2013, foram selecionadas para esta pesquisa.

O estudo de caso realizado partiu do problema de que o MJ publicasse como *datasets* documentos arquivísticos em vez de dados brutos na seção “Dados abertos” de seu sítio na Internet. Após a análise dos dados coletados, pôde-se verificar que o *dataset* publicado trata-

se de um recorte do relatório quadrimestral sobre contratação de mão de obra terceirizada produzido pela área responsável.

No caso específico do MJ, não são documentos arquivísticos que estão sendo publicados como *datasets* e essa conclusão apoia-se na aplicação do modelo de análise diplomática adotado pelo Projeto InterPARES 2. Das cinco características que um documento precisa ter para ser reconhecido como arquivístico, o ideal é que ele apresente todas; caso contrário, é um *bad record*. Porém, a relação orgânica devido à participação numa atividade é obrigatória – característica esta que o *dataset* “Lista de terceirizados do Ministério da Justiça” não tem.

Não obstante, um *dataset* é passível de ser reconhecido como um documento arquivístico, dependendo do contexto de produção no qual está inserido, como as tabelas de um banco de dados com registros de identificação de funcionários, de salários, tempo de serviço, frequência, entre outros dados, para compor a folha de pagamento de pessoal, por exemplo, conforme apresentado na seção 2.

Com relação aos autores consultados na elaboração da fundamentação teórica, para fins deste trabalho, compreende-se que a definição de *dataset* aplicada no âmbito da iniciativa de dados governamentais abertos é aquela de dados estruturados compreensíveis por máquina, adotada pelas áreas computacionais.

O termo *dataset* é utilizado por diferentes grupos de usuários, de diferentes áreas do conhecimento para fins estatísticos ou censitários, por exemplo. À Arquivologia não cabe aprofundar-se na definição de *dataset*, porém, seu objeto de estudo – o documento arquivístico – é uma das principais fontes desses conjuntos de dados que, para terem utilidade no reuso por cidadãos neles interessados, precisam ter como qualidades acurácia, completeza, acessibilidade e confiabilidade; as mesmas qualidades inerentes a um documento arquivístico.

Outra reflexão necessária diz respeito ao princípio de que um *dataset* precisa ser atual, ou seja, quanto antes ele for publicado para reuso do consumidor de dados, maior será a sua utilidade. Em um arquivo, são atuais aqueles documentos em idade corrente cujo responsável é a unidade organizacional produtora ou o Arquivo Central. Por isso, faz-se necessário o envolvimento da Comissão Permanente de Avaliação de Documentos (CPAD) nas iniciativas de governo aberto e dados governamentais abertos, pois documentos arquivísticos também são fontes de dados governamentais.

Com relação à gestão arquivística de documentos digitais, presume-se a necessidade de ampliar a cooperação e a interação dos órgãos do Poder Executivo federal com o Arquivo Nacional (instituição arquivística da mesma esfera governamental), o que significa um afinamento do debate e a consequente definição de políticas públicas com a temática de governo aberto e dados governamentais abertos.

Outros estudiosos da Arquivologia também vêm abordando essa temática, ainda que timidamente. O que se observa, no entanto, é que há de se estimular estudos sobre a contribuição das funções arquivísticas para com as iniciativas de dados abertos, seja numa administração em âmbito público ou privado.

Nesse sentido, a elaboração de uma proposta de um instrumento de *benchmark* para ser adotado no Ministério da Justiça, campo empírico desta pesquisa, busca contribuir para que haja produção e manutenção de documentos arquivísticos autênticos no âmbito do MJ, com base em política e procedimentos arquivísticos e não apenas recursos computacionais.

O desafio de realizar um estudo de caso sobre *datasets* publicados dentro da iniciativa de dados abertos consiste no entendimento das diferentes abordagens para o mesmo objeto de estudo, considerando, por exemplo, o que é um dado ou um *dataset* na tradução adequada da literatura estrangeira e ser uma temática ainda em fase de consolidação conceitual.

A incipiência de literatura nacional sobre o tema e, sobretudo, a partir de um olhar arquivístico, proporciona uma sensação ambígua de insegurança intelectual e estímulo na busca por resposta ao questionamento do título desta pesquisa: “Um estudo de caso sobre *datasets* do Ministério da Justiça: dados brutos ou documentos arquivísticos?”. Os *datasets* são dados brutos e, por isso, não exigem procedimentos arquivísticos para o seu gerenciamento.

Governo eletrônico, governo aberto, dados governamentais abertos são as iniciativas em que o governo brasileiro está inserido, acompanhando o cenário internacional, na busca de transparência das ações governamentais, de participação cidadã nas tomadas de decisões e de colaboração da sociedade para definição de políticas públicas que atendam as suas necessidades. Mas o questionamento sobre quem é esse cidadão brasileiro a fazer uso de tais iniciativas precisa ser feito.

A lei brasileira de acesso à informação pública, a Lei nº 12.527/2011, prevê que os órgãos públicos, nas diferentes esferas governamentais, disponibilizem a informação pública de forma passiva ou ativa, isto é, o cidadão a obtém por meio da transparência passiva (quando ele faz o pedido de acesso) ou da transparência ativa (quando o órgão publica em seu sítio na *Web*, por exemplo). Nesse sentido, a publicação de *datasets* é uma forma de transparência ativa.

Os preceitos da Arquivologia quanto à difusão dos e acesso aos arquivos colaboram para que haja transparência ativa (difusão) e passiva (acesso) das informações públicas. É relevante que os órgãos e entidades públicos estejam atentos para o correto tratamento de seus arquivos correntes, intermediário e permanentes que também são fontes de dados governamentais.

Uma outra inquietação causada pela temática de dados governamentais abertos é com relação ao princípio de que esse dado precisa ser completo, isto é, “todos os dados públicos devem ser disponibilizados” na definição feita pelo Consórcio W3C (2011). Contudo, a prática do Ministério da Justiça é selecionar alguns dados de temas pré-determinados, não seguindo ao princípio do dado governamental aberto ser completo.

Portanto, para o aprimoramento deste estudo, torna-se relevante aprofundar a compreensão dos princípios dos dados governamentais abertos, do perfil de seu consumidor e a contribuição da gestão arquivística de documentos para o Governo Aberto brasileiro.

REFERÊNCIAS

- ALBERCH Y FULGUERAS, Ramón. **Archivos, calidad, memória y derechos democráticos: los desafíos del siglo XXI**. [19--]. 23p.
- ALVES, William Pereira. **Banco de dados: teoria e desenvolvimento**. Editora Érica, 2009.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ISO 9000**. Sistemas de gestão da qualidade: fundamentos e vocabulário. Rio de Janeiro, 2000. Disponível em: <<http://www.cefetsp.br/edu/jcaruso/apostilas/iso9000.pdf>>. Acesso em: 17 mai. 2014.
- BAIÃO, Fernanda. **Gerência de dados e informações.pdf**. Unirio/DIA/PPGI: Disciplina Gestão de Dados, 2014, (apresentação em eslaides). Disponível em: <<http://uniriodb2.uniriotec.br/login/index.php>>. Acesso restrito aos discentes.
- BARRAN, Alicia Casas de; GRIFFIN, Andrew; MILLAR, Laura; ROPER, Michael. **Gestion de documentos del sector publico desde una perspectiva archivística II: Organización y control de documentos administrativos**. Consejo Internacional de Archivos, Universidad de La Republica Archivo General, International Records Management Trust, [2008], 202p. Disponível em: <http://www.irmt.org/documents/educ_training/spanish/IRMT_Perspectiva_Archivistica_II.pdf>. Acesso em: 30 dez. 2013.
- BENCHMARK REQUIREMENTS SUPPORTING THE PRESUMPTION OF AUTHENTICITY OF ELECTRONIC RECORDS (APPENDIX 21A). In: DURANTI, Luciana; PRESTON, Randy. **International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records**. Pádova: Associazione Nazionale Archivistica Italiana (Anai), 2008. Disponível em: <<http://www.interpares.org>>. Acesso em: 17 mar. 2015.
- BENSON, Peter. ISO 8000: a new international standard for data quality. **Data quality PRO**, 2009. Disponível em: <<http://dataqualitypro.com/data-quality-pro-blog/iso-8000-new-international-standard-data-quality>>. Acesso em: 25 jun. 2014.
- BITTENCOURT, Paola Rodrigues. **Implementação da lei de acesso à informação no Poder Executivo Federal: uma análise a partir dos serviços arquivísticos**. Dissertação (Mestrado Profissional em Gestão de Documentos e Arquivos) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2014.
- BRASIL. Lei nº 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. **Conselho Nacional de Arquivos – legislação arquivística brasileira**. Disponível em: <http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?from_info_index=21&infoid=146&sid=54>. Acesso em: 07 dez. 2013.

BRASIL. Lei nº 8.666, de 21 de junho de 1993. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. **Planalto**. Brasília, 1993. Disponível em:

<http://www.planalto.gov.br/ccivil_03/leis/l8666compilado.htm>. Acesso em: 26 jun. 2015.

BRASIL. Lei complementar nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. **Planalto**. Brasília, 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm>. Acesso em: 25 set. 2014.

BRASIL. Decreto nº 4.073, de 3 de janeiro de 2002. Regulamenta a Lei nº 8.159, de 8 de janeiro de 1991, que dispõe sobre a política nacional de arquivos públicos e privados.

Planalto. Brasília, 2002. Disponível em:

<http://www.planalto.gov.br/ccivil_03/decreto/2002/d4073.htm>. Acesso em: 06 jun. 2015.

BRASIL. Decreto nº 4.915, de 12 de dezembro de 2003. Dispõe sobre o Sistema de Gestão de Documentos de Arquivo SIGA, da administração pública federal, e dá outras providências.

Conselho Nacional de Arquivos – legislação arquivística brasileira. Disponível em:

<http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?from_info_index=21&infoid=146&sid=54>. Acesso em: 07 dez. 2013.

BRASIL. Decreto s/nº, de 15 de setembro de 2011. Institui o Plano de Ação Nacional sobre Governo Aberto e dá outras providências. **Planalto**. Brasília, 2011a. Disponível em:

<https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/dsn/dsn13117.htm>. Acesso em: 28 jan. 2014.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Planalto**. Brasília, 2011b. Disponível em:

<http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm>. Acesso em: 07 dez. 2013.

BRASIL. Secretaria de Logística e Tecnologia da Informação. Instrução normativa nº 4, de 13 de abril de 2012. Institui a Infraestrutura Nacional de Dados Abertos – INDA. **GOV.BR**.

Brasília, 2012. Disponível em:

<<http://www.governoeletronico.gov.br/biblioteca/arquivos/instrucao-normativa-da-infraestrutura-nacional-de-dados-abertos-2013-inda/view?searchterm=Instru%C3%A7%C3%A3o%20normativa%20INDA>>. Acesso em: 27 jun. 2014.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Conheça o Programa de Governo Eletrônico Brasileiro**. Brasília, 2013a. Disponível em:

<<https://www.governoeletronico.gov.br/o-gov.br>>. Acesso em: 07 dez. 2013.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Histórico do Governo Eletrônico**. Brasília, 2013b. Disponível em: <<https://www.governoeletronico.gov.br/o->

gov.br/historico>. Acesso em: 23 abr. 2013.

BRASIL. Câmara dos Deputados. Projeto de Lei PL 7804, julho de 2014. **Câmara dos Deputados**. Brasília, 2014a. Disponível em:

<http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra;jsessionid=4806F6AEF7B620F08192EEB73C91F6F.proposicoesWeb1?codteor=1266800&filename=PL+7804/2014>. Acesso em: 20 jun. 2015.

BRASIL. Decreto nº 8.189, de 21 de janeiro de 2014. Aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções Gratificadas do Ministério do Planejamento, Orçamento e Gestão e remaneja cargos em comissão e funções comissionadas técnicas. **Planalto**. Brasília, 2014b. Disponível em:

<http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/Decreto/D8189.htm>. Acesso em: 29 abr. 2015.

BRASIL. Ministério da Justiça. **Institucional**. Brasília, 2014c. Disponível em:

<<http://www.justica.gov.br/Acesso/institucional>>. Acesso em: 23 ago. 2014.

BRASIL. Ministério da Justiça. Portaria nº 1.378, de 20 de agosto de 2014. Institui, no âmbito do Ministério da Justiça, a Política de Dados Abertos e Espaciais e aprova o Plano Institucional de Dados Abertos e Espaciais. **Diário Oficial da União**. Brasília, 2014d.

Disponível em: <<http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&pagina=33&data=21/08/2014>>. Acesso em: 25 set. 2014.

BRASIL. Secretaria de Logística e Tecnologia da Informação. **Portal Brasileiro de Dados Abertos**. Brasília, 2014e. Disponível em: <<http://www.dados.gov.br>>. Acesso em: 25 set. 2014.

BRASIL. Lei nº 13.080, de 02 de janeiro de 2015. Dispõe sobre as diretrizes para a elaboração e execução da Lei Orçamentária de 2015 e dá outras providências. **Planalto**. Brasília, 2015a. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13080.htm>. Acesso em: 30 jul. 2015.

BRASIL. Ministério da Justiça. Portaria nº 1.246, de 31 de dezembro de 2014. Institui o Comitê Gestor do Sistema Eletrônico de Informações - SEI e estabelece os procedimentos de gestão de documentos, processos e arquivos pelo SEI, no âmbito do Ministério da Justiça. **Diário Oficial da União**, 2015b. Disponível em:

<<http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=02/01/2015&jornal=1&pagina=78&totalArquivos=128>>. Acesso em: 05 jul. 2015.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Governo Eletrônico. **Logística e Tecnologia da Informação**. Brasília, 2015c. Disponível em:

<<https://www.planejamento.gov.br/ministerio.asp?index=7&ler=s897>>. Acesso em: 5 abr. 2015.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Processo eletrônico nacional – PEN**. Brasília-DF, 2015d. Encontro sobre Protocolo Integrado e Processo Eletrônico

Nacional: Novos Paradigmas para a Administração Pública Federal. Brasília-DF, março de 2015. Disponível em: <<http://www.siga.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm>>. Acesso em: 15 maio 2015.

BRASIL. Ministérios da Justiça e do Planejamento, Orçamento e Gestão. **Protocolo integrado**. Brasília-DF, 2015e. Encontro sobre Protocolo Integrado e Processo Eletrônico Nacional: Novos Paradigmas para a Administração Pública Federal. Brasília-DF, março de 2015. Disponível em: <<http://www.siga.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm>>. Acesso em: 15 maio 2015.

BRASIL. Ministério da Justiça. **Dados abertos**. Brasília, 2015f. Disponível em: <<http://www.justica.gov.br/dados-abertos>>. Acesso em: 10 out. 2015.

BROCHU, Frédéric. La gestion des archives comme ressources informationnelles stratégiques: reconnaître les priorités et établir des alliances. In: **Actes du 5ème Symposium du Groupe de Reserche en Archivistique (GIRA)**, Montréal, 24 mars, 2006.

BUCKLAND, Michael. What is a “digital document”? In: **Preprint of article published in Document Numérique (Paris) 2, no. 2 (1998): 221-230**, [1998]. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/digdoc.html>>. Acesso em: 20 jan. 2015.

BUSINESS DICTIONARY. [201-]. Disponível em: <<http://www.businessdictionary.com/definition/data-management.html#ixzz320MQnKIJ>>. Acesso em: 17 mai. 2014.

CALDERÓN, César; LORENZO, Sebastián (coord.). **Open Government: Gobierno Abierto**. Algón Editores, MMX, abril, 2010. Disponível em: <<http://pt.scribd.com/doc/39496858/Open-Government-Gobierno-Abierto>>. Acesso em: 25 jun. 2014.

CAMARGO, Liriane Soares de Araujo de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Arquitetura da informação para repositórios científicos digitais. In: SAYÃO, Luís et al (Org.) **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. 365 p. il. ISBN 978-85-232-0655-0.

CENTER FOR RESEARCH LIBRARY (CRL); ONLINE COMPUTER LIBRARY CENTER (OCLC). **Trustworthy Repositories Audit & Certification: Criteria and Checklist**. Estados Unidos da América (USA), versão 1.0, february, 2007. Disponível em: <http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf>. Acesso em: 25 jun. 2015.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). Câmara Técnica de Documentos Eletrônicos – CTDE. **Glossário**: versão 5.1, março de 2010. Disponível em: <http://www.documentoseletronicos.arquivonacional.gov.br/media/publicacoes/glossario/2010glossario_v5.1.pdf>. Acesso em: 14 fev. 2015.

_____. Câmara Técnica de Documentos Eletrônicos – CTDE. **e-ARQ Brasil: Modelo de**

Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos, versão 1.1 – Rio de Janeiro: Arquivo Nacional, 2011. (Anexo da Resolução nº 25, de 27 de abril de 2007). CDD 025.04. Disponível em:

<<http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm>>. Acesso em: 13 maio 2015.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). Resolução nº 37, de 19 de dezembro de 2012. Aprova as diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. **Diário Oficial da União**, 2012. Disponível em: <conarq.arquivonacional.gov.br>. Acesso em: 25 maio 2015.

_____. Resolução nº 39, de 29 de abril de 2014. Estabelece diretrizes para a implementação de repositórios digitais confiáveis para a transferência e recolhimento de documentos arquivísticos digitais para instituições arquivísticas dos órgãos e entidades integrantes do Sistema Nacional de Arquivos – SINAR. **Diário Oficial da União**, 2014. Disponível em: <conarq.arquivonacional.gov.br>. Acesso em: 25 set. 2014.

COSTA, Sely Maria de Souza; LEITE, Fernando César Lima. Insumos conceituais e práticos para iniciativas de repositórios institucionais de acesso aberto à informação científica em bibliotecas de pesquisa. In: **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Organizadores: Luís Sayão et al. Salvador: EDUFBA, 2009. 365 p. il.

DE ROBBIO, Antonella. Datti aperti nella Pubblica Amministrazione tra crescita e trasparenza. In: **DigItalia**, 2013. Disponível em: <<http://digitalia.sbn.it/article/view/718>>. Acesso em: 23 jun. 2014.

DIPLOMATIC ANALYSIS TEMPLATE (APPENDIX 7). In: DURANTI, Luciana; PRESTON, Randy. **International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records**. Pádova: Associazione Nazionale Archivistica Italiana (Anai), 2008. Disponível em: <<http://www.interpares.org>>. Acesso em: 15 mar. 2015.

DODDS, Leigh. **Blog Lost Boy**, postado em 09 fevereiro de 2013. Disponível em: <<http://blog.ldodds.com/2013/02/09/what-is-a-dataset>>. Acesso em: 01 maio 2015.

DURANTI, Luciana. Registros documentais contemporâneos. **Revista Estudos Históricos**. Rio de Janeiro: CPDOC/FGV, 1994.

_____. The long-term preservation of accurate and authentic digital data: the InterPares Project. In: **Data Science Journal**, volume 4, 25 october 2005. Disponível em: <https://www.jstage.jst.go.jp/article/dsj/4/0/4_0_106/_pdf>. Acesso em: 28 ago. 2014.

_____. Concepts, principles, and methods for the management of electronic records. **The Information Society**, number 17, p. 271-279, 2001. Disponível em: <<http://210.48.147.73/silibus/conceptprinciples.pdf>>. Acesso em: 1 jan. 2014.

DURANTI, Luciana; THIBODEAU, Ken. The concept of record in interactive, experiential and dynamic environments: the view of InterPARES. In: DURANTI, Luciana; PRESTON, Randy. **International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records**. Pádua: Associazione Nazionale Archivistica Italiana (Anai), 2008. Disponível em: <<http://www.interpares.org>>. Acesso em: 15 mar. 2015.

FISHER, Tony. Making the case for better data. In: **The data asset: how smart companies govern their data for business success**. SAS Institute Inc., Cary, North Carolina, USA, 2009, capítulo 1.

FLORES, Daniel. **Os repositórios arquivísticos digitais**. Santa Maria-RS, [2014]. No prelo.

FONSECA, Maria Odila Kahl. **Arquivologia e ciência da informação**. Rio de Janeiro: Editora FGV, 2007.

FREE SOFTWARE FOUNDATION. **O sistema operacional GNU: O que é o software livre?** Disponível em: <<https://www.gnu.org/philosophy/free-sw.html>>. Acesso em: 10 out. 2015.

HACKETT, Yvette; UNDERWOOD, William; EPPARD, Philip. Part one – Case and general studies in the artistic, scientific and governmental sectors: focus task force report. In: DURANTI, Luciana; PRESTON, Randy. **International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records**. Pádua: Associazione Nazionale Archivistica Italiana (Anai), 2008. Disponível em: <<http://www.interpares.org>>. Acesso em: 15 mar. 2015.

HALL, Wendy; SHADBOLT, Nigel; TIROPANIS, Thanassis; O'HARA, Kieron; DAVIES, Tim. **Open data and charities**. Oxford, GB, Nominet Trust, 87p. 2012 (Nominet Trust State of the Art Reviews). Disponível em: <http://eprints.soton.ac.uk/341346/1/Open_Data_and_Charities.pdf>. Acesso em: 27 jun. 2014.

HEREDIA HERRERA, Antonia. **¿Qué es un archivo?**. Gijón: Ediciones Trea, S. L., Archivos Siglo XXI – 1, 2007.

INDOLFO, Ana Celeste. **Dimensões político-arquivísticas da avaliação de documentos na Administração Pública Federal (2004-2012)**. Tese (Doutorado em Ciência da Informação). – Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Escola de Comunicação, Rio de Janeiro, 2013.

INTERNATIONAL BUSINESS MACHINE. **IBM Terminology**, 2015. Disponível em: <<http://www-01.ibm.com/software/globalization/terminology/d.html>>. Acesso em: 23 jun. 2015.

INTERNATIONAL RECORDS MANAGEMENT TRUST (IRMT). Governo Aberto e documentos arquivísticos confiáveis: instrumento de referência para marcos regulatórios/institucionais e capacidades. Tradução: Vitor Manoel Marques Fonseca. (s.l.), [2013]. Disponível em: <http://www.irmt.org/wp2/wp-content/uploads/2013/11/PORT_Benchmark-_revisto_20_-mar_2014_final.docx>. Acesso em: 5 jul. 2015.

JARDIM, José Maria. O conceito e a prática de gestão de documentos. Rio de Janeiro: **Aervo** – Revista do Arquivo Nacional, v. 2, n.2, jul./dez, 1987. Disponível em: <<http://www.arquivonacional.gov.br/media/v.2,n.2,jul.dez.1987.pdf>>. Acesso em: 1 jan. 2014.

JARDIM, José Maria. A construção do e-gov no Brasil: configurações político-informacionais. In: CIFORM – ENCONTRO DE CIÊNCIA DA INFORMAÇÃO, VI, 2005, Salvador. **Anais...** Salvador: UFBA. Disponível em: <http://www.ciform.ufba.br/v_anais/artigos/jose-mariajardim.html>. Acesso em: 10 dez. 2013.

_____. Em torno de uma política nacional de arquivos: os arquivos estaduais brasileiros na ordem democrática (1988-2011). In: MARIZ, Anna Carla Almeida; JARDIM, José Maria; SILVA, Sérgio Conde de Albite (organizadores). **Novas dimensões da pesquisa e do ensino da arquivologia no Brasil**. Rio de Janeiro: Móbile: Associação dos Arquivistas do Estado do Rio de Janeiro, 2012.

JOHANNESON, Paul. The role of business model in enterprise modelling. In: Krogstie, John; Opdahl, Andreas Lothe; Brinkkemper, Sjaak. **Conceptual Modelling in Information Systems Engineering**. Springer, Berlin Heidelberg, 2007.

MADNICK, Stuart E. et al. Data and Information Quality Research: Its Evolution and Future. In: **Computing Handbook Set**. Chapman and Hall, [2012]. Disponível em: <http://mitiq.mit.edu/Documents/Publications/Papers/2012/Madnick_2012_Data%20and%20Information%20Quality.pdf>. Acesso em: 17 mai. 2014.

MARTINS, Daniel Barbosa. **Extração personalizada e incremental de dados em ambientes de BI em tempo real**. 2009. 112f. Dissertação (Mestrado) - Programa de pós-graduação em Informática, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2009.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data**: a revolution transform how we live, work, and think. Nova Iorque: Houghton Mifflin Harcourt Publishing Company, 2013.

MONTVILOFF, Victor. **Políticas nacionales de información**: Manual para la formulación, aprobación, ejecución y funcionamiento de una política nacional de información. Programa General de Información y UNISIST. Paris: UNESCO, 1990. Disponível em: <<http://unesdoc.unesco.org/images/0008/000869/086995sb.pdf>>. Acesso em: 22 set. 2014.

MOSLEY, Mark. **DAMA-DMBOK Guide**: functional framework, Version 3.02, september 10, 2008. The data Management Association – DAMA INTERNATIONAL. Disponível em: <<http://www.dama.org/content/body-knowledge>>. Acesso em: 12 ago. 2014.

NESMITH, Tom. Conhecimento e educação para a sociedade dos arquivos. **Arquivo & Administração**, Rio de Janeiro: AAB, v.9, n.2, jul./dez. 2010, p. 15-30.

OPEN DATA INSTITUTE (ODI). What is open data? In: **Guide**. Disponível em: <<http://theodi.org/guides/what-open-data>>. Acesso em: 06 dez. 2013.

OPEN DEFINITION. **The definition**. Disponível em: <<http://opendefinition.org/od/>>. Acesso em: 25 mar. 2015.

OPEN GOVERNMENT PARTNERSHIP. **About**. United States: San Francisco, CA, 2013. Disponível em: <<http://www.opengovpartnership.org>>. Acesso em: 29 nov. 2013.

OPEN KNOWLEDGE FOUNDATION (OKF). **Open data handbook documentation**. (Release 1.0.0). Novembro, 2012. Disponível em: <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>. Acesso em: 25 jun. 2014.

PEARSE-MOSES, Richard. **A Glossary of Archival and Records Terminology**. Chicago: The Society of American Archivists, 2005. Disponível em: <<http://www2.archivists.org/glossary/terms/d/data-set>>. Acesso em: 01 maio 2015.

PROJETO INTERPARES 2. **Diretrizes do produtor** – a elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. Tradução e revisão: Arquivo Nacional; Câmara dos Deputados. Brasília, DF: Câmara dos Deputados, 2010. Disponível em: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet--portuguese.pdf>. Acesso em: 13 mai. 2014.

PROJETO INTERPARES 3. Team Brasil. **Estudo de caso BR07** – Diretoria geral de Recursos Humanos da Universidade Estadual de Campinas (Unicamp) – Demonstrativo de pagamento de pessoal: relatório final. (s.l.), versão 1.3, 2012. Disponível em: <http://www.siarq.unicamp.br/siarq/images/siarq/pesquisa/pdf/relatorio_interpares_unicamp_2013.pdf>. Acesso em: 26 jan. 2015.

RAYWARD, W. Boyd. The Origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). **Journal Of The American Society For Information Science**. Nº 48(4):289–300, 1997. Disponível em: <<http://people.lis.illinois.edu/~wrayward/OriginsofInfoScience.pdf>>. Acesso em: 20 jan. 2015.

RÊGO, Bergson Lopes. **Gestão e governança de dados**: promovendo dados como ativo de valor nas empresas. Rio de Janeiro: Brasport, 2013.

RENEAR, Allen H.; SACCHI, Simone; WICKETT, Karen M. Definitions of *Dataset* in the Scientific and Technical Literature. **ASIST 2010**, October 22–27, 2010, Pittsburgh, PA, USA, 2010. Disponível em: <https://www.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/240_Final_Submission.pdf>. Acesso em: 01 maio 2015.

ROCHA, Claudia Lacombe. Gestão de documentos digitais: as ações do Arquivo Nacional e da Câmara Técnica de Documentos eletrônicos – Conarq. In: NEGREIROS, Leandro Ribeiro; NEVES, Marta Eloísa Melgaço (Org.). **Documentos eletrônicos: fundamentos teóricos arquivísticos para uma prática de pesquisa em gestão e preservação**. Belo Horizonte: Secretaria de Estado de Cultura de Minas Gerais, Arquivo Público Mineiro, 2008.

RONDINELLI, Rosely Curi. **O documento arquivístico ante a realidade digital: uma revisão conceitual necessária**. Rio de Janeiro: Editora FGV, 2013.

ROUSSEAU, Jean-Yves; COUTURE, Carol. **Os fundamentos da disciplina arquivística**. Lisboa: Publicações Dom Quixote. Tradução: Magda Bigotte de Figueiredo, 1998.

SADIQ, Shazia; YEGANEH, Naiem Khodabandehloo; INDULSKA, Marta. 20 Years of Data Quality Research: Themes, Trends and Synergies. Australasian Database Conference. 22nd (ADC 2011), Perth, Australia, January 2011. In: **Conferences in Research and Practice in Information Technology (CRPIT)**, Vol. 115, Heng Tao Shen and Yanchun Zhang, Ed. Disponível em:

<http://www.researchgate.net/publication/49515275_20_years_of_data_quality_research_The_mes_trends_and_synergies/file/3deec5179ddf947a29.pdf>. Acesso em: 26 mai.2014.

SANTOS, Henrique Machado dos; FLORES, Daniel. Repositórios digitais confiáveis para documentos arquivísticos: ponderações para a preservação em longo prazo. **Perspectivas em ciência da Informação**, v. 20, n. 2, p. 198-218, abr./jun. 2015. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2341/1604>>. Acesso em: 14 jul. 2015.

SETZER, Valdemar W. Dado, Informação, Conhecimento e Competência. In: **DataGramZero**. Revista de Ciência da Informação, n. zero, dez/99, artigo 01. Disponível em: <http://datagramazero.org.br/dez99/Art_01.htm>. Acesso em: 25 mai. 2014.

SILVA, Armando Malheiro da; MARTINS, Neire Rossio. Estudo sobre o sistema de arquivos da Universidade Estadual de Campinas. **Arquivística.net**, 2007. Disponível em <<http://www.arquivistica.net/ojs/viewissue.php?id=7#Artigo>>. Acesso em: 19 jan. 2015.

SILVA, Armando Malheiro da. [et al]. **Arquivística – teoria e prática de uma ciência da informação**. Porto: Afrontamento, 3ª edição, 2009.

SUBER, Peter. **Open access overview: focusing on open access peer-reviewed research articles and their preprints**. *Blog* revisado, 2013. Disponível em: <legacy.earlham.edu/~peters/fos/overview.htm>. Acesso em: 5 abr. 2015.

_____. Timeline of Open Access. History of Open Access. In: **Open access directory OAD**. Página WIKI, 2015. Disponível em: <oad.simmons.edu/oadwiki/Main_Page>. Acesso em: 5 abr. 2015.

THE WHITE HOUSE. Office of the Press Secretary. **Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government**, 2009. Disponível em: <http://nsarchive.gwu.edu/news/20090121/2009_transparency_memo.pdf>. Acesso em: 31 maio 2015.

TOGNOLI, Natália Bolfarini. **A contribuição epistemológica canadense para a construção da arquivística contemporânea**. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2010, CDD 025.171.

VAZ, José Carlos. A delicada disputa pelas cabeças na evolução do serviço público no Brasil pós-democratização. **Cadernos ENAP**, edição especial, 2006.

VIEIRA, Vânia Lúcia Ribeiro. Transparência e acesso à informação no âmbito do Governo Federal brasileiro. In: RESENDE, André (org.). **Governo brasileiro no futuro: sugestões e desafios para o Estado (2012-2022)**. São Paulo: Cubzac, 2012.

WANG, Richard Y; STRONG, Diane M. Beyond accuracy: What data quality means to data consumers. **Journal of Management Information Systems**. vol. 12, número 4 (Spring 1996), p.5-34. Disponível em: <http://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf>. Acesso em: 7 abr. 2014.

WORLD WIDE WEB CONSORTIUM – W3C (Escritório Brasil) et al. **Manual dos dados abertos: governo**. Traduzido e adaptado de opendatamanual.org. São Paulo: Comitê gestor da Internet no Brasil, 2011. [cooperação técnica científica entre o Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC.br)].

_____. **Dados Abertos Governamentais**. Brasília, 2012. Disponível em: <<http://www.w3c.br/divulgacao/pdf/dados-abertos-governamentais.pdf>>. Acesso em: 23 abr. 2015.

ZINS, Chaim. Conceptual approaches for defining data, information and knowledge. **Journal of the American Society for Information Science and Technology**. JASIST, v. 58, n. 4, p. 479-493, 2007. (Wiley Periodicals, Inc.; Wiley InterScience; DOI: 10.1002/asi.20508).

ANEXO A – INTERPARES PROJECT 2: DIPLOMATIC ANALYSIS TEMPLATE

Diplomatic Analysis

CS[##]

[Title of Case Study] Case Study

INTRODUCTION

Paragraph 1 (*Introduction to the case study*) - The InterPARES case study # ...

Paragraph 2 (*Introduction to the case study subject*) - The ...
project/program/system/database ...

Paragraph 3 (*Introduction to the diplomatic analysis, including the indication of the subject of the diplomatic analysis, i.e., the digital entity in the report that needs to be analyzed*) - The following text presents the results of the diplomatic analysis on the digital entity identified in the case study report. The purpose of the diplomatic analysis is to assess the status of the identified digital entity as record, and based on the analysis, Domain 3 of InterPARES 2 could propose applicable preservation strategies. The digital entity identified in the case study report is ... This diplomatic analysis therefore centers on the identification of the [digital entity] as record.

IDENTIFICATION OF RECORD(S)

The current version of the InterPARES glossary definition of record defines a record¹ as “a document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference.”² This definition implies that, to be considered as a record, a digital entity must comprise five indispensable elements: fixed content and form, embedded action, archival bond, persons and contexts, the establishment of which is based on diplomatic analysis, archival science and findings from InterPARES. The application of the definition on the [digital entity] is therefore analyzed accordingly:

1. TO BE IDENTIFIED AS A RECORD, THE DIGITAL ENTITY MUST POSSESS FIXED CONTENT AND FORM,³ AND BE AFFIXED TO A STABLE MEDIUM (OR PHYSICAL CARRIER).

- The content of the [digital entity] is fixed or not and why
- The documentary form⁴ of the [digital entity] is fixed or not and why

¹ Current version of the definition of record as assessing criterion, which starts the process of analyzing the 5 components.

² Glossary definitions, in Terminology Database, accessible through the InterPARES website research restricted area.

³ The InterPARES1 Authenticity Task Force has defined fixed form as the following: 1) binary content of the record, including indicators of documentary form, must be stored in a manner that ensures it remains complete and unaltered, and 2) technology must be maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when set aside. (See ATF Research Methodology Statement, available at: http://www.interpares.org/documents/interpares_ResearchMethodologyStatement.pdf).

⁴ Definition of documentary form from Glossary Definitions, Terminology Database, InterPARES Web site: The rules of representation according to which the content of a record, its administrative and documentary

- The digital entity is fixed to a stable medium or not and why

2. A RECORD MUST ALSO PARTICIPATE IN AN ACTION, DEFINED AS THE CONSCIOUS EXERCISE OF WILL BY THE AUTHOR OR BY AN EXTERNAL PERSON, AIMED TO CREATE, MAINTAIN, MODIFY OR EXTINGUISH SITUATIONS. A RECORD RESULTS AS A NATURAL BY-PRODUCT OF THE ACTION.

Statement of the name of action that generated the [digital entity] and how it participates in the action.

3. THIRD, A RECORD MUST POSSESS AN ARCHIVAL BOND, WHICH IS THE RELATIONSHIP THAT LINKS EACH RECORD TO THE PREVIOUS AND SUBSEQUENT RECORD OF THE SAME ACTION AND, INCREMENTALLY, TO ALL THE RECORDS WHICH PARTICIPATE IN THE SAME ACTIVITY. THE ARCHIVAL BOND IS ORIGINARY (I.E., IT COMES INTO EXISTENCE WHEN THE RECORD IS MADE OR RECEIVED AND SET ASIDE), NECESSARY (I.E., IT EXISTS FOR EVERY RECORD), AND DETERMINED (I.E., IT IS CHARACTERISED BY THE PURPOSE OF THE RECORD).

The [digital entity] possesses (or does not possess) archival bond and why.

4. FOURTH, RECORD CREATION MUST INVOLVE AT LEAST THREE PERSONS, WHETHER OR NOT THEY EXPLICITLY APPEAR IN THE RECORD ITSELF. THESE PERSONS ARE THE AUTHOR, ADDRESSEE AND WRITER; IN THE ELECTRONIC ENVIRONMENT, ONE MUST ALSO TAKE INTO ACCOUNT TWO ADDITIONAL NECESSARY PERSONS: THE CREATOR AND THE ORIGINATOR.

- The record's author is the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

Name of the author and why.

- The writer is the physical or juridical person having the authority and capacity to articulate the content of the record.

Name of the writer and why.

- The addressee is the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

Name of the address and why.

- The creator is the person in whose fonds the record exists.

Name of the creator and why.

- The originator is the person to whom the Internet account issuing or the server holding the record belongs.

Name of the originator and why.

5. FINALLY, A RECORD MUST POSSESS AN IDENTIFIABLE CONTEXT, DEFINED AS THE FRAMEWORK IN WHICH THE ACTION IN WHICH THE RECORD PARTICIPATES TAKES PLACE. THE TYPES OF CONTEXT INCLUDE

context, and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements.

JURIDICAL-ADMINISTRATIVE, PROVENANCIAL, PROCEDURAL, DOCUMENTARY, AND TECHNOLOGICAL.

- The juridical-administrative context is the legal and organizational system in which the creating body belongs.

Juridical context:

Administrative context:

- The provenancial context refers to the creating body, its mandate, structure and functions.

Provenancial context:

- The procedural context comprises of the business procedure in the course of which the record is created.

Procedures:

Diplomatic analysis of Procedural Phases in the Creation of the [digital entity]

a) **Initiative**: The introductory phase of any procedure is “constituted by those acts, written and/ or oral, which start the mechanism of the procedure.”⁵

b) **Inquiry**: This preliminary phase “is constituted by the collection of the elements necessary to evaluate the situation.”⁶

c) **Consultation**: This phase is “constituted by the collection of opinions and advice after all the relevant data has been assembled.”⁷

d) **Deliberation**: This phase is “constituted by the final decision-making.”⁸

e) **Deliberation control**: This phase is “constituted by the control exercised by a physical or juridical person different from the author of the document embodying the transaction, on the substance of the deliberation and / or on its forms.”⁹

f) **Execution**: “The documents created in this phase are the originals of those embodying the transactions.”¹⁰ In other words, the execution phase results in the issuing of the first record capable of producing the consequences intended by its author.

- The documentary context is defined as the archival fonds to which a record belongs and its internal structure.

Documentary context:

- The technological context is defined as the characteristics of the technological components of an electronic computing system in which records are created.

Technological context:

5 Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Maryland and London: The Scarecrow Press in association with the Society of American Archivists and the Association of Canadian Archivists, 1998), 115.

6 Ibid.

7 Ibid.

8 Ibid.

9 Ibid.

10 Ibid., 116.

CONCLUSIONS

Overall conclusions of the status of the digital entity under examination:

If it is not a record:

Summary of the digital entity as a publication:

Summary of the digital entity as a potential record:

ANEXO B – INTERPARES PROJECT 2: BENCHMARK REQUIREMENTS SUPPORTING THE PRESUMPTION OF AUTHENTICITY OF ELECTRONIC RECORDS

Benchmark excerpted from: Authenticity Task Force (2002), “Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records,” in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Luciana Duranti, ed. (San Miniato, Italy: Archilab, 2005), 204-219. Online reprint available at http://www.interpares.org/book/interpares_book_k_app02.pdf.

Preamble

The benchmark requirements are the conditions that serve as a basis for the preserver’s assessment of the authenticity of the creator’s electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record’s authenticity on the basis of the manner in which the records have been created, handled and maintained by the creator.

Within the benchmark requirements, Requirement A.1 identifies the core information about an electronic record – the immediate context of its creation and the manner in which it has been handled and maintained – that establishes the record’s identity and lays a foundation for demonstrating its integrity. Requirements A.2–A.8 identify the kinds of procedural controls over the record’s creation, handling and maintenance that support a presumption of the record’s integrity.

Benchmark Requirements (Requirement Set A)

To support a presumption of authenticity the preserver must obtain evidence that:

<p>REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record</p>	<p>the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.</p>
<p>A.1.a Identity of the record: A.1.a.i Names of the persons concurring in the formation of the record, that is:</p> <ul style="list-style-type: none"> • name of author¹ • name of writer² (if different from the author) • name of originator³ (if different from name of author or writer) • name of addressee⁴ 	

-
- 1 The name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.
 - 2 The name of the physical or juridical person having the authority and capacity to articulate the content of the record.
 - 3 The name of the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.
 - 4 The name of the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

<p>A.1.a.ii Name of action or matter</p> <p>A.1.a.iii Date(s) of creation and transmission, that is:</p> <ul style="list-style-type: none"> • chronological date⁵ • received date⁶ • archival date⁷ • transmission date(s)⁸ <p>A.1.a.iv Expression of archival bond⁹ (e.g., classification code, file identifier)</p> <p>A.1.a.v Indication of attachments</p> <p>A.1.b Integrity of the record:</p> <p>A.1.b.i Name of handling office¹⁰</p> <p>A.1.b.ii Name of office of primary responsibility¹¹ (if different from handling office)</p> <p>A.1.b.iii Indication of types of annotations added to the record¹²</p> <p>A.1.b.iv Indication of technical modifications;¹³</p>
--

<p>REQUIREMENT A.2: Access Privileges</p>	<p>the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;</p>
---	---

<p>REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records</p>	<p>the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;</p>
---	--

-
- 5 The date, and possibly the time, of compilation of a record included in the record by the author or the electronic system on the author's behalf.
 - 6 The date, and possibly the time, when a record is received by the addressee.
 - 7 The date, and possibly the time, when a record is officially incorporated into the creator's records.
 - 8 The date and time when a record leaves the space in which it was generated.
 - 9 The archival bond is the relationship that links each record, incrementally, to the previous and subsequent ones and to all those [that] participate in the same activity. It is ordinary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).
 - 10 The office (or officer) formally competent for carrying out the action to which the record relates or for the matter to which the record pertains.
 - 11 The office (or officer) given the formal competence for maintaining the authoritative record, that is, the record considered by the creator to be its official record.
 - 12 Annotations are additions made to a record after it has been completed. Therefore, they are not considered elements of the record's documentary form.
 - 13 Technical modifications are any changes in the digital components of the record as defined by the Preservation Task Force. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods (software) applied to reproduce the record from the stored digital components; that is, any changes that might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.

REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;
REQUIREMENT A.5: Establishment of Documentary Forms	the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;
REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;
REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;
REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

ANEXO C – INTERNATIONAL RECORDS MANAGEMENT TRUST: GOVERNO ABERTO E DOCUMENTOS ARQUIVÍSTICOS CONFIÁVEIS – INSTRUMENTO DE REFERÊNCIA PARA MARCOS REGULATÓRIOS/INSTITUCIONAIS E CAPACIDADES.



International Records Management Trust

Governo Aberto e documentos arquivísticos confiáveis Instrumento de referência para marcos regulatórios/institucionais e capacidades

Documentos arquivísticos públicos confiáveis e acessíveis são um meio de se demonstrar transparência e responsabilização; eles constituem a base legal sobre a qual o governo aberto é construído. Muitos crêm que documentos arquivísticos confiáveis sejam disponibilizados para apoiar iniciativas que fortalecem essa abertura, mas, com frequência, não é esse o caso. Quando documentos arquivísticos não podem ser acessados, um governo realmente aberto é inalcançável.

Um bem sucedido Governo Aberto, que inclui dados abertos e acesso à informação, fundamenta-se na evidência das decisões, ações e transações governamentais que derivam, em grande parte, de documentos oficiais. Uma boa gestão de documentos assegura que documentos acurados e confiáveis sejam produzidos e permaneçam acessíveis, utilizáveis e autênticos pelo tempo que for necessário, de forma a fundamentarem a melhoria dos serviços, o controle da corrupção e o fortalecimento da democracia. Quando os documentos arquivísticos são confiáveis, os dados abertos e o acesso à informação tornam-se meios poderosos para assegurar a transparência governamental e possibilitar que os cidadãos assumam e participem mais integralmente de seus governos. Os dados podem ser relacionados a fontes acuradas de evidência, e é possível encontrar e confiar na informação requerida.

No entanto, em muitos países, documentos arquivísticos públicos não são gerenciados de acordo com padrões internacionais e, em alguns casos, nem mesmo procedimentos básicos de controle de gestão estão disponíveis, especialmente onde a adoção e o uso de tecnologias digitais ultrapassaram a capacidade dos governos de gerenciar documentos arquivísticos digitais. Se os governos devem ser transparentes e passíveis de responsabilização e os cidadãos devem se envolver com seus governos de maneira significativa, então o acesso contínuo a documentos arquivísticos confiáveis, acurados e autênticos é essencial.

Este instrumento fornece parâmetros de boas práticas em gestão de documentos que os governos e as organizações podem utilizar para identificar áreas que necessitam de aperfeiçoamento para abertura de dados e responsabilização. São estabelecidos marcos de referência para as áreas-macro das estruturas institucionais/ regulatórias (leis, políticas, normas e infraestrutura) e de capacidades (pessoal, educação e treinamento) em três estágios: básico, intermediário e avançado¹.

¹ Para esta pesquisa foi adotado o Estágio 1: básico.

Estágio 1: Básico**Marcos regulatórios/ institucionais**

<i>Recomendação</i>	<i>Meios de verificação</i>
1. A autoridade arquivística (uma instituição arquivística com abrangência nacional, estadual ou local, com autoridade equivalente) tem poder de assessorar o governo quanto à política, estabelece normas e define controles de qualidade para a gestão de documentos públicos em todos os formatos.	Legislação arquivística; publicação das competências e políticas
2. A autoridade arquivística está administrativamente posicionada de forma a cumprir sua missão no que diz respeito a todas as áreas e funções do setor público.	Competência da autoridade arquivística
3. O governo estimula uma cultura de acurácia de dados e de acesso do cidadão à informação, baseada em fontes documentais confiáveis e íntegras, em conformidade com leis, regulamentos e normas.	Programas e declarações de políticas governamentais
4. O governo reconhece a interdependência da gestão de documentos/ informações e a abertura de dados e a lei de acesso à informação. Há um diálogo contínuo entre a autoridade arquivística e as autoridades de Governo Aberto/ governo eletrônico/ acesso à informação e Tecnologias de Informação e Comunicação (TIC), no tocante a iniciativas acerca das informações governamentais.	Programas e declarações de políticas do governo; documentos de reuniões da autoridade arquivística e as autoridades de Governo Aberto/ governo eletrônico/ acesso à informação e TIC
5. Uma política governamental ampla de gestão de documentos (ou código de prática equivalente) atribui responsabilização e define as relações entre a autoridade arquivística e os organismos responsáveis por Governo Aberto / governo eletrônico / acesso à informação e TIC; há um cronograma para a implementação e revisão periódica da política de gestão de documentos.	Declarações de políticas e programas do governo e cronogramas de implementação
6. A autoridade arquivística assume um papel de liderança junto ao governo na garantia de que os documentos arquivísticos estarão disponíveis para fornecer informações íntegras e confiáveis ao público, e no desenvolvimento de normas de gestão de documentos; ela coopera com autoridades em dados abertos / acesso à informação, orientando órgãos do governo sobre aprimoramento do acesso aos documentos e sobre as condições em que controles legítimos de acesso podem ser implementados, sem comprometer os objetivos gerais do Governo Aberto e do acesso à informação.	Programas, declarações publicadas e relatórios de auditoria da autoridade arquivística; declarações de políticas do governo

7. A autoridade arquivística trabalha com autoridades de Governo Aberto / acesso à informação para identificar formas de aperfeiçoar o acesso, tais como eliminação ou redução do impacto do período legal de restrição ao acesso a documentos públicos.	Documentos de reuniões entre a autoridade arquivística e autoridades em Governo Aberto/ acesso; declarações e orientações públicas conjuntas.
8. A Lei de Acesso à Informação, a legislação arquivística e outros atos legais relacionados com a liberação de informações do governo são compatíveis.	Lei de Acesso à Informação, legislação arquivística e quaisquer outros atos legais relacionados com a liberação de informações
9. Caso ainda não tenha sido implementada, há cronograma para a entrada em vigor da Lei de Acesso à Informação; a autoridade arquivística está diretamente envolvida nas atividades de implementação, tais como treinamento e normalização.	Programas e declarações de políticas do governo; programas e declarações públicas da autoridade arquivística
10. A autoridade arquivística adota uma norma de gestão de documentos, como a ISO 15489, e uma norma para a funcionalidade da gestão de documentos em sistemas de TIC, como ICA-Req ² .	Declarações públicas da autoridade arquivística; política de gestão de documentos
11. A autoridade arquivística segue um cronograma para a produção e atualização de normas e orientações voltadas para a gestão dos principais tipos de documentos públicos, em papel ou digitais.	Planejamentos e relatórios publicados pela autoridade arquivística
12. A autoridade arquivística segue um cronograma para a produção e atualização de tabelas de temporalidade para os documentos públicos, em papel ou digitais.	Planejamentos e relatórios publicados pela autoridade arquivística; política de gestão de documentos
13. A autoridade arquivística implementa um programa de inspeção de gestão de documentos em órgãos públicos; ela efetua inspeções regulares e possui autoridade para exigir a observância de normas e políticas de gestão de documentos (p. ex. regras para a destinação de documentos)	Relatórios de inspeção e documentos de acompanhamento
14. A Lei de Acesso à Informação possibilita acesso a documentos públicos e se sobrepõe a restrições de acesso pré-existentes, como o período regulamentado de restrição de acesso. Este requisito é divulgado publicamente.	Lei de Acesso à Informação e declarações públicas

2 N. T.: No Brasil, o Conselho Nacional de Arquivos recomenda o uso do e-ARQ Brasil, que é um documento semelhante ao ICA-Req, porém ajustado às práticas e normas arquivísticas brasileiras.

Capacidades

<i>Recomendação</i>	<i>Meios de verificação</i>
1. A autoridade arquivística tem, como uma de suas principais funções, um programa de conscientização e treinamento em gestão de documentos, para o setor público.	Material didático da autoridade arquivística
2. A autoridade arquivística oferece, como uma de suas principais funções, um programa de conscientização e treinamento para todos os funcionários envolvidos na produção e guarda de documentos públicos.	Material didático da autoridade arquivística
3. A autoridade arquivística fornece, como uma de suas principais funções, um programa de conscientização das autoridades governamentais sobre a importância fundamental da gestão de documentos no setor público, bem como sobre suas responsabilidades de guarda desses documentos como bens públicos.	Material didático da autoridade arquivística
4. Universidades ou outras instituições de ensino nacionais oferecem cursos de graduação e pós-graduação em gestão de documentos e arquivos permanentes, com, no mínimo, uma introdução à gestão de documentos arquivísticos digitais e ao relacionamento entre a gestão de documentos e dados abertos/ acesso à informação e TIC/ governo eletrônico.	Folhetos e descrições de cursos
5. Pelo menos, 50% dos funcionários na área-fim da autoridade arquivística, concluíram cursos de pós-graduação em gestão de documentos e arquivos permanentes.	Certificados acadêmicos
6. Todo órgão público tem uma unidade dedicada à gestão de documentos e informação, com a função de orientar quanto à gestão do ciclo vital dos documentos (digitais ou em papel), estabelecer normas e práticas e avaliar seu cumprimento.	Estruturas organizacionais, materiais de treinamento; reuniões com as unidades de gestão de documentos
7. Há um programa de treinamento em boas práticas de gestão de documentos, para as equipes das unidades de gestão de documentos e informação.	Material de treinamento da autoridade arquivística ou de órgãos públicos
8. Funcionários da autoridade arquivística e das unidades de gestão de documentos e informação estão familiarizados e têm experiência no uso de uma ampla variedade de sistemas informatizados (por exemplo, aplicativos em rede, sistemas corporativos, programas para escritório), e em sistemas e tecnologias	Verificação do acesso dos funcionários a sistemas de gestão de documentos digitais e sistemas corporativos; currículos

destinadas a gerenciar a integridade e a fidedignidade de documentos ao longo do seu ciclo vital.	
9. A autoridade arquivística oferece treinamento e orientação a órgãos públicos na avaliação para acesso a documentos e informações, e na preparação e atualização de declarações quanto ao acesso à informação (categorias de documentos e informações acessíveis ao público).	Materiais de treinamento e orientação da autoridade arquivística