

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
ESPECIALIZAÇÃO EM INFORMÁTICA: ÁREA DE CONCENTRAÇÃO: GESTÃO DE
TECNOLOGIA DA INFORMAÇÃO

Milton Pereira de Souza

**Mecanismos para Viabilizar Transparência Ativa e Abertura de Dados Adequados à
Governança de Informações Digitais da ANAC**

Brasília
2019

MILTON PEREIRA DE SOUZA

MECANISMOS PARA VIABILIZAR TRANSPARÊNCIA ATIVA E ABERTURA DE DADOS ADEQUADOS À GOVERNANÇA DE INFORMAÇÕES DIGITAIS DA ANAC

Monografia apresentada ao Curso de Especialização em Informática do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do grau de Especialista em Informática.

Área de Concentração: Gestão de Tecnologia da Informação

Orientador: Flávio Vinícius Diniz de Figueiredo

Brasília-DF
2019

© Milton Pereira de Souza
Todos os direitos reservados

Ficha catalográfica elaborada pela Biblioteca do ICEX - UFMG

Souza, Milton Pereira de

S729m Mecanismos para viabilizar transparência ativa e abertura de dados adequados à governança de informações digitais da ANAC / Milton Pereira de Souza– Brasília, 2019.
x, 121 f. : il.

Monografia (especialização) – Universidade Federal de Minas Gerais. Departamento de Ciência da Computação.

Orientador: Flávio Vinícius Diniz de Figueiredo

1. Computação – Monografias. 2. Informações digitais.
3. Governança. 4. Gestão. 5. Transparência. 6. Agência Nacional de Aviação Civil. I. Orientador. II. Título.

CDU 519.6*



UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
ESPECIALIZAÇÃO EM INFORMÁTICA: ÁREA DE CONCENTRAÇÃO GESTÃO EM
TECNOLOGIAS DA INFORMAÇÃO

Mecanismos para viabilizar transparência ativa e abertura de dados adequados à
governança de informações digitais da ANAC

MILTON PEREIRA DE SOUZA

Monografia apresentada aos Senhores:

Prof. Flávio Vinícius Diniz de Figueiredo
Orientador
DCC - ICEx – UFMG

Prof. José Nagib/Cotrim Árabe
DCC - ICEx - UFMG

Prof. José Marcos Silva Nogueira
DCC - ICEx - UFMG

Belo Horizonte, 14 de março de 2019

RESUMO

Este trabalho apresenta mecanismos implementados na Agência Nacional de Aviação Civil, complementares e dependentes entre si, que vêm transformando uma determinação normativa (que, de acordo com o Decreto, basta a ação de abertura de dados de forma reativa para estar em conformidade), em um ecossistema de governança e gestão para abertura e transparência de dados dentro de um contexto de informações digitais na Agência, que entrega à sociedade um ferramental para controle social, e com potencial para fomentar a pesquisa científica, promover o desenvolvimento tecnológico e a inovação. Para tanto, apresentou a evolução histórica de temas relacionados a transparência e abertura de dados e sua importância para o fortalecimento das democracias.

Palavras chaves: Governança; Gestão; Transparência ativa; Abertura de dados; Informações digitais ; Agência Nacional de Aviação Civil

ABSTRACT

This work presents mechanisms implemented in the National Civil Aviation Agency, complementary and dependent on each other, which have been transforming a normative determination (which, according to the act, it is sufficient to open data in a reactive way to be in compliance), in a governance and management ecosystem for openness and transparency of data within a context of digital information in the Agency, which provides society with a tool for social control, with the potential to foster scientific research, promote technological development and innovation. For that, it presented the historical evolution of issues related to transparency and data openness and its importance for the strengthening of democracies.

Keywords: Governance; Management; Active transparency; Opening of data; Digital information; National Civil Aviation Agency

LISTA DE FIGURAS

Figura 1 - Evolução da duração dos Termos de copyright nos Estados Unidos	22
Figura 2 - Diferentes níveis de governança “gravitando” simultaneamente	35
Figura 3 - Relação entre governança e gestão	38
Figura 4 - Sistema de Governança em órgãos e entidades da administração pública	39
Figura 5 - Frequência de problemas encontrados no Brasil	44
Figura 6 - taxonomia de modelos de recuperação de informação	55
Figura 7 - evolução do patrimônio de candidatos	62
Figura 8 - imagens na página principal do portal Tesouro Transparente	63
Figura 9 - diagrama: “como são aplicadas as receitas federais?”	64
Figura 10 - Grid de visualização dos dados de despesas com o sistema integrado de Administração.....	64
Figura 11 - Imagem do Balanço Patrimonial do BSPN 2017, exemplificativa de disponibilização de acesso ao código de geração do gráfico.....	65
Figura 12 - imagem representativa dos mecanismos para viabilizar transparência ativa e abertura de dados na ANAC.....	68
Figura 13 - imagem representativa do mecanismo 1 para viabilizar transparência ativa e abertura de dados na ANAC.....	68
Figura 14 - avaliação de maturidade dos processos da ANAC com relação a governança de dados e informações	75
Figura 15 - modelo esquemático visual da Política de Governança de Informações Digitais	80
Figura 16 - imagem representativa do mecanismo 2 para viabilizar transparência ativa e abertura de dados na ANAC.....	81
Figura 17 - imagem representativa do ciclo do processo para abertura dos dados da ANAC	84
Figura 18 - Estrutura de governança do Plano de Dados Abertos da ANAC.....	91
Figura 19 - imagem representativa do mecanismo 3 para viabilizar transparência ativa e abertura de dados na ANAC.....	92
Figura 20 - imagem representativa do fluxo de atividades para publicação de dados e transparência.....	96
Figura 21 - Fluxo de Projeto de Desenvolvimento e Manutenção ajustado.....	98
Figura 22 - imagem representativa do mecanismo 4 para viabilizar transparência ativa e abertura de dados na ANAC.....	101
Figura 23 – interação entre produtos da suíte ELK.....	105

Figura 24 - gráfico do tipo “heat map” como resultado do piloto.....	107
Figura 25 - gráfico "heat map" incorporado em dashboard construído no piloto	108

LISTA DE SIGLAS

AGA - Association of Government Accountants
AGID - Área de Governança de Informações Digitais
ANAC - Agência Nacional de Aviação Civil
BI - Business Intelligence
BSPN - Balanço do Setor Público Nacional
CKAN - Comprehensive Knowledge Archive Network
CMA - Compilado de Movimentação Aeroportuária
DAC - Departamento de Aviação Civil
DIGM - Data and Information Governance Maturity
DMBOK - Data Management Body of Knowledge
DMM - Data Management Maturity
EGD - Estratégia de Governança Digital
ELK - Elasticsearch, Logstash e Kibana
ERIC - Centro de Informações sobre Recursos Educacionais
INDA - Infraestrutura Nacional de Dados Abertos
INDE - Infraestrutura Nacional de Dados Espaciais
ISO/IEC- International Standard Organization
OGP - Open Government Partnership
PDA - Plano de Dados Abertos
PGID - Política de Governança de Informações Digitais
SAF - Superintendência de Administração e Finanças
STI - Superintendência de Tecnologia da Informação
STN – Secretaria do Tesouro Nacional
TCU - Tribunal de Contas da União
TI - Tecnologia da Informação
TIC - Tecnologia da Informação e Comunicação
WMO - World Meteorological Organization

SUMÁRIO

1	INTRODUÇÃO	9
2	JUSTIFICATIVA.....	11
3	REFERENCIAL TEÓRICO	13
3.1	Dados, informação e conhecimento como bens públicos:.....	15
3.2	Mas e os Direitos Autorais?.....	19
3.3	Transparência e Dados abertos.....	23
3.3.1	Os Princípios dos dados abertos – encontro de Sebastopol	24
3.4	Transparência e Dados Abertos Governamentais no Brasil	25
3.5	Governo aberto, dados abertos governamentais e os compromissos com a OGP	28
3.6	Compromisso do Governo Brasileiro com a OGP	30
3.7	Dados e Informações no contexto das Agências reguladoras	32
3.8	Governança de dados e informações no setor público	33
3.9	Reflexão sobre dados abertos.....	39
3.10	Tecnologias disruptivas para uso com dados abertos	48
3.10.1	Internet das Coisas e cidades inteligentes	48
3.10.2	Big Data.....	51
3.10.3	Data Lake.....	53
3.10.4	Sistemas de recuperação, recomendação e relevância de informação.....	54
3.10.5	Conclusão	58
3.11	Experiências de sucesso no uso de dados abertos com tecnologias disruptivas	58
3.11.1	O caso Big Data do Governo Federal: GovData.....	59
3.11.2	O caso “Operação Serenata de Amor”	60
3.11.3	O caso “Portal Tesouro Transparente”	62
4	MECANISMOS PARA VIABILIZAR TRANSPARÊNCIA ATIVA E ABERTURA DE DADOS NA ANAC	66
4.1	Mecanismo 1 - Implantação de um modelo de Governança de Informações digitais na ANAC	68
4.1.1	Projeto “Governança de Informações Digitais”	70
4.1.2	Política de Governança de Informações Digitais da ANAC.....	78
4.2	Mecanismo 2 - Institucionalização de um processo para gestão de Plano de Abertura de dados	81
4.2.1	O Processo para abertura de dados	82
4.3	Mecanismo 3 - Ativos de processos da Agência (normas, diretrizes e processos) para abertura de dados	92
4.3.1	Captação e recebimento de dados.....	93
4.3.2	Dados gerados a partir de sistemas de gestão.....	96
4.3.3	Divulgação de dados e informações.....	99
4.4	Mecanismo 4 – Desenvolvimento de plataforma para publicação e pesquisa de dados abertos, análise e publicação de informações com a suíte ELK.....	101
4.4.1	Realização de piloto com a suíte ELK	106
4.4.2	Conclusão	108
5	CONCLUSÃO E TRABALHOS FUTUROS	109
6	REFERÊNCIAS.....	112

1 INTRODUÇÃO

A partir da implantação do o Programa Nacional de Desestatização (BRASIL, 1997), o papel do Estado deixou de ser o executor exclusivo da prestação de serviços ao cidadão passando a atuar em atividades de fiscalização, regulamentação, e controle dos produtos e serviços gerados pelas empresas concessionárias desses serviços.

Para que haja a entrega da titularidade da prestação de serviços públicos a um ente particular, são necessários vários estudos de viabilidade, análise de impacto, regulamentação e preparação de corpo técnico de servidores especializados para fiscalizar a atuação do particular.

Além disso, o Estado deve garantir a participação da sociedade nas decisões de regulamentação, na fiscalização do setor “privatizado” e das ações do ente regulador, os quais são formas de controle dos atos da administração.

Um dos mecanismos de controle social sobre as ações das agências reguladoras, bem como sobre as ações dos entes regulados é por meio da disponibilização de dados gerados por esses regulados. São dados de desempenho, financeiros, de composição de custos, de receita, de investimentos, de mercado, etc. Entretanto, esses dados nem sempre são inteligíveis ao cidadão comum. A simples disponibilização desse acervo de dados à população não é efetiva para diminuição da assimetria de informação.

Um dos papéis das agências reguladoras é receber os dados das empresas prestadoras dos serviços públicos sobre as atividades prestadas, avaliar sua consistência e veracidade e divulgar esses dados à sociedade, porém, em uma estrutura inteligível, de forma que a sociedade comum consiga interpretar a informação disponibilizada.

Um ponto importante para a sociedade no que diz respeito a acesso aos dados da Administração pública é que desde 2016 todos os órgãos do Poder Executivo Federal (BRASIL, 2016) são obrigados a publicarem os dados de interesse público em formato aberto e seguindo critérios de livre acesso a esses dados. O processo de abertura de dados à sociedade torna possível para públicos mais especializados sua utilização, controle, pesquisa, desenvolvimento de soluções de software e divulgação em formato mais compreensível ao cidadão.

Entretanto a abertura de dados no Governo Federal, e não poderia ser diferente na ANAC, é tratada de forma passiva, ou seja, por provocação. Somente são considerados a partir da existência dos dados. Não há mecanismos que tratem essas questões de forma ativa,

em momento de concepção de solução, ou durante o levantamento de necessidades. Surgem então as seguintes questões a serem tratadas neste trabalho:

Como tornar essa abertura de dados uma solução eficiente de forma a diminuir a assimetria de informação entre o ente regulado e o cidadão, além de clarear as ações de regulação pelo órgão regulador?

É possível implementar um modelo de governança de informações digitais que promova, de forma ativa, ações de transparência e suporte aos conceitos de dados abertos?

Para responder a essas questões este trabalho apresenta a implementação de mecanismos, dentro de um contexto de governança de informações digitais, que promova, de forma ativa, ações de transparência, aos moldes dos conceitos de dados abertos, e como tornar a abertura de dados uma solução eficiente no controle social e na diminuição da assimetria de informação entre a Agência, a firma regulada e o cidadão, além de clarear as ações de regulação da ANAC.

Este trabalho está estruturado da seguinte forma: a seção REFERENCIAL TEÓRICO aborda a questão histórica e filosófica sobre transparência e abertura de dados e sua importância para o fortalecimento da democracia.

Nessa seção são apresentadas as similaridades existentes entre dados público e bens públicos, as diferenças conceituais entre Governo aberto e dados abertos governamentais, um panorama sobre a situação de governança de dados e informações no Governo.

E ainda, tecnologias disruptivas que podem ser utilizadas para aumentar a potencialidade de dados abertos, além de experiências de sucesso no uso de dados abertos com essas tecnologias.

A seção MECANISMOS PARA VIABILIZAR TRANSPARÊNCIA ATIVA E ABERTURA DE DADOS NA ANAC apresenta 4 mecanismos, complementares e dependentes entre si, que respondem à questão posta; demonstra como a ANAC vem transformando uma determinação normativa (que, de acordo com o Decreto, basta a ação de abertura de dados de forma reativa para estar em conformidade), em um ecossistema de governança e gestão para transparência e abertura de dados dentro de um contexto de informações digitais na Agência, que entrega à sociedade um ferramental para controle social, e com potencial para fomentar a pesquisa científica, promover o desenvolvimento tecnológico e a inovação.

A seção CONCLUSÃO E TRABALHOS FUTUROS traz uma avaliação dos 4 mecanismos e a oportunidade de continuidade deste trabalho em amplitudes de maior impacto, no setor regulado de transporte (em sentido amplo), e nas políticas públicas para o setor de transporte.

2 JUSTIFICATIVA

As agências reguladoras são entidades autárquicas especiais criadas pelo Governo com o objetivo institucional de controlar os responsáveis pela prestação de serviços públicos, sob a forma de concessão, permissão ou autorização, além de intervir no domínio econômico, quando necessário, para evitar abusos dos entes regulados provenientes da iniciativa privada.

O Legislador, ao admitir a criação de agências reguladoras, visava resolver questões de falhas de mercado, em especial as relacionadas à assimetria de informação.

Para que o agente regulador consiga equilibrar o mercado (objeto de sua atuação) e estimular a eficiência do prestador dos serviços públicos, ele necessita de informações técnicas e de atividades econômico-financeiras e operacionais das firmas.

As melhores fontes de obtenção dessas informações são as próprias firmas reguladas, pois elas são as detentoras das informações consideradas relevantes. Às empresas é interessante ocultar as informações a seu respeito, de forma que suas concorrentes não se beneficiem dessas informações.

A consequência da assimetria de informação são outras falhas de mercado ou de governo, como: captura regulatória e externalidade negativa.

A captura regulatória ocorre quando uma agência reguladora, criada para atuar no interesse público, acaba atuando para favorecer grupos de interesses que dominam a indústria ou o setor que estaria encarregada de regular.

Já a externalidade negativa é o custo gerado por uma atividade econômica que não é arcado ou apropriado pelo agente responsável pela atividade econômica, e, em geral, a atividade de fiscalização ou regulatória não consegue identificar previamente, em função da assimetria de informação, como foi o ocorrido no rompimento represa de Mariana em 2016.

No que tange a regulação, o acesso à informação é importante, pois é através dela que os órgãos reguladores podem tomar ações que estimulem a firma regulada a operar de forma eficiente e dentro de parâmetros contratuais e legais.

Na mesma esteira, o acesso aos dados de regulação pela sociedade é a principal forma para diminuir a assimetria de informação entre o ente regulador (a Agência, o cidadão (usuário dos serviços) e as empresas concessionárias e prestadoras desses serviços públicos.

O acesso a esses dados tem sido sistematizado a partir da publicação do Decreto 8.777 de 2016, da Presidência da República, que institui a Política de Dados Abertos do Poder Executivo federal (BRASIL, 2016).

Ocorre que as ações de abertura de dados no Governo Federal, e não poderia ser diferente na ANAC, são tratadas de forma passiva, ou seja, por provocação. Somente são considerados a partir da existência dos dados. Não há mecanismos que tratem essas questões de forma ativa (ou proativa), em momento de concepção de solução, ou durante o levantamento de necessidades, e o pior: não há uma avaliação prévia sobre a eficiência da disponibilização desses dados e muito menos uma avaliação posterior sobre o resultado para a sociedade.

Dessa forma, este trabalho apresenta mecanismos, dentro de um contexto de governança de informações digitais, que promova, de forma ativa, ações de transparência, aos moldes dos conceitos de dados abertos, e como tornar a abertura de dados um mecanismo de apoio para diminuição da assimetria de informação entre a Agência, a firma regulada e o cidadão, além de clarear as ações de regulação da ANAC.

3 REFERENCIAL TEÓRICO

Nesta seção apresentamos fundamentação conceitual que embasa este trabalho, além de apresentar, de forma pontual, marcos de destaque sobre a evolução dos temas de transparência e abertura de dados em alguns países.

Para situarmos melhor o leitor antes de adentrarmos no conteúdo que dá base a este trabalho, convém apresentar conceitos sobre transparência e abertura, visto que há estreita relação entre os dois e que serão constantemente referenciados neste trabalho:

Se o objetivo for um entendimento amplo e geral, a definição de transparência poderia ser simplesmente como “levantar o véu do sigilo” (DAVIS, 1998, p. 121) ou “a capacidade de olhar com clareza através das janelas de uma instituição” (BOER 1998, p. 105). Bovens (2010) define transparência do governo como “uma relação entre um ator (um servidor ou um funcionário) e um fórum (pode ser uma pessoa específica, como um superior, um ministro ou um jornalista, ou pode ser uma instituição, como o parlamento, um tribunal ou o escritório de auditoria), em que o ator tem a obrigação de explicar e justificar sua conduta, o fórum pode colocar questões e avaliar; e o ator enfrenta as consequências dessa avaliação”. Moser (2001, p.3), por exemplo, define ser transparente como “abrir os procedimentos de trabalho não imediatamente visíveis para aqueles que não estão diretamente envolvidos para demonstrar o bom funcionamento de uma instituição”. Oliver (2004, p.2) indica que a transparência pode ser descrita como tendo três elementos: um observador, algo disponível para ser observado e um meio ou método para observação. Este tipo de definição baseia-se na teoria principal-agente: um principal requer informação sobre o agente para verificar se o agente adere ao “contrato”, é uma teoria alinhada com economistas (PRAT, 2005).

Cornelia Moser (2001) discorre sobre a relação existente entre transparência e abertura de dados. Para ela os conceitos são muito próximos, mas não há dúvidas quanto sua importância - desde que haja amplo acesso a informação -, para legitimação da democracia e para elaboração de políticas. E isso fica claro ao analisarmos o conceito que ela apresenta sobre transparência, que é o acesso popular ao processo decisório das instituições.

“O denominador comum dessas diferentes abordagens é sua ligação de transparência e abertura à legitimidade (democrática). Eles estão se concentrando na transparência das instituições políticas em sua relação com a sociedade civil, sejam elas pessoas físicas ou interesses comerciais, a mídia ou representantes eleitos. No entanto, as definições de transparência variam entre as várias abordagens e enfatizam características particulares da ideia subjacente. A transparência é vista e definida de forma diferente: variando de uma perspectiva estreita de procedimentos simples e claros até o mais amplo “acesso popular efetivo ao processo de tomada de decisão” (MOSER, 2001 apud BOER 1998, p. 99).

Transparência, em sua forma mais simples, é “a capacidade de ver claramente através das janelas de uma instituição” (MOSER, 2001 apud BOER, 1998, p. 105). A própria ideia é esclarecida por esta metáfora: abrir os procedimentos de trabalho não imediatamente visíveis àqueles que não estão diretamente envolvidos para demonstrar o bom funcionamento de uma instituição.

Adaptada aos sistemas políticos, transparência é uma característica da relação entre governante e governado, possibilitando “que a operação do governo seja suficientemente aberta à opinião pública e suficientemente simples em seus fundamentos para que os cidadãos possam entender prontamente como e o que está fazendo.” (MOSER, 2001 apud DAHL, 1998, p.126).

Dados esses atributos, transparência é geralmente usada em mesmo contexto de abertura e frequentemente vinculada à simplicidade e à compreensão. A abertura e a transparência também podem ser vistas como conceitos ligeiramente diferentes (MOSER, 2001 apud DAHL, LARSSON, 1998, p. 40), mas geralmente são igualadas, pois ambas se referem a uma determinada qualidade.

Assim, “abertura e transparência são claramente uma e a mesma coisa (...), a atividade de levantar o véu do sigilo” (MOSER, 2001 apud DAVIS, 1998, p. 121). A transparência e / ou a abertura são vistas como um pré-requisito para a formulação de políticas modernas, acrescentando o sabor distinto da legitimidade democrática.”

Para entendermos sobre “transparência” recorreremos à sociologia histórica, que é um ramo da sociologia que analisa as mudanças sociais ao longo do tempo, posicionando-as em tendências e desenvolvimentos históricos, as situações atuais só podem ser compreendidas se soubermos como elas se desenvolveram ao longo do tempo. Por meio dela é que podemos entender a origem das instituições atuais dentro do contexto das relações de poder, padrões de valor e crenças do tempo de origem para compreendermos sua composição atual (MEIJER, 2019 apud ZWAAN, 2001).

Conforme demonstra Meijer (2015), uma análise histórica demonstra como a transparência se transformou ao longo do tempo, de uma ideia normativa para uma prática garantida, como ela assumiu um papel no desenvolvimento do sistema de freios e contrapesos, daí a necessidade de retrocedermos 200 anos até a origem dos conceitos sobre transparência.

Meijer (2015) aponta que o ponto de partida sobre transparência remonta 250 anos, com a Lei Sueca da Liberdade de Imprensa em 2 de dezembro de 1766, importante marco revolucionário do século XVIII contra a prepotência do absolutismo monárquico europeu e a

intolerância religiosa; seguido 200 anos depois pelos Estados Unidos (1966), pelo Japão (1999), Peru e México (2002), Reino Unido e Índia (2005) e Brasil (2011).

Em 1789, houve uma ampliação da coisa pública sobre a vida privada com a Revolução Francesa. Desde muito antes já havia grande discussão sobre o que é considerado público ou privado. Pode-se dizer que o marco que radicaliza o princípio de direito de acesso aos documentos de arquivo foi 14 de julho de 1789, auge da Revolução Francesa, com a Queda da Bastilha.

Após a Queda da Bastilha a Assembleia Nacional Constituinte da França revolucionária aprovou em 26 de agosto de 1789 a Declaração dos Direitos do Homem e do Cidadão (FRANÇA, 1789), sintetizado em dezessete artigos, sendo que o artigo 15 cita que *“La société a le droit de demander compte à tout agent public de son administration”* (a sociedade tem o direito de pedir contas a todo agente público pela sua administração).

A Declaração Universal dos Direitos Humanos foi proclamada pela Assembleia Geral da Organização das Nações Unidas em resposta aos atos ocorridos na Segunda Guerra mundial. Ela assegura ao ser humano, além de outros, o direito fundamental de acesso à informação, e ao Estado o dever da transparência, independentemente de limites territoriais tanto entre os povos dos próprios Países-Membros quanto entre os povos dos territórios sob sua jurisdição, com base em seu artigo 19:

“Todos têm direito à liberdade de opinião e de expressão; este direito inclui a liberdade de expressar opiniões sem interferência e de buscar, receber e transmitir informações e ideias por quaisquer meios e sem limitações de fronteiras”. (ASSEMBLÉIA GERAL DAS NAÇÕES UNIDAS, 1948).

3.1 Dados, informação e conhecimento como bens públicos:

A ideia de bem comum aplicada ao conhecimento já havia sido teorizada, bem antes da invenção da Internet. Robert King Merton foi um dos pais da sociologia da ciência. A teoria que levou seu nome mostra os benefícios sobre a abertura de dados científicos. Já em 1942, Merton mostrou a importância de que os resultados de pesquisas fossem livremente acessíveis a todos. Cada pesquisador deveria também contribuir para um “pote comum” e abrir mão dos direitos de propriedade intelectual para permitir que o conhecimento avançasse (HESS & OSTROM, 2007, p. 124 - 140).

Conforme publicado pela UNESCO (2015), a primeira iniciativa para criar um banco de dados de bibliografia eletrônica com acesso aberto foi do Centro de Informações sobre Recursos Educacionais (ERIC) em 1966. Ainda no mesmo ano, o MEDLINE foi criado - um banco de dados on-line gratuito gerenciado pela National Library of Medicine e National

Instituto de Saúde (EUA) com citações bibliográficas de periódicos da área biomédica, que mais tarde seria chamado de PubMed, atualmente com mais de 14 milhões de artigos completos.

O termo dados aberto apareceu pela primeira vez em 1995 quando o Comitê de Dados Geofísicos e Ambientais - National Research Council publicou na *On the Full and Open Exchange of Scientific* seu posicionamento com relação ao compartilhamento e abertura de dados para o progresso das pesquisas:

A atmosfera, os oceanos e a biosfera da Terra formam um sistema integrado que transcende as fronteiras nacionais. Para entender os elementos do sistema, a maneira como eles interagem e como eles mudaram com o tempo, é necessário coletar e analisar dados ambientais de todas as partes do mundo. Estudos do ambiente global requerem colaboração internacional por muitas razões:

para tratar de questões globais, é essencial ter conjuntos de dados globais e produtos derivados desses conjuntos de dados;

é mais eficiente e econômico para cada nação compartilhar seus dados e informações do que coletar tudo o que precisa de forma independente; e

a implementação de políticas eficazes que abordem questões do ambiente global requer o envolvimento desde o início de quase todas as nações do mundo.

Os programas internacionais para pesquisa de mudança global e monitoramento ambiental dependem crucialmente do princípio da troca de dados completa e aberta (ou seja, dados e informações são disponibilizados sem restrições, de forma não discriminatória, para não mais do que o custo de reprodução e distribuição). (NATIONAL RESEARCH COUNCIL. 1995, p.8).

Esse posicionamento ratificou o compromisso publicado pelo World Meteorological Organization (WMO) em junho de 1990, citado por National Research Council (1995, p.8): “Solicita que os membros reafirmem seu compromisso com o ‘intercâmbio internacional livre e irrestrito de dados e produtos meteorológicos básicos’ entre os serviços meteorológicos nacionais”.

Cabe ressaltar que essa publicação, *On the Full and Open Exchange of Scientific*, embora seja de áreas de pesquisas relacionadas dados meteorológicos, percebeu a necessidade da adoção da prática de abertura de dados e informações em todos os demais setores ambientais:

“Embora este relatório se concentre em dados relativos à mudança ambiental global em geral e à proposta da Organização Meteorológica Mundial (OMM) em particular, as questões discutidas têm relevância para todos os tipos de dados ambientais”. (NATIONAL RESEARCH CONCIL, 1995, p.1).

A tendência dessa época ainda não era de abertura de dados de forma ampla. No entanto, no meio acadêmico isso já ocorria, já reconheciam a necessidade de compartilhamento de informações. como demonstrado acima. Roberta Balstad (2012, p. 18), sumariza os benefícios advindos de políticas de compartilhamento de dados científicos:

- O acesso aberto à ciência contribui para a inovação e o crescimento econômico.

- Os avanços científicos, tanto substantivos quanto metodológicos, são agora intensivos em dados e requerem acesso aberto a dados científicos.
- O custo da pesquisa é reduzido. Isto é muito importante agora na maioria dos países, porque muitas vezes há menos dinheiro disponível para pesquisa. Para manter a ciência viva e vital, o acesso aberto aos dados é uma vantagem real. (BALSTAD, 2012, p. 18).

Por outro lado, a autora também menciona as desvantagens trazidas pela restrição de compartilhamento de informações:

Limitar o acesso a dados - o outro lado dessa moeda - resulta em custos mais altos de pesquisa, oportunidades perdidas, barreiras à inovação, cooperação científica menos eficaz, qualidade sub-ótima dos dados (já que como ninguém está trabalhando com eles não há como fornecer correções) e um fosso crescente entre os países da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) e os países em desenvolvimento. (BALSTAD, 2012, p. 18).

Mas a abertura de dados no meio científico não deve ser vista como algo absoluto, não deve ser tratado de forma extremista. A autora ainda cita as exceções em que se deve manter a restrição à abertura de dados:

No entanto, ao exigir acesso aberto aos dados, devemos reconhecer que existem alguns motivos legítimos para limitar o acesso a dados públicos:

- Segurança nacional e segurança pública.
- Privacidade pessoal e confidencialidade, que são protegidos em muitos países.
- Direitos proprietários de partes do setor privado. Ninguém está falando sobre forçar acesso aberto a pesquisas que uma empresa fez para promover seu produto. (BALSTAD, 2012, p. 18).

É de se observar que mesmo sem um referencial teórico ainda discorrido sobre dados abertos e transparência, em sua forma ampla, é possível inferir que as citações da autora Roberta Balstad são perfeitamente aplicáveis a qualquer área de conhecimento produtora ou mantenedora de dados e informações, inclusive suas vantagens, desvantagens e exceções à transparência e abertura de dados.

Muito antes de ser um objeto técnico ou movimento político, os dados abertos estavam enraizados na praxis da comunidade científica. Os pesquisadores foram os primeiros a perceber o benefício da abertura e do compartilhamento de dados.

Ainda na comunidade acadêmica outro movimento de acesso aberto pode ser verificado em 1960 com o lançamento do "Project Gutenberg" por Michel Hart. Este é considerado o marco principal do movimento. O surgimento de periódicos de acesso aberto ocorreu no final da década de 1980, com a publicação da revista "Psychology", um jornal online gratuito de Stevan Harnad em 1989. No mesmo ano, outra revista "Computer Systems Review" foi publicada também como um jornal online gratuito (UNESCO, 2015, p.7).

Em meados de 1990 começou um movimento no sentido de reconhecer a informação como um bem comum, compartilhável. Nessa época a informação digital era

fechada, mercantilizada e excessivamente patenteada. Artigos científicos surgiram com estudos sobre a situação da proteção dos conteúdos digitais da época, citados por Charlotte Hess and Elinor Ostrom (2007, p. 4). Um número crescente de estudiosos descobriu que o conceito de “Bens comuns” ou “commons” os ajudariam a desvendar novos dilemas que estavam observando com o surgimento de informações digitais distribuídas. O Commons tornou-se uma palavra de ordem para a informação digital.

As tecnologias da informação também deram um novo fôlego a essa filosofia dos bens comuns. Simon Chignard (2013) cita pesquisa em que o Prêmio Nobel de Economia de 2009, a economista Elinor Ostrom, mostrou a especificidade das informações como bens comuns. “Eles são muito semelhantes aos bens públicos, porque o uso deles por uma pessoa não impede seu uso pelos outros. No entanto, estes são bens públicos de um novo tipo: não apenas o seu uso não diminui o estoque comum, mas o enriquece”. (CHIGNARD, 2013).

Na esteira do “commons”, KOLLOCK & SMITH (1996, p.2), compararam a informação digital aos tradicionais bens comuns conhecidos (por exemplo, televisão pública ou um projeto de melhoria comunitária). Um bem público ou bem comum é um recurso do qual todos podem se beneficiar, independentemente de terem ajudado a cria-lo ou não. A tentação é desfrutar de um bem público sem contribuir para sua produção, mas se todos chegam a esta decisão, o bem nunca é produzido e o prejuízo é para todos.

No caso de bens comuns, quando uma pessoa não pode ser excluída dos benefícios que os outros proporcionam, cada pessoa é motivada a não contribuir para o esforço conjunto, mas a caronear-se dos esforços dos outros. Se todos os participantes optarem por caronear, o benefício coletivo não será produzido. A tentação de caronear, no entanto, pode dominar o processo de decisão e, assim, tudo acabará em lugar algum, sem o benefício coletivo. Isso é o que KOLLOCK & SMITH (1996, p. 1) chamaram de “a tragédia dos comuns”.

Os conceitos dos bens comuns, até então conhecidos, poderiam ser incorporados aos bens comuns digitais (ou novos commons), inclusive os princípios descobertos a partir de estudos de pesquisa de bens comuns citados por Kollock & Smith (1996 apud Ostrom 1990, 90-102) são sugestivos de serem aplicáveis em bens digitais, são eles:

- Limites em vigor claramente definidos.
- As regras em uso são compatíveis com as necessidades e condições locais.
- Indivíduos afetados por essas regras podem participar de modificações das regras.
- O direito dos membros da comunidade de elaborar suas próprias regras é respeitado por autoridades externas.

- Estabelecimento de sistema para o comportamento dos membros.
- Existência de sistema graduado de sanções.
- Acesso, pelos membros da comunidade, a mecanismos de baixo custo para resolução de conflitos.
- Mecanismos de gestão relacionados - isto é, apropriação, provisão, monitoramento e sancionamento, resolução de conflitos e outras atividades de governança - são organizadas em uma estrutura relacionada com várias camadas de atividades.

Esses princípios foram descobertos após a realização de um grande conjunto de estudos empíricos sobre governança de recursos comuns.

Entretanto, a maioria dos tipos de conhecimento tradicionalmente tem sido relativamente não-subtrativa, ou seja, quanto mais pessoas compartilham conhecimento útil, maior é o bem comum, isso diferente de outros tipos de bens que quanto mais são utilizados e compartilhados, maior é a probabilidade de eles se esvaírem.

3.2 Mas e os Direitos Autorais?

Com a exploração do conhecimento e da informação digital, tornando-a fechada, mercantilizada e excessivamente patenteada, os produtores de Commons digitais - sejam eles dados, informação, conhecimento, sistema, código ou qualquer outra propriedade intelectual -, contam com diversos mecanismos de apoio ao direito autoral e intelectual, por exemplo: copyright, copyleft, Open Database License, GNU Lesser General Public License, GNU General Public License, GNU Free Documentation License, Affero General Public License, creative Commons, as legislações locais, entre outros (WIKIPEDIA, 2013).

Cada um desses mecanismos tem características próprias, porém, de forma geral, todas objetivam garantir de alguma forma, ao autor, o direito que tem sobre suas obras, sejam elas literárias, artísticas ou científicas, dados, informação, conhecimento, sistema, código ou qualquer outra propriedade intelectual com maior ou menor grau de restrição.

Embora, o objetivo deste trabalho não seja focar nesses mecanismos para garantia do direito sobre a produção intelectual, dados, informação ou conhecimento, haveria uma sensação de vazio caso não apresentássemos, mesmo que de forma superficial, ao menos um desses mecanismos, inclusive sua jornada desde a motivação da criação até sua materialização.

Cabe salientar que este assunto encontra alinhamento com o tema deste trabalho, pois como será detalhado adiante, na seção Os Princípios dos dados abertos – encontro de

Sebastopol, “dados abertos” são sustentados por oito princípios (princípios do Open Government Data), sendo que destes, três referem-se a direito autoral:

Não-discriminatório: os dados devem estar disponíveis para qualquer pessoa, sem necessidade de registro.

Não-proprietário: os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tem o controle exclusivo.

Licença livre: os dados não estão sujeitos a quaisquer direitos de autor, patentes, marcas comerciais ou regulamento secreto. Pode ser permitida uma razoável privacidade e restrições de privilégio e segurança.

O conceito de direitos autorais pode ter começado com o rei Diarmait Mac Cerbhaill, no final do século VI na Irlanda. Ele estava presidindo uma disputa sobre a propriedade do Cathach - um dos primeiros exemplares da escrita irlandesa que contém uma versão dos Salmos 10 a 13. Um missionário irlandês, Saint Columba, recebeu um livro de salmos de Saint Finian, e tinha feito uma cópia. Finian argumentou que, como ele possuía o original, ele deveria possuir a cópia, mas Columba discordou.

Diarmait Mac Cerbhaill tentou resolver o assunto com o julgamento: "Para cada vaca pertence o seu bezerro, portanto, para cada livro pertence a sua cópia", marcando a primeira decisão documentada no que diz respeito aos direitos de autor. (THE ROYAL IRISH ACADEMY, 2002).

Aproximadamente 900 anos após a disputa sobre a propriedade do Cathach, com o surgimento da imprensa, facilitando a produção de várias cópias de um mesmo livro, começou a rápida disseminação de informações e, para remediar isso, foi criada uma situação em que poderosas gráficas, limitadas aos membros da Company of Stationers of London, receberam o monopólio sobre a capacidade de publicar informações, a qual passou a ser a detentora do monopólio de impressão na Inglaterra.

O lado negativo disso era que apenas a literatura que o grupo quisesse imprimir poderia ser impressa e, como resultado, eles rapidamente se apropriaram do conteúdo - os autores e o público não tinham poder nenhum contra essa poderosa distorção comercial do mercado e, portanto, sem direitos. O poder da Company of Stationers of London aumentou a ponto que “o Judiciário não interferia nas políticas públicas de censura”. (ZANINI, 2010).

Além disso, o monopólio limitava a utilização das obras dos autores por 14 anos, contados do dia de sua primeira publicação, sendo também oferecida a possibilidade de renovação do direito por outros 14 anos (ZANINI, 2010). Criou-se o conceito de domínio

público pela primeira vez, onde o público em geral é dono de um trabalho criativo, desde que findos esses prazos.

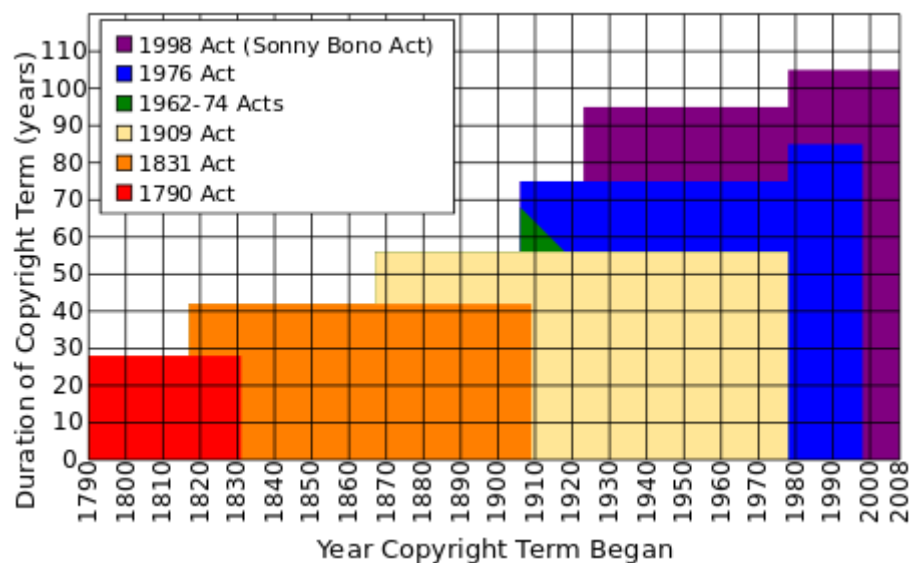
O início do século XVIII foi marcado com o surgimento da primeira norma legal que reconheceu o que se conhece no direito anglo-saxão como copyright (GEERE, 2011). O Estatuto da Rainha Ana, promulgado em 10 de abril de 1710, logo após a invenção da imprensa, constitui um antecedente histórico de grande importância no âmbito dos direitos de propriedade intelectual.

Este normativo foi focado a corrigir os problemas existentes nesse momento em torno da reprodução e venda de obras literárias. Ele reconhece ao autor o seu direito de propriedade, os quais poderiam ser transferidos ao editor (LESSIG, 2005, p. 103-104).

O surgimento da Creative Commons iniciou 289 anos depois do Estatuto da Rainha Anne. Seu marco inicial foi com a aprovação pelo Congresso dos EUA da Lei de Extensão de Direitos Autorais de Sonny Bono, também conhecida como Lei de Sonny Bono, ou (com desprezo) a Lei de Proteção de Mickey Mouse, a qual acrescentava mais 20 anos à duração dos direitos autorais, é o que nos narra Geere (2011).

Atualmente a duração dos direitos autorais para a maioria dos trabalhos produzidos nos Estados Unidos é de 70 anos após a morte do autor, ou 120 anos após a criação ou 95 anos após a publicação (o que for menor). Os EUA aprovaram uma série de outras extensões de prazo ao longo dos anos 60, 70 e 80 (GEERE, 2011) - geralmente da mesma forma que os direitos autorais do Mickey Mouse que estavam prestes a expirar. Termo de copyright é o período em que o direito autoral subsiste em um trabalho antes de passar para o domínio público. A evolução da duração dos Termos de copyright nos Estados Unidos é apresentada no gráfico abaixo:

Figura 1 - Evolução da duração dos Termos de copyright nos Estados Unidos



Fonte: wikipedia (2019)

Embora esses atos tenham sido implementados para proteger empresas, a insegurança jurídica com relação aos termos de direitos autorais, trazia problemas e prejuízos para empresas que se aguardavam obras entrarem no domínio público em função do término de seus prazos de copyright.

Eric Eldred era uma dessas pessoas - ele administrava um site que reimprimia obras cujos direitos autorais haviam expirado, tornando-os mais amplamente disponíveis em vários formatos. Um desses Atos do Congresso Americano iria destruir o seu negócio pelos próximos vinte anos, então ele foi ao tribunal e à Suprema Corte para argumentar que o ato era inconstitucional.

De acordo com Geere (2011), o cerne de seus argumentos foram: (1) que estender termos existentes violava a requisição de “tempo limitado” da Constituição, e (2) que estender termos por mais 20 anos violava a Primeira Emenda.

Eldred teve a companhia de uma série de outras pessoas com interesses comerciais e não comerciais, e seu advogado foi Lawrence Lessig, ativista político e professor da Harvard Law School. Em 17 de fevereiro de 1999, Lessig formou uma legião de interessados em ajudar a combater o caso, que, além de Eldred, Hal Abelson, professor de engenharia elétrica e ciência da computação no MIT. O movimento que inicialmente recebeu o nome de Copyrights Commons (GEERE, 2011) foi renomeado para Creative Commons.

Embora eles não tivessem obtido êxito na empreitada jurídica contra os atos do Congresso Americano, tendo conseguido levar o caso até a Suprema Corte, o movimento Copyrights Commons deixou um legado de grande importância: a Creative Commons, fundada

em 2001 com o do Centro para o Domínio Público. É liderado por um Conselho de Administração que inclui especialistas em propriedade intelectual e cibernética James Boyle, Michael Carroll, Molly Shaffer Van Houweling e Lawrence Lessig, professor de ciência da computação do MIT Hal Abelson, advogado que se tornou documentarista e especialista em cibercrime Eric Saltzman, o renomado documentarista Davis Guggenheim, e o editor de web de domínio público Eric Eldred.

A Creative Commons é uma empresa com o objetivo de compartilhar o uso da criatividade e do conhecimento por meio de instrumentos jurídicos gratuitos. As licenças da Creative Commons estão adaptadas ao ordenamento jurídico brasileiro e, conforme será detalhado adiante, atendem aos ditames previstos pelo Decreto Nº 8.777, de 11 de maio de 2016 que institui a Política de Dados Abertos do Poder Executivo federal.

3.3 Transparência e Dados abertos

Dados abertos foram definidos por um conjunto de princípios estabelecidos em um encontro realizado nos dias 7 e 8 de dezembro de 2007 em Sebastopol, na Califórnia, que reuniu 30 advogados, pesquisadores, representantes de organizações da sociedade civil e ativistas norte-americanos. (BRAINER, 2015).

Entre eles estavam Lawrence Lessig, famoso criador do termo “cultura livre”; Tim O’Reilly, publisher de editora da área de tecnologia da informação que leva seu sobrenome e criador da expressão “Web 2.0”; Ethan Zuckermann, diretor do Center of Civic Media MIT e criador do Trip Advisor; Joseph Hall, Center for Democracy & Technology; Aaron Schwartz, conhecido ativista hacker, cuja vida é retratada no filme “The Internet’s Own Boy”, além dos fundadores de organizações Sunlight Foundation, My Society e GovTrack – pioneiras no uso dos dados abertos na promoção da transparência.

A autora demonstra a evolução do tema (BRAINER, 2015), embora o foco do encontro fosse a abertura de dados governamentais, nos anos que se seguiram, o conceito passou a ter seu uso ampliado, passando a incluir dados científicos ou mesmo de organizações privadas. Cabe mencionar que a sociedade acadêmica já estava passos à frente no compartilhamento de dados científicos, conforme descrito na seção “Dados, informação e conhecimento como bens públicos” deste documento.

Ainda naquele mesmo mês, de acordo com as publicações do Congresso dos Estados Unidos (U.S. GOVERNMENT, 2007), em 31 de dezembro de 2007, foi promulgado pelo Congresso Americano o OPEN Government Act of 2007, que altera a Lei de Liberdade de

Informação (FOIA) para revisar os requisitos para divulgações de agências federais de informações solicitadas sob essa Lei. De modo geral, essa lei propunha promover a acessibilidade, responsabilização e abertura no Governo, fortalecendo a seção 552 do título 5, o Código dos Estados Unidos (comumente chamado de Lei de Liberdade de Informação).

3.3.1 **Os Princípios dos dados abertos – encontro de Sebastopol**

O encontro ocorrido nos dias 7 e 8 de dezembro de 2007 em Sebastopol, na Califórnia, contou com a participação diversas mentes brilhantes, como citado nos parágrafos anteriores. O resultado histórico foi a produção dos 8 princípios, conforme apresentado abaixo:

Completo - todos os dados públicos devem ser disponibilizados. Dados públicos são dados que não estão sujeitos a restrições de privacidade, segurança ou privilégios de acesso.

Embora os dados físicos, não estejam sujeitos aos princípios do Open Government Data, é recomendável que tais recursos sejam disponibilizados eletronicamente na medida do possível.

Primários - os dados devem ser coletados na fonte com o maior nível de granularidade possível, e não de forma agregada ou modificada.

Tempestivo - sua disponibilidade deve ser feita tão rapidamente quanto necessário para preservar o valor dos dados.

Acessibilidade - os dados devem estar disponíveis para o maior número de usuários e as mais diversas finalidades. Os dados devem ser disponibilizados na Internet de forma a permitir um amplo acesso de usuários.

Isso significa considerar como as escolhas na preparação e publicação de dados afetam usuários portadores de necessidades especiais e como isso pode afetar usuários de uma variedade de plataformas de software e hardware. Os dados devem ser publicados com os atuais protocolos e formatos comuns (considerados “padrão”), bem como protocolos e formatos alternativos, quando os padrões da indústria impuserem ônus à ampla reutilização dos dados.

Processável por máquinas - os dados devem ser razoavelmente estruturados de modo a permitir o processamento automatizado.

A capacidade de os dados serem amplamente utilizados exige que eles sejam codificados ou estruturados adequadamente. Texto de forma livre não é um substituto para registros tabulares e normalizados. Imagens de texto não são substitutos para o texto em si. Devem ser disponibilizadas documentações sobre o formato de dados e significados dos itens de dados para os usuários.

Não-discriminatório - os dados devem estar disponíveis para qualquer pessoa, sem necessidade de registro. Deve ser permitido acesso anônimo aos dados públicos, incluindo o acesso por meio de proxies anônimos. O acesso aos dados não deve ser dificultado com a existência de vários passos e registros para habilitar o usuário.

Não-proprietário - os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tem o controle exclusivo. Formatos proprietários adicionam restrições desnecessárias sobre quem pode usar os dados, como eles podem ser usados e compartilhados, e se os dados serão utilizáveis no futuro. Embora alguns formatos proprietários sejam quase onipresentes, não é aceitável usar apenas formatos proprietários. Da mesma forma, os formatos não proprietários relevantes podem não atingir um público amplo. Nesses casos, convém disponibilizar os dados em vários formatos.

Licença livre - os dados não podem estar sujeitos a quaisquer direitos de autor, patentes, marcas comerciais ou regulamento secreto. Privacidade razoável, restrições de segurança e privilégio podem ser permitidas.

Como as informações do governo são uma combinação de registros públicos, informações pessoais, trabalhos protegidos por direitos autorais e outros dados não abertos, é importante ter clareza sobre quais dados estão disponíveis e quais licenças, termos de serviço e restrições legais se aplicam. Os dados para os quais não se aplicam restrições devem ser marcados claramente como sendo de domínio público. (BRAINER, 2015).

Embora esses 8 princípios sejam um marco na busca dos resultados previstos para dados abertos governamentais, inclusive são adotados pelo Governo Brasileiro, há ainda a recomendação de adoção de outros princípios com igual importância, conforme divulgado no sítio do Open Government Data (2007):

Gratuito e online - a informação não é significativamente pública se não estiver disponível na Internet gratuitamente, ou pelo menos não mais do que o custo marginal de reprodução. Também deve ser localizável.

Permanente - os dados devem ser disponibilizados em um local estável na Internet, indefinidamente e em um formato de dados estável pelo maior tempo possível.

Confiável - o conteúdo publicado deve ser assinado digitalmente ou incluir atestado de data de publicação / criação, autenticidade e integridade.

Presunção de abertura - A presunção de abertura se baseia em leis como a Lei de Liberdade de Informação, procedimentos que incluem gerenciamento de registros e ferramentas, como catálogos de dados.

Documentado - a documentação sobre o formato e o significado dos dados ajuda muito a tornar os dados úteis. Os sites do governo devem fornecer aos usuários informações suficientes para fazer avaliações sobre a precisão e o valor das informações publicadas no site.

Seguro para abrir - os órgãos governamentais que publicam dados online devem sempre procurar publicar usando formatos de dados que não incluem conteúdo executável.

Projetado com participação da sociedade - o público está na melhor posição para determinar quais tecnologias de informação serão mais adequadas para as aplicações que o público pretende criar para si mesmo. A contribuição pública é, portanto, crucial para disseminar informação de tal forma que ela tenha valor. (OPEN GOVERNMENT DATA, 2007)

Cabe observar que os instrumentos no arcabouço jurídico brasileiro equivalentes à Lei de Liberdade de Informação (FOIA), incluindo os 8 princípios propostos no encontro em Sebastopol e os 7 princípios recomendados no sítio do Open Government Data, são o Decreto 8.777, (que Institui a Política de Dados Abertos do Poder Executivo federal (BRASIL, 2016) e a Lei de Acesso a Informação (BRASIL, 2011a). Conforme será demonstrado adiante, esses dois conjuntos de princípios foram incorporados ao arcabouço jurídico brasileiro.

3.4 Transparência e Dados Abertos Governamentais no Brasil

Transparência e dados abertos, além de terem conotações semelhantes, objetivam alcançar o mesmo e principal resultado: o fortalecimento da democracia (MOSER, 2001, p.3). Ambos os conceitos estão também relacionados à publicidade dos atos da administração pública:

A publicação de informações pela Administração Pública brasileira é um dever previsto no princípio da publicidade (art. 37 da Constituição Federal brasileira de 1988), bem como direito do cidadão em receber informações públicas, previsto no inciso XXXIII do art. 5º da mesma Constituição. É muito mais que mera questão burocrática a ser atendida pelos órgãos públicos. O seu efeito positivo garante a saúde da democracia e mantém a confiança nas instituições, uma vez que possibilita a qualquer cidadão ser um fiscal dos atos Administração, é o que nos ensina Demócrito Ramos Reinaldo (1997):

"O princípio da publicidade obriga a Administração Pública a expor todo e qualquer comportamento que lhe diga respeito. É esse princípio que confere certeza às condutas estatais e segurança aos administrados. A publicidade resulta, no Estado contemporâneo, do princípio democrático segundo o qual sendo o poder do povo (art. 1º, parágrafo único, da CF/88) e, conseqüentemente, sendo o Estado o próprio povo reunido e constituído sob determinado modelo de Direito, para atingir seus objetivos definidos sistematicamente, tudo o que a pessoa estatal faça ou deixe de fazer, enfim, todos os seus comportamentos, deve ser do conhecimento público. Considerando-se que a democracia que se põe à prática contemporânea conta com a participação direta dos cidadãos, especialmente para efeito de fiscalização e controle da juridicidade e da moralidade administrativa, há que se concluir que o princípio da publicidade adquire, então, valor superior àquele antes constatado na história, pois não se pode cuidar de exercerem os direitos políticos sem o conhecimento do que se passa no Estado" (REINALDO, 1997).

Autores envolvidos com o tema, associam transparência com a divulgação e o acesso às informações da Administração Pública, ou seja, ao princípio da publicidade previsto na Constituição Federal (MELLO, 2008; ARAUJO, 2006; DI PIETRO, 2006):

evidencia-se também que todos têm direito de ter conhecimento do que os agentes fazem no gerenciamento da coisa pública. É o sentido geralmente designado por transparência, bastante referido pelos políticos, por influência de processo nesse sentido em passado recente (glasnost), nos países ex-socialistas. (VAZ, RIBEIRO & MATHEUS, 2010 apud ARAUJO, 2006, p. 58)

Ora, se o Princípio da Publicidade está expresso na Constituição Federal e é requisito para validade e eficácia dos atos da Administração Pública, então concluímos que a transparência se relaciona com os atos da Administração e é a evidência dos atos válidos, sendo uma das formas de confirmar a legalidade dos atos da Administração. Essa mesma conclusão pode também ser obtida a partir da leitura das diretrizes prescritas no Art. 3º da Lei nº 12.527, de 18 de novembro de 2011 que regula o acesso a informações da Administração Pública:

Art. 3º Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:

I - observância da publicidade como preceito geral e do sigilo como exceção;

II - divulgação de informações de interesse público, independentemente de solicitações;

III - utilização de meios de comunicação viabilizados pela tecnologia da informação;

IV - fomento ao desenvolvimento da cultura de transparência na administração pública;

V - desenvolvimento do controle social da administração pública. (BRASIL, 2011a)

Ressalta-se, entretanto, a exceção expressa com relação a dados sigilosos, os quais encontram-se expressos tanto nas Diretrizes da Lei 12.527/2011, citado no parágrafo anterior, quanto na Constituição Federal brasileira de 1988, Art. 37, Inciso XXXIII:

“todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”. (BRASIL, 1988).

Em linha gerais, esses comandos tratam de informações com as seguintes características:

Informações Pessoais: informações relacionadas à pessoa natural identificada ou identificável relativas à intimidade, vida privada, honra e imagem.

Informações imprescindíveis à segurança da sociedade ou do Estado.

Informações cujo sigilo é garantido por outras legislações: por exemplo, sigilo bancário ou comercial, segredo de justiça, hipóteses de segredo industrial, entre outras.

Na outra esteira, de acordo com a definição da Open Knowledge Internacional, dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura. Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta.

Dados abertos no governo, dados abertos, ou dados abertos governamentais têm sido os termos indiscriminadamente utilizados para denominar o processo de “disponibilização, através da Internet, de informações e dados governamentais de domínio público para a livre utilização pela sociedade” (AGUNE, GREGORIO FILHO, BOLLIGER, 2010). Porém há algumas diferenças conceituais que merecem ser apresentadas.

O que difere dados abertos de outras informações governamentais abrangidas pelo Princípio da Transparência é o padrão do formato aberto. A Lei que regula o acesso a informações da Administração Pública, a Lei 12.527/2011, não traz exigências sobre requisitos

técnicos, estruturais, de padrão ou de formato da informação (exceto a utilização de meios de comunicação viabilizados pela tecnologia da informação). Já no Decreto 8.777/2016, que institui a Política de Dados Abertos do Poder Executivo federal há, sim, previsão de critérios técnicos a serem atendidos de forma a possibilitar que os dados sejam acessados, reutilizados, lidos por máquinas e distribuídos livremente, além de ter foco em dados brutos, no menor nível de agregação possível.

Além disso, dados abertos são uma metodologia para a publicação de dados do governo em formatos reutilizáveis, visando o aumento da transparência e a maior participação política por parte do cidadão, além de gerar diversas aplicações desenvolvidas de forma colaborativa pela sociedade (OPEN KNOWLEDGE FOUNDATION, p.04).

Então, na mesma esteira temos o comando legal que insta o Governo a abrir seus dados, respeitando critérios e princípios, em formato específico, a qualquer cidadão, para que sejam acessados, reutilizados, lidos por máquinas e distribuídos livremente, e a condição fundamental da transparência para que realmente haja conformidade com o Princípio da Transparência e o Princípio Constitucional da Publicidade, resguardado, porém, as exceções que justificam sigilo da informação, tanto para dados abertos quanto para transparência da informação.

3.5 Governo aberto, dados abertos governamentais e os compromissos com a OGP

A Parceria para Governo Aberto ou OGP (do inglês Open Government Partnership) é uma iniciativa internacional que pretende difundir e incentivar globalmente práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social.

A OGP foi lançada em 20 de setembro de 2011, quando os oito países fundadores da Parceria (África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido) assinaram a Declaração de Governo Aberto e apresentaram seus Planos de Ação. Atualmente, 75 países integram a Parceria.

Conforme apresentado no sítio www.governoaberto.gov.br a OGP define o conceito de governo aberto por meio de quatro compromissos que devem ser assumidos pelos Governos (OGP BRASIL, 2011):

a) aumentar a disponibilidade de informações sobre atividades governamentais;

- b) apoiar a participação social;
- c) implementar os mais altos padrões de integridade profissional na Administração;
- d) aumentar o acesso a novas tecnologias que promovam a transparência e a prestação de contas.

Cabe salientar que o Governo Brasileiro assumiu os quatro compromissos listados em setembro de 2011, sendo que o comprometimento dos países perante a OGP e suas diretrizes se materializa na forma de um plano de ação nacional, contendo as ações concretas e mensuráveis para o fortalecimento da transparência e para a construção de um governo aberto no âmbito de cada país membro.

Governo Aberto também é visto como uma forma de estabelecer canais de comunicação e cooperação entre políticos, administração pública, indústria e cidadãos.

Em uma visão de tecnicista Governo Aberto é um modelo de governo interoperável no qual pessoas e sistemas podem se comunicar facilmente. O papel das TIC é fundamental para a abertura do governo, embora a tecnologia não possa explicar todas as mudanças recentes.

A Controladoria Geral da União – CGU (BRASIL, CGU, 2016) apresenta os conceitos da OGP sobre Governo Aberto. Dados abertos governamentais não devem ser confundidos com Governo Aberto. Enquanto aquele (dados abertos governamentais) refere-se a dados disponibilizados pelo Governo com facilitadores para reaproveitamento, tratamento, atendimento de critérios técnicos de formato aberto e com a participação da sociedade na escolha dos dados que devem ser abertos, este (Governo aberto) é um status adquirido por governos que assumem os 04 compromissos e se tornam membros Open Government Partnership – OGP.

O Brasil é reconhecido como um protagonista no cenário internacional no que diz respeito a governo aberto, tendo sido convidado para compor, juntamente com outros países, a OGP, à qual foi o primeiro a presidir, juntamente com os Estados Unidos.

Em 2011 o Brasil firmou o compromisso de Governo Aberto e passou a ser membro da parceria de Governo Aberto. O documento de compromisso assinado pelos representantes brasileiros tem em seu último parágrafo um fechamento enfático de assunção de compromisso:

“Comprometemo-nos a adotar estes princípios em nossos compromissos internacionais e a trabalhar para cultivar uma cultura global de governo aberto que dê autonomia aos cidadãos e lhes apresente resultados e avance os ideais do governo

aberto e participativo do século XXI”. (OPEN GOVERNMENT PARTNERSHIP, 2011).

Para serem considerados dados abertos governamentais, devem também atender aos 8 princípios definidos no Encontro de Sebastopol em 2007 (veja na seção “Os Princípios dos dados abertos – encontro de Sebastopol”)

Cabe destacar que o comando normativo sobre dados abertos é o Decreto 8.777/2016, que Institui a Política de Dados Abertos do Poder Executivo federal (BRASIL, 2016), define que os dados devem ser acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte; enquanto os ditames sobre transparência de informação vêm da já citada Lei de Acesso à Informação - LAI, Lei 12.527/2011 (BRASIL, Lei, 2011a).

3.6 Compromisso do Governo Brasileiro com a OGP

O principal ato normativo publicado a partir da assinatura da declaração de Governo Aberto, em que o Governo brasileiro se comprometeu na parceria OGP, foi a instituição do Plano de Ação Nacional sobre Governo Aberto por meio do Decreto S/N (BRASIL, Lei, 2011) de 15 de setembro de 2011. Nesse Decreto, em seu Art. 1º são listadas as diretrizes para promoção de ações e medidas que legitimam a assunção dos 4 compromissos com a OGP, ou seja, um total alinhamento por meio do Decreto:

Art. 1º Fica instituído o Plano de Ação Nacional sobre Governo Aberto destinado a 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, que serão pautadas, entre outras, pelas seguintes diretrizes:

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”.

O Decreto 8.777/2016 detalha o Plano de Ação Nacional sobre Governo Aberto e o Inciso III do Art. 1º é derivado nos seguintes objetivos:

Art. 1º Fica instituída a Política de Dados Abertos do Poder Executivo federal, com os seguintes objetivos:

V - fomentar o controle social e o desenvolvimento de novas tecnologias destinadas à construção de ambiente de gestão pública participativa e democrática e à melhor oferta de serviços públicos para o cidadão;

VI - fomentar a pesquisa científica de base empírica sobre a gestão pública;

VII - promover o desenvolvimento tecnológico e a inovação nos setores público e privado e fomentar novos negócios. (BRASIL, Decreto, 2011).

Ao incluir esses objetivos no Decreto o Governo demonstra a intenção de estimular organizações, cidadãos, acadêmicos, empresas e instituições públicas na criação de soluções capazes de transformar os dados governamentais em plataformas e produtos com valor social e econômico, utilizando os dados abertos.

E é por meio de dados com qualidade, íntegros, seguros, confiáveis e com tecnologia capaz de serem disponibilizados aos seus públicos-alvo é que esse objetivo pode ser alcançado, além de o Governo utilizar e apoiar o uso de tecnologias que possibilitem que esses dados possam ser utilizados e alcancem os resultados esperados, seja em processos decisórios, uso e integração com outros dados do próprio Governo ou de outros Estados, ou para o exercício do papel de fiscalização dos atos da Administração pelo cidadão.

Além disso, ao disponibilizar os dados o Governo estimula a criação de novos negócios e o desenvolvimento de soluções inovadoras, uma vez que a sociedade privada tem mais capacidade e agilidade para desenvolver soluções que o próprio governo, somando-se o fato de que a sociedade pode construir soluções para produção e compartilhamento de conhecimento e serviços de acordo com sua necessidade. Há na verdade uma cooperação de esforços e interesses entre o ente privado e governo na oferta de serviços públicos.

Nessa Esteira o compromisso brasileiro com a parceria OGP para ser um governo aberto sinalizou o seguinte com relação a acesso a novas tecnologias para abertura de dados e prestação de contas:

“Ampliar o acesso a novas tecnologias para fins de abertura e prestação de contas.

As novas tecnologias oferecem oportunidades para a troca de informações, participação pública e colaboração. Pretendemos dominar tais tecnologias para tornar públicas as informações a fim de possibilitar sua compreensão pelo povo sobre as atividades de seu governo e para influenciar decisões.

Comprometemo-nos a criar espaços online acessíveis e seguros como plataformas para prestar serviços, cativar o público e compartilhar informações e ideias. Reconhecemos que o acesso igualitário e fácil à tecnologia representa um desafio e comprometemo-nos a buscar conectividade crescente online e móvel, enquanto também identificamos e promovemos o uso de mecanismos alternativos para participação cívica.

Comprometemo-nos a atrair a sociedade civil e a comunidade empresarial para que identifiquem práticas eficazes e abordagens inovadoras para alavancar novas tecnologias que capacitem as pessoas e fomentem transparência no governo.

Reconhecemos também que o crescente acesso à tecnologia implica o apoio à capacidade do governo e dos cidadãos de utilizá-la. Comprometemo-nos a apoiar e a desenvolver o uso de inovações tecnológicas tanto por funcionários públicos como por cidadãos. Entendemos também que a tecnologia representa um complemento para

informações claras, acessíveis e úteis, não um substituto.” (OPEN GOVERNMENT PARTNERSHIP, 2011)

3.7 Dados e Informações no contexto das Agências reguladoras

As agências reguladoras são entidades autárquicas especiais criadas pelo Governo com o objetivo institucional de controlar os responsáveis pela prestação de serviços públicos, sob a forma de concessão, permissão ou autorização, além de intervir no domínio econômico, quando necessário, para evitar abusos dos entes regulados provenientes da iniciativa privada.

O Legislador, ao admitir a criação de agências reguladoras, visava resolver questões de falhas de mercado, em especial as relacionadas à assimetria de informação. Assimetria de informação é quando dois ou mais agentes econômicos estabelecem entre si uma transação, sendo que uma das partes detém informações superiores aos da outra parte.

Assimetria de informação é quando dois ou mais agentes estabelecem entre si uma transação sendo que uma das partes envolvidas detém informações superiores às demais. A assimetria de informações está presente sempre que se faz preciso tomar decisões em relacionamentos em que uma parte tem mais, ou melhor, informação que a outra. Uma das modalidades de assimetria informacional trata do problema do tipo agente-principal, também conhecido como dilema da agência, que cuida das dificuldades que surgem em condições de informação incompleta e assimétrica quando uma das partes, chamada de “principal”, relaciona-se com outra, chamada de “agente”, em um contexto que impõe custos e ônus para o agente, sem que o principal tenha certeza em relação às características ou ações do agente. Dessa forma, torna-se difícil para o principal garantir que o comportamento do agente esteja em conformidade com seus interesses. (ANP, 2015).

No que tange a regulação, o acesso a informação é vital, pois é por meio dela que os órgãos reguladores podem tomar ações que estimulem a firma regulada a operar de forma eficiente e dentro de parâmetros contratuais e legais.

Para que o agente regulador consiga equilibrar o mercado (objeto de sua atuação) e estimular a eficiência do prestador dos serviços públicos, ele necessita de informações técnicas e de atividades econômico-financeiras e operacionais das firmas.

As melhores fontes de obtenção dessas informações são as próprias firmas reguladas, pois elas são as detentoras das informações consideradas relevantes. Às empresas é interessante ocultar as informações a seu respeito, de forma que suas concorrentes não se beneficiem dessas informações, além de possibilitar aumento de seus lucros pelo desconhecimento, por parte do órgão regulador, de seus reais custos.

Em contraposição, cabe ao ente público regulador aplicar a lei de forma a garantir que, além das informações sobre seus próprios atos, os dados e informações que envolvem as operações das empresas reguladas sejam também disponibilizados à sociedade.

Cabe lembrar que acesso a informação clara, irrestrita, atualizada, completa e tempestiva, é um direito do usuário ou consumidor dos serviços, que tem previsão explícita em diversos instrumentos normativos, alguns exemplos:

Lei nº 8.987, de 13 de fevereiro de 1995 (BRASIL1995) Art. 7º: “Sem prejuízo do disposto na Lei no 8.078, de 11 de setembro de 1990, são direitos e obrigações dos usuários: [...] II - receber do poder concedente e da concessionária informações para a defesa de interesses individuais ou coletivos”;

Lei 12.527, de 18 de novembro de 2011 (BRASIL, Lei, 2011), Art. 3º:

“Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes: I - observância da publicidade como preceito geral e do sigilo como exceção; II - divulgação de informações de interesse público, independentemente de solicitações”;

Decreto 8.777, de 11 de maio de 2016 (BRASIL, 2016), Art. 1º:

“Fica instituída a Política de Dados Abertos do Poder Executivo federal, com os seguintes objetivos: [...] III - franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso”

Decreto 8.638, de 15 de janeiro de 2016 (BRASIL, 2016a), Art. 1º:

“Fica instituída a Política de Governança Digital para os órgãos e as entidades da administração pública federal direta, autárquica e fundacional, com as seguintes finalidades: [...] III - assegurar a obtenção de informações pela sociedade, observadas as restrições legalmente previstas.”

3.8 Governança de dados e informações no setor público

Antes de discorrer sobre a forma como é o sistema de governança de dados e informações no setor público, convém explanar sua fundamentação, práticas, recomendações e pilares normativos.

A Administração Pública rege-se pelo princípio da legalidade, ou seja: só é permitido fazer o que a lei expressamente autoriza. Nesse prisma, a construção da proposta encontra força normativa no princípio da eficiência o qual, num grau não inferior aos demais, vem ditar ao gestor público o dever de celeridade, eficácia, economicidade, efetividade e qualidade por ocasião da concretização de seus atos administrativos.

“O princípio da eficiência apresenta, na realidade, dois aspectos: pode ser considerado em relação ao modo de atuação do agente público, do qual se espera o melhor desempenho possível de suas atribuições, para lograr os melhores resultados; e em relação ao modo de organizar, estruturar, disciplinar a Administração Pública, também com o mesmo objetivo de alcançar os melhores resultados na prestação do serviço público.” (DI PIETRO, 2002, p. 83).

Além dos princípios constitucionais, outros instrumentos surgiram para fortalecer a governança pública, entre eles:

(a) o Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal - Decreto 1.171, de 22 de junho de 1994 (BRASIL, 1994) - e a Lei de Responsabilidade Fiscal - Lei Complementar 101, de 4 de maio de 2000 (BRASIL, 2000) -, que têm por objeto aspectos éticos e morais e o comportamento da liderança;

(b) o Programa Nacional de Gestão Pública e Desburocratização (GesPública), instituído em 2005 e revisado em 2009 e em 2013, cujos treze fundamentos norteiam-se pelos princípios constitucionais da administração pública e pelos fundamentos da excelência gerencial contemporânea (BRASIL, 2005);

(c) a Lei 12.813, de 16 de maio de 2013 (BRASIL, 2013), que dispõe sobre o conflito de interesses no exercício de cargo ou emprego do Poder Executivo Federal;

(d) os instrumentos de transparência, como a Lei de Acesso à Informação - Lei 12.527, de 18 de novembro de 2011 (BRASIL, 2011a) -, que asseguram o direito fundamental de acesso à informação e facilitam o monitoramento e o controle de atos administrativos e da conduta de agentes públicos;

(e) Lei Complementar 131 de 2009, que determina a disponibilização, em tempo real, de informações digitais pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios;

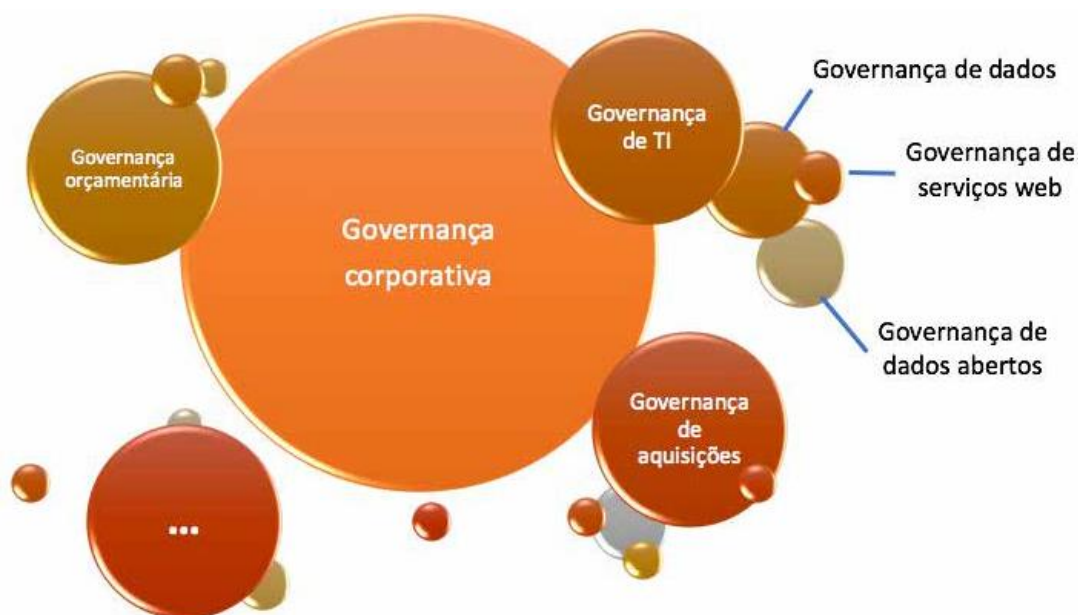
(f) Decreto 8.777 de 11 de maio de 2016, que institui a política de dados abertos;

e

(g) Decreto 8.638 de 15 de janeiro de 2016, que institui a política de governança digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. (BRASIL, 2016a)

A essência dos conceitos difundidos para a boa governança, de forma ampla, pode ser utilizada em outras vertentes de governança, por exemplo em governança corporativa, governança de TI, governança de aquisições, governança orçamentária, governança de dados, governança de informações digitais, e outras mais, entretanto todas agindo harmonicamente entre si:

Figura 2 - Diferentes níveis de governança “gravitando” simultaneamente



Fonte: (BRASIL, TCU, 2016, p. 137)

O documento, GOVERNANÇA PÚBLICA: Referencial Básico de Governança Aplicável a Órgãos e Entidades da Administração Pública e Ações Indutoras de Melhoria (BRASIL, TCU, 2014), sintetizou vários conceitos para a compreensão e a melhoria da governança e da gestão no contexto do setor público, que resultou na seguinte formulação:

“Governança no setor público compreende essencialmente os mecanismos de liderança, estratégia e controle postos em prática para avaliar, direcionar e monitorar a atuação da gestão, com vistas à condução de políticas públicas e à prestação de serviços de interesse da sociedade.”

Nessa mesma publicação o tribunal (BRASIL, TCU, 2014 abud IFAC, 2013), fundamentado por estudo realizado pela IFAC - International Federation of Accountants, defende que a boa governança permite:

- a. garantir a entrega de benefícios econômicos, sociais e ambientais para os cidadãos;
- b. garantir que a organização seja, e pareça, responsável para com os cidadãos;
- c. ter clareza acerca de quais são os produtos e serviços efetivamente prestados para cidadãos e usuários, e manter o foco nesse propósito;
- d. ser transparente, mantendo a sociedade informada acerca das decisões tomadas e dos riscos envolvidos;
- e. possuir e utilizar informações digitais de qualidade e mecanismos robustos de apoio às tomadas de decisão;

- f. dialogar com e prestar contas à sociedade;
- g. garantir a qualidade e a efetividade dos serviços prestados aos cidadãos;
- h. promover o desenvolvimento contínuo da liderança e dos colaboradores;
- i. definir claramente processos, papéis, responsabilidades e limites de poder e de autoridade;
- j. institucionalizar estruturas adequadas de governança;
- k. selecionar a liderança tendo por base aspectos como conhecimento, habilidades e atitudes (competências individuais);
- l. avaliar o desempenho e a conformidade da organização e da liderança, mantendo um balanceamento adequado entre eles;
- m. garantir a existência de um sistema efetivo de gestão de riscos;
- n. utilizar-se de controles internos para manter os riscos em níveis adequados e aceitáveis;
- o. controlar as finanças de forma atenta, robusta e responsável; e
- p. prover aos cidadãos dados e informações digitais de qualidade (confiáveis, tempestivas, relevantes e compreensíveis).

Cabe observar a atenção quanto a dados e informações digitais (item p) dentro de um modelo de governança. Ele não só tem a missão de garantir que a estratégia esteja bem orquestrada como também tem foco na transparência de seus atos de forma confiável, tempestiva, relevante e compreensível.

Um sistema de governança no setor público, para alcançar a boa governança, é composto por vários atores agindo de forma integrada para consecução de objetivos confluentes. São também atividades, processos, comunicação e outros elementos executados por funções segregadas (governança e gestão), entretanto alinhadas:

“O sistema de governança reflete a maneira como diversos atores se organizam, interagem e procedem para obter boa governança. Envolve, portanto, as estruturas administrativas (instâncias), os processos de trabalho, os instrumentos (ferramentas, documentos etc), o fluxo de informações digitais e o comportamento de pessoas envolvidas direta, ou indiretamente, na avaliação, no direcionamento e no monitoramento da organização.” (BRASIL, TCU, 2014, p. 27)

As funções do sistema de governança (governança e gestão) possuem papéis e responsabilidades distintos, de forma a proporcionar eficiência na execução operacional, do dia-a-dia, com o controle necessário. Esse modelo fornece maior transparência e segurança às decisões, fluidez no processo de comunicação e clareza dos papéis de cada participante.

De acordo com o Banco Mundial, governança diz respeito a estruturas, funções, processos e tradições organizacionais que visam garantir que as ações planejadas (programas) sejam executadas de tal maneira que atinjam seus objetivos e resultados de forma transparente (WORLD BANK, 2013). Busca, portanto, maior efetividade (produzir os efeitos pretendidos) e maior economicidade (obter o maior benefício possível da utilização dos recursos disponíveis) das ações. (TCU, 2014, p. 31)

Ainda de acordo com a Corte de Contas, são funções da governança:

- a) definir o direcionamento estratégico;
- b) supervisionar a gestão;
- c) envolver as partes interessadas;
- d) gerenciar riscos estratégicos;
- e) gerenciar conflitos internos;
- f) auditar e avaliar o sistema de gestão e controle; e
- g) promover a accountability (prestação de contas e responsabilidade) e a transparência.

Utilizando-se dos conceitos apresentados pelo Tribunal de Contas da União: de modo complementar, gestão diz respeito ao funcionamento do dia a dia de programas e de organizações no contexto de estratégias, políticas, processos e procedimentos que foram estabelecidos pelo órgão; preocupa-se com a eficácia (cumprir as ações priorizadas) e a eficiência das ações (realizar as ações da melhor forma possível, em termos de custo-benefício). Ela é inerente e integrada aos processos organizacionais, sendo responsável pelo planejamento, execução, controle, ação, enfim, pelo manejo dos recursos e poderes colocados à disposição de órgãos e entidades para a consecução de seus objetivos.

São funções da gestão:

- a) implementar programas;
- b) garantir a conformidade com as regulamentações;
- c) revisar e reportar o progresso de ações;
- d) garantir a eficiência administrativa;
- e) manter a comunicação com as partes interessadas; e
- f) avaliar o desempenho e aprender

Figura 3 - Relação entre governança e gestão



Fonte: (BRASIL, TCU, 2014, p. 32)

As funções do sistema de governança (governança e gestão) possuem papéis e responsabilidades distintos, de forma a proporcionar eficiência na execução operacional, do dia-a-dia, com o controle necessário. Esse modelo fornece maior transparência e segurança às decisões, fluidez no processo de comunicação e clareza dos papéis de cada participante.

Já a governança da informação surge em momento de supervalorização da informação como um ativo. Cada vez mais empresas estão suportando seus negócios e processos e apoiando seus objetivos com base em informações. Daí a necessidade de ter informações confiáveis, íntegras e com qualidade. Em essência, seu objetivo é assegurar informações confiáveis às organizações.

De acordo com Davenport (1998, p. 13 a 15) com o aumento da importância da informação, há uma necessidade de se pensar além das máquinas, e focado no objetivo principal da informação: informar as pessoas. Governança da informação é um modelo que fornece às organizações uma forma de lidar com esse desafio.

Em artigo publicado pela EnAnpad (LAJARA & MAÇADA, 2013), encontramos uma adequada definição de governança de informação, a qual subsidiará à frente o conceito de governança de informações digitais:

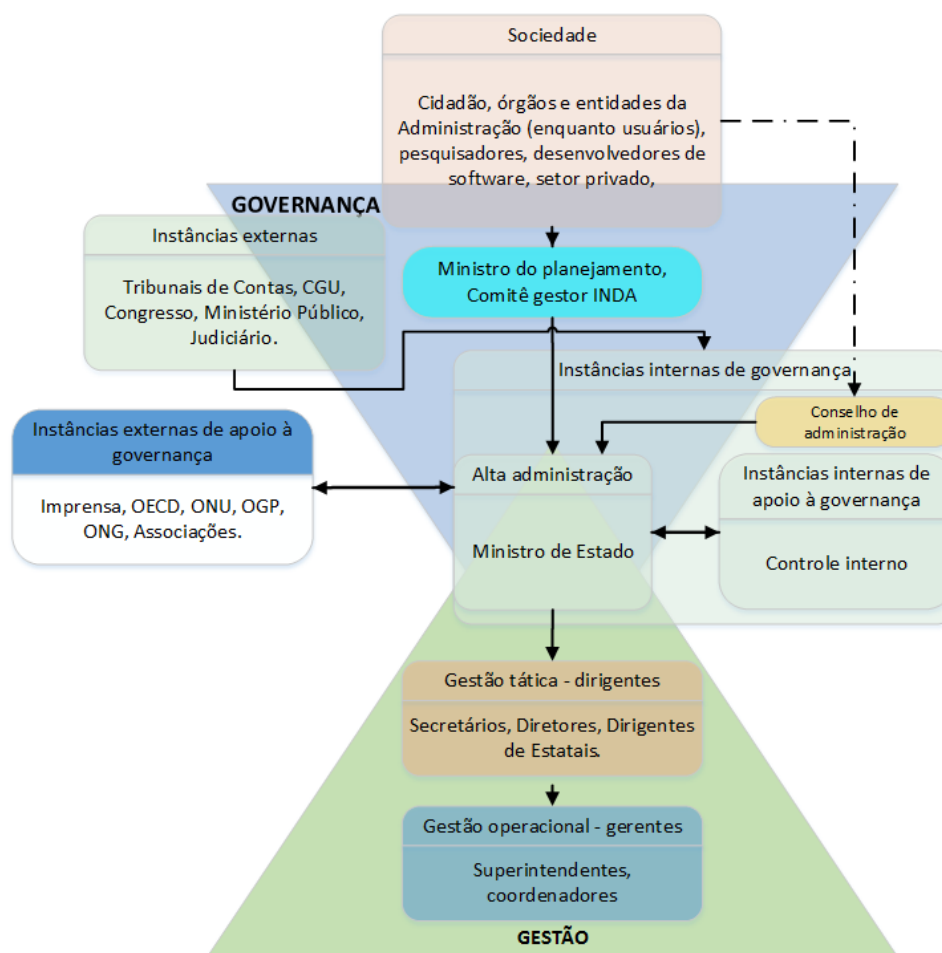
Alinhada com a governança corporativa e a governança de Tecnologia da Informação (TI), a governança de informação é um conjunto de normas, diretrizes e controles de responsabilidades desenvolvido para assegurar o valor, a qualidade e o compliance das informações. Com essas dimensões, a gestão da informação fornece normas, diretrizes e controles de responsabilidade para empresas que estão enfrentando problemas como a explosão da informação, para que essas organizações possam obter e garantir a qualidade, valor e compliance de seu ambiente informacional. (LAJARA & MAÇADA, 2013)

Para o Data Governance Institute (DGI, 2007, p. 1), a governança de dados “é um sistema de direitos decisórios e responsabilidades para processos relacionados com a

informação, executados conforme modelos acordados. Esses processos descrevem quem pode tomar quais ações, com qual informação, e quando, em que circunstâncias, usando que métodos”.

Conforme citado no início desta seção a Administração Pública rege-se pelo princípio da legalidade, assim sendo, só pode atuar nos limites autorizados pela Lei. Portanto, com base nos instrumentos normativos que disciplinam o direito do cidadão à informação, chegamos ao modelo de governança de dados e informações praticado no setor público e que viabiliza a realização de um ecossistema em que as partes interessadas interajam entre si, mas ao mesmo tempo haja clara fronteira de papéis e responsabilidades:

Figura 4 - Sistema de Governança em órgãos e entidades da administração pública



Fonte: (BRASIL, TCU, 2014, p. 28)

3.9 Reflexão sobre dados abertos

Embora as iniciativas de dados abertos tenham demonstrado grande evolução para aproximar a sociedade às ações e atos de governos, possibilitando controle social,

fiscalização, geração de informação com base nesses dados, desenvolvimento de aplicativos, e tantas outras iniciativas que contribuem para o fortalecimento de democracias, alguns pontos de atenção não podem ser desconsiderados.

Questões como complexidade técnica das bases de dados, tratamento a posteriori dos dados abertos, custos e continuidade da disponibilização de dados e privacidade e proteção dos dados pessoais são temas em discussão em alguns segmentos. Sobre esses assuntos o TCU (BRASIL, TCU, 2014) realizou fiscalização com objetivo de conhecer iniciativas de publicação de dados abertos na Administração Pública Federal (APF), considerando ainda o uso atual de soluções de coleta, armazenamento e processamento de grandes volumes de dados (Big Data) pelo governo e sociedade como estratégias de transparência e modernização da gestão pública.

Nesse trabalho o Tribunal verificou fatores inibidores do uso de dados abertos, um deles está relacionada à complexidade técnica das bases de dados. Essa dificuldade pode ser um fator inibidor de uso e de exploração de dados abertos. Para o órgão é um dos principais desafios às políticas de dados abertos. Esse mesmo levantamento destacou também problemas de qualidade dos dados disponibilizados, dados incompletos e desatualizados como desafios a serem vencidos.

Outro estudo partícipe do Acórdão do TCU, subsidiariamente relacionado à questão da complexidade técnica das bases de dados, publicado pela Association of Government Accountants (AGA), destaca duas “escolas de pensamento” em transparência.

A primeira pensa dados no seu formato bruto e em forma granular, disponibilizados na web, de forma que os usuários possam criar seus próprios relatórios. Eles acreditam que dados disponibilizados de forma agregada contém o viés da entidade governamental que realizou as agregações.

A segunda escola acredita que os dados deveriam ser agregados e apresentados de forma amigável e compreensível, além de serem distribuídos por diversos meios e mídias, inclusive a internet. Esse grupo acredita que o “dado” não organizado não é genuinamente “informação”.

Para a AGA, os dois caminhos são válidos e deveriam existir esforços nos dois sentidos por parte do governo: o de fornecer dados primários e também dados analíticos.

Outro fator dificultador, antagonicamente ao conceito de abertura de dados, diz respeito ao seu excesso, ou simplesmente a falta de sentido dos dados disponibilizados. Esse problema é denominado por Andrea DiMaio, do Gartner Group, em seu artigo Open Data and the New Divide (BRASIL, TCU, 2014 apud DIMAIO, 2012), de “dilúvio dos dados abertos”, e o risco de “afogamento” do usuário devido ao volume avassalador dos dados brutos

disponíveis ou simplesmente devido a sua completa falta de sentido quando apresentados sem a preparação adequada.

A questão observada pelo TCU (BRASIL, TCU, 2014), a partir do apontamento de DiMaio diz respeito a: quem deve processar esses dados para que eles façam sentido? O próprio governo? seria a sociedade organizada, por meio dos seus ativistas, militantes, hackers e suas associações não governamentais? Uma importante questão a ser considerada é a interpretação por conveniência por parte de quem processa os dados. A falta de isenção a depender do interesse do ator pode levar a risco de interpretações errôneas ou direcionadas ao convencimento de determinados públicos. Esse fenômeno o autor chamou de “o lado negro dos dados abertos”, o que pode ser uma forte divisão entre quem tem e quem não tem recursos e habilidades para manipular os dados abertos. Conclui-se que a capacidade de processamento é um diferencial de acordo com a disponibilidade de recursos tecnológicos por parte de quem processará os dados disponibilizados pelo governo.

Embora as tecnologias de consumo se tornem cada vez mais poderosas e os recursos de processamento massivos se tornem disponíveis como uma commodity, dar sentido a dados grandes e abertos não é para os fracos, e exigirá investimentos significativos para os tempos vindouros. (DIMAI, 2012)

O cidadão comum tem acesso diferenciado a hardware e software necessários para baixar e processar conjuntos de dados abertos, bem como níveis variados de habilidades necessárias para analisar, contextualizar e interpretar os dados. Diferentemente dos recursos dotados aos poderosos. Rob Kitchin (Kitchin, 2013) destaca o potencial efeito dos dados abertos no sentido de municiar os poderosos: embora a intenção intrínseca à transparência seja o fortalecimento da democracia e o acesso igual e indiscriminado aos dados públicos, a força de grupos e empresas, já preparadas e dotadas de recursos tecnológicos para a análise e exploração de grandes quantidades de dados, faz com que o resultado a partir dos dados abertos, favoreça segmentos detentores de maiores recursos tecnológicos TCU (BRASIL, TCU, 2014).

Sobre a preocupação com a proteção dos dados pessoais, Varun Singh, Ishan Srivastava e Vishal Johri (2014) demonstram no artigo “Big Data and the Opportunities and Challenges for Government Agencies” desafios relacionados a privacidade e segurança. O papel da salvaguarda e privacidade dos dados dos cidadãos é de responsabilidade do Governo dentro dos ditames e limites legais. O uso de Big Data introduz mais um elemento para ser considerado no gerenciamento de risco da segurança da informação, que são os dados relacionadas a privacidade e intimidade de dados pessoais. Além da informação armazenada, deve-se atentar para as fontes de dados e os mecanismos de troca de dados entre órgãos que podem ser alvos de hackers.

Eles ainda recomendam uma análise cuidadosa com relação aos dados disponibilizados em Big Data. Muitos governos têm políticas de governança abertura de dados, sob as quais podem liberar uma grande quantidade de dados legíveis por máquina. Além disso, os Big Data agregam valor combinando diversos conjuntos de dados relevantes, diferentes, que podem ser vinculados e analisados como um todo para revelar novos padrões, tendências e insights. Mas antes de empregar as técnicas de análise de big data, o governo deve demonstrar a seus cidadãos que isso pode ser realizado sem comprometer seus direitos de privacidade.

Já o TCU (BRASIL, TCU, 2014) demonstra preocupação com questões relacionadas a anonimização de dados pessoais, que são as práticas de ofuscação de dados que permitam a identificação individualizada. Para o Tribunal a “reidentificação” desses dados em muitos casos é possível por meio de triangulação de informações até que se encontre coincidências tão restritivas que acabem isolando as informações de um indivíduo em meio aos demais. Para o Tribunal convém amplo debate antes de disponibilização de bases de dados “anonimizadas”, assim como com relação às tecnologias envolvidas no processo.

No Brasil o assunto é regulamentado recente publicada Lei 13.709 de 14 de agosto de 2018 (BRASIL, 2018) que dispõe sobre a proteção de dados pessoais. Ela é fundamentada em diversos valores, como o respeito à privacidade; à autodeterminação informativa; à liberdade de expressão, de informação, de comunicação e de opinião; à inviolabilidade da intimidade, da honra e da imagem; ao desenvolvimento econômico e tecnológico e a inovação; à livre iniciativa, livre concorrência e defesa do consumidor e aos direitos humanos liberdade e dignidade das pessoas.

Conforme registrado no sítio da Wikipédia (2018), um dos motivadores da aprovação do projeto de lei foram os escândalos de privacidade do Facebook – em que a empresa Cambridge Analytica utilizou de dados dos usuários para que pudessem fazer uma campanha política mais assertiva e customizada na eleição de Donald Trump em 2016 – também trouxeram visibilidade para o assunto. A segurança de dados e a privacidade passaram a ser pautas recorrentes, recaindo uma cobrança sobre os políticos brasileiros no sentido de tratar dessas questões, já que não havia no país legislação com objetivo específico de defender os dados dos usuários e definir responsabilidades relativas ao tratamento destes. Além desse fator, houve também investigações no Brasil sobre um possível comércio entre órgãos públicos, onde o produto eram os dados das pessoas, sem que houvesse consentimento dos donos.

A Lei prevê ainda a responsabilização de qualquer empresa ou órgão público que atue no tratamento de dados pessoais que violar a legislação, com sanções a serem aplicadas após procedimento administrativo que possibilite a oportunidade da ampla defesa, que vão

desde advertência até o pagamento de multa, porém não substitui a aplicação de sanções administrativas, civis ou penais definidas em legislação específica.

Além da Lei 13.709 (BRASIL, 2018), de forma explícita a Lei de Acesso a Informação (BRASIL, LEI, 2011), corroborando com aquela Lei, o traz restrições ao uso de informações pessoais:

Art. 31. O tratamento das informações pessoais deve ser feito de forma transparente e com respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais.

§ 1º As informações pessoais, a que se refere este artigo, relativas à intimidade, vida privada, honra e imagem:

I - terão seu acesso restrito, independentemente de classificação de sigilo e pelo prazo máximo de 100 (cem) anos a contar da sua data de produção, a agentes públicos legalmente autorizados e à pessoa a que elas se referirem; e

II - poderão ter autorizada sua divulgação ou acesso por terceiros diante de previsão legal ou consentimento expresso da pessoa a que elas se referirem.

§ 2º Aquele que obtiver acesso às informações de que trata este artigo será responsabilizado por seu uso indevido. (BRASIL, LEI, 2011).

A partir da publicação dessa Lei o Brasil passa a fazer parte dos países que contam com legislação específica para proteção de dados e da privacidade dos seus cidadãos. Outros regulamentos similares ao LGPD são o General Data Protection Regulation (GDPR) na União Europeia (UE, 2018), que passou a ser obrigatório em 25 de maio de 2018 e aplicável a todos os países da União Europeia (UE) e o California Consumer Privacy Act of 2018 (CCPA), nos Estados Unidos da América, implementado através de uma iniciativa em âmbito estadual, na Califórnia, onde foi aprovado no dia 28 de junho de 2018 (AB 375).

Apesar das sugestões de melhorias e críticas, a realidade é que a transparência e publicação de dados no Governo Brasileiro ainda tem um longo a percorrer, é o que nos mostra o resultado do Índice de Dados Abertos para Brasil (FGV, 2017). Esse trabalho subsidia um estudo global sobre dados abertos conduzido pela Open Knowledge Foundation.

O Brasil teve um bom desempenho no índice quando comparado aos demais países. No entanto, a pontuação de 65% indica que ainda há muito espaço para aprimoramentos. Apenas 40% das dimensões do índice receberam a nota total e três delas nem mesmo existem publicamente. A maior parte dos problemas identificados está mais relacionada à usabilidade (completude, atualização e formato) do que a processo (meios de acesso e licença de uso dos dados).

O problema mais recorrente é a dificuldade de se trabalhar os dados e, em segundo lugar, a não disponibilização da base de dados completa para download. Isso indica que os órgãos responsáveis pelo provimento das informações talvez não tenham a compreensão de que a facilitação do processo de edição e manipulação dos dados é também sua responsabilidade e parte importante da transparência. (FGV, 2017)

O estudo da FGV avaliou 15 dimensões de bases de dados no nível Federal, sendo elas relacionadas a informações sobre: Orçamento Público, Resultados Eleitorais, Mapas

Nacionais, Estatísticas Socioeconômicas, Leis em vigor, Atividade Legislativa, Gastos Públicos, Limites Administrativos, Previsões Climáticas, Compras Públicas, Registro de Empresas, Qualidade do Ar, Localizações, Qualidade da Água, Propriedade da Terra.

O objetivo do índice é avaliar o estado da política de dados abertos, levando em consideração todas as suas características: quais dados são divulgados sobre determinado assunto, como eles são divulgados em relação ao acesso, ao formato, à granularidade, aos custos e ao ritmo de sua atualização.

Nessa avaliação Brasil está posicionado em um escore de 65%, 9º no do ranking, uma colocação antes dos Estados Unidos, e o primeiro dentre os países da América Latina.

A avaliação das dimensões considerou uma metodologia de contagem de problemas encontrados relacionados a usabilidade e processo. Problemas de usabilidade estão vinculados a dificuldade de uso do dado em si e à capacidade de transformação deste em informação. Como questões de processo relacionam-se com dificuldades de acessar o dado e omissão/inadequação da licença.

A tabela abaixo sumariza o estudo sob a ótica dos problemas e apresenta de imediato dois problemas que se destacam: a inexistência da informação e a restrição do acesso a ela:

Figura 5 - Frequência de problemas encontrados no Brasil

GARGALOS MAPEADOS	FREQUÊNCIA DE IDENTIFICAÇÃO DOS GARGALOS
Dificuldade de trabalhar dados	9
Download da base completa indisponível	5
Dificuldade de localizar dados	3
Indisponibilidade de formato aberto	3
Dataset desatualizado	3
Licença não transparente	2
Dataset incompleto	2
Restrição de Acesso	1

Fonte: FGV (2017) Índice de Dados Abertos no Brasil

Corroborando com esse resultado, outra pesquisa também conduzida pela FGV, Índice de Dados Abertos para Cidades 2018 (FGV, 2018), contemplando metodologia semelhante envolvendo 8 cidades brasileiras, apresenta números convergentes:

Apenas 25% das bases de dados avaliadas estão 100% de acordo com a definição de dados abertos. De todo o universo de obstáculos encontrados (429), 62% são problemas de usabilidade e 38% de processo.

Os problemas mais comuns das bases de dados são: dificuldade de trabalhar dados (incluindo os metadados insuficientes), indisponibilidade de download da base de dados completa, dataset incompleto e ausência da informação em formato aberto.

Uma reflexão mais apurada da metodologia de ambos os relatórios da FGV traz-nos a uma inferência pessimista: não é escopo desses trabalhos outros aspectos comuns em avaliação de dados abertos, como contexto, uso, impacto ou qualidade de mensuração dos dados. O que nos leva a pensar que o escore de 9ª posição, à frente dos Estados Unidos, não atende plenamente aos compromissos de dados abertos (ver seção “Governo aberto, dados abertos governamentais e os compromissos com a OGP”), uma vez que uma boa avaliação sob a ótica procedimentos e processos para abertura de dados (avaliado em 65%) não garante a mesma satisfação no alcance de resultados de efetividade.

Essa reflexão é pertinente, pois a abertura de dados não é um processo em si só. Não é um ato isolado nem sequer uma iniciativa de gestão tecnológica ou afeta a uma área de tecnologia.

As conclusões apresentadas pela FGV foram vivenciadas na prática por Francisco Nauber Bernardo Gois, Analista do Serpro, ao realizar uma prova de conceito com o objetivo de demonstrar que é possível aplicar as técnicas apresentadas em dados governamentais agregando valor ao governo brasileiro e à sociedade. (GOIS, 2017)

Entretanto, Nauber esbarrou com os seguintes problemas ao tentar utilizar dados abertos do Governo Federal:

Apesar de o governo brasileiro disponibilizar diversos dados abertos em diversos portais institucionais (<http://dados.gov.br/>), a falta de metadados adequados, e o uso de formatos proprietários dificultam o uso desses dados em ferramentas de análise de dados. (GOIS, 2017).

Para resolver o problema de forma que conseguisse chegar a termo com seu trabalho, Nauber lançou mão de dados base de dados com informações de todos os crimes na cidade de Chicago desde 2001.

O tipo de situação vivenciada por Francisco Gois, (falta de metadados adequados) não é o único ou o principal tipo de problema aos quais os processos de abertura de dados estão sujeitados. Além dos listados nos trabalhos realizados pela FGV, o Tribunal de Contas da União (BRASIL, TCU, 2014, p. 37), apresentou uma síntese de relatos de órgãos sobre fatores limitadores da disponibilização de dados governamentais, (p. 37):

as dificuldades relatadas pelos órgãos entrevistados foram: sustentabilidade das iniciativas geradas, em especial, das soluções construídas pela sociedade ou de forma colaborativa; falta de pessoal e de governança na gestão do processo; e falta de padronização. (BRASIL, TCU, 2014, p. 37).

A situação apresentada no parágrafo anterior não é de difícil comprovação. Uma rápida pesquisa no portal de dados abertos do Governo Federal (www.dados.gov.br) podemos encontrar alguns exemplos para duas situações:

1^a) conjuntos de dados desatualizados ou sem acessibilidade adequada, como é o caso do BNDES, que tem conjuntos de dados somente até 2015; da CEF, que tem somente 1 conjunto de dados divulgado (indicadores sobre minha casa melhor) e até o ano de 2014; do IBRAM, que disponibiliza manuais de acesso a sistemas, em formato pdf, ao invés dos dados abertos; do TCU, que para acessar os dados é necessário realizar um cadastramento prévio (Dados coletados no Perfil de Governança de TI); e

2^a) aplicações desatualizadas desenvolvidas pela sociedade, como é o caso do Observatório de Votações da Câmara dos Deputados, que encontra-se indisponível; do Basômetro, que não tem dados atualizados e teve foco nas eleições de 2012; do Saúde Acessível, que não retornou informação válida em todos os testes; do Para onde foi o meu dinheiro? que está com dados atualizados até 2016; do Obras do PAC, que o endereço divulgado para acesso ao site apresenta erro.

Observa-se a fragilidade do processo de abertura de dados, o que, aparentemente, visou apenas conformidade normativa sem ações para continuidade da prestação do serviço.

Machado Junior (2015, p.9) comenta esse tipo de situação, opinando que abertura de dados é mais do que uma questão ou uma perspectiva técnica e deve ser tratada como uma plataforma completa, um ecossistema que inclui componentes políticos, legais, avanços tecnológicos e mecanismos de classificação dos dados segundo diversos critérios.

Para abrir os seus dados, uma instituição deve estar preparada para disponibilizá-los a um público amplo e diverso, de maneira a atender às demandas não só dos cidadãos, mas também dos diversos intermediários de dados. Para isso, o processo de preparação para a abertura dos dados envolve mudanças institucionais, em termos de cultura, metodologia, capacitação (interna à instituição e dos usuários), envolvimento dos cidadãos, infraestrutura

tecnológica e financiamento da inovação. A essa preparação Machado Junior (2015, p.30) intitulou de **prontidão para a abertura de dados**.

3.10 Tecnologias disruptivas para uso com dados abertos

A abertura de dados do governo traz consigo potenciais diversos, como a melhoria da gestão pública, provimento da transparência, o fomento ao controle social, a promoção da participação social e o fomento a novos negócios e inovação. É perceptível e inegável que a quantidade de dados custodiados ao governos brasileiro é de dimensões (arrisco estimar) na ordem de petabytes (um petabyte = 10^{15} bytes ou 1 terabyte x 1024), e um fator crítico de sucesso para que se consiga resultados no uso desses dados e com relação à tecnologia a ser usada para processamento de grandes volumes de dados. (WIKIPÉDIA, PETABYTE).

Esse é um momento coetâneo ao surgimento de novas tecnologias capazes de gerar, coletar, processar e disseminar dados eletrônicos em grande volume e alta velocidade. O grande desafio é conhecer e adotar a tecnologia adequada para o fim desejado.

Por questões de limitação de escopo de trabalho, as potenciais tecnologias capazes de suportar dados abertos que serão citadas neste trabalho são: Internet das Coisas e Cidades Inteligentes, Big Data, Data Lake, e Recuperação, recomendação e relevância de informação.

Entretanto, este trabalho não tem a intenção de aprofundar nos conceitos dessas tecnologias, pelo contrário, o objetivo desses próximos tópicos é apresentar uma visão superficial sobre tecnologias que podem agregar valor ao cidadão quando utilizadas com dados abertos.

3.10.1 Internet das Coisa e cidades inteligentes

Em 2005 a International Telecommunications Union – ITU (2005) publicou um relatório com o conceito de Internet das Coisas (ou IoT, do termo *Internet of Things*), com uma visão abrangente e holística: Internet das Coisas é uma infraestrutura global para a sociedade da informação, que habilita serviços avançados por meio da interconexão entre coisas (físicas e virtuais), com base nas tecnologias de informação e comunicação (ITU International Telecommunication Union, 2005). Em sentido amplo, trata-se não apenas de conectar coisas, mas também de dotá-las do poder de processar dados, tornando-as “inteligentes”.

Os componentes de IoT estão relacionados a 1) recebimento de dados digitais vindos de sensores e indo para atuadores (ex. sensor de umidade na terra, enviando dados para atuadores que controlam Pivots de irrigação); conexão a uma rede fora do objeto (exemplo: envio para bases da rede local da Iot, ou para fog computing – computação em névoa ou de

borda, intermediária -, ou para cloud computing – computação em nuvem); e capacidade de processar (ferramentas para previsão e predição com base em dados coletados e armazenados).

Para tornar mais claro o conceito, usarei exemplos (bem reais) e que dão uma perfeita noção do que vem a ser IoT: O trator que não só ara a terra, mas também passa a coletar uma extraordinária quantidade de dados, que serão posteriormente analisados por uma aplicação hospedada em um data center, produzindo relatórios que permitem que um agricultor tome decisões sobre onde e quando plantar; ou dispositivos vestíveis (wearables) que fornecem informações ao sistema médico sobre indicadores relacionados à saúde de pacientes;

Outros exemplos foram listados pelo BNDES no “Relatório de Roadmap Tecnológico” (BNDES, 2017), p.246), que é parte de seu estudo para o diagnóstico e a proposição de plano de ação estratégico para o país em Internet das Coisas:

Construções inteligentes: soluções que atuam na gestão de prédios e construção, comerciais ou residenciais, no intuito de melhorar a eficiência no uso de recursos como energia, água e vagas de estacionamento. Atuam, também, na automatização de funções no âmbito de sua construção, como a refrigeração de ar. Estas soluções demandam conectividade para transmissão dos dados e gerenciamento centralizado;

Gerenciamento de frotas: soluções voltadas para o aumento da eficiência no uso de veículos de frota, como caminhões, ônibus ou automóveis particulares. Atendem veículos autônomos, ou não, e são baseadas em dados de telemetria e sensoriamento enviados por redes sem fio. Estas soluções demandam conectividade nas cidades, estradas ou a utilização de transmissores nos veículos para transmissão dos dados e gerenciamento centralizado;

Gestão de infraestruturas municipais: soluções de telemetria e sensoriamento de infraestrutura das cidades, tais como rede de água, rede de esgoto, energia para centralizar o gerenciamento destas estruturas e gerar análises preditivas de falhas;

Gestão de serviços e meio ambiente no município: soluções específicas de sensoriamento de variáveis associadas à prestação de serviços no âmbito municipal e do controle do meio ambiente. Alguns exemplos são o controle de veículos de distintas frotas de transporte público, com vistas à melhoria da eficiência do sistema de transporte público como um todo, gestão da iluminação pública e monitoramento de condições críticas de poluição;

Gestão da segurança pública e patrimonial (2): soluções de vigilância patrimonial, de edificações públicas ou privadas, e soluções de vigilância de espaços públicos (ruas, praças, etc.) que detectam situações de ameaças por meio da análise de sons (disparos de armas), imagens, vibrações;

Serviços avançados à população: soluções para uso pelos cidadãos por meio de seus aparelhos portáteis móveis (celulares, tablets), baseadas na captura de informações, diversas publicadas como dados abertos por governos, de sensores trânsito; rotas e localização de ônibus e trens; previsão de tempo; turismo e outras.

Expandindo esses exemplos e os conceitos e imaginando-os aplicados em toda uma cidade, contando ainda com governança envolvendo múltiplos atores (iniciativa privada e governo), teremos a aplicação do conceito de “cidade inteligente”.

Para Namrata Nagpal (2017) cidades inteligentes são conjuntos de soluções que conectam vários aspectos do dia a dia, como transporte, energia e edifícios, de uma maneira inteligente e eficaz, melhorando, assim, a qualidade de vida das pessoas da cidade. Os dados

em cidades inteligentes são reunidos em grande abundância a partir dos dispositivos em tempo real, sensores, vídeo / áudio, arquivos de log, redes, web, aplicativos transacionais e mídias sociais. Assim, as cidades inteligentes incorporam IoT, infraestrutura inteligente, conectividade M2M para gerenciar energia, recursos e melhorar o planejamento urbano

Nas cidades inteligentes, os volumes de dados crescem exponencialmente em todos os domínios, corporativo, governamental e global. Elas são geradas a partir de várias fontes, como sistemas corporativos, sistemas de programas de governo, registros de sites, serviços da Web, mídias sociais, dispositivos móveis, sensores, GPS, etc. e de várias formas - estruturada, semiestruturada e não estruturada. Tornou-se extremamente importante entender os conjuntos de dados e organizá-los. (KUMAR & PRAKASH, 2013)

Outro conceito foi formulado por Khan Z (Khan Z,2013) como “sendo uma cidade que investe em governança aprimorada de TIC e processos participativos para definir investimentos adequados em serviços públicos e transporte que possam garantir desenvolvimento socioeconômico sustentável, melhor qualidade de vida e gerenciamento inteligente de recursos naturais”.

Ora, vê-se que os conceitos apresentados sobre cidades inteligentes nem são restritos a soluções específicas nem são exclusivos para um setor ou setores específicos. Muito pelo contrário: prevê a participação do cidadão, da iniciativa privada e do governo. De acordo com o Cities in Motion Index, do IESE Business School na Espanha (IESE, 2018), 10 dimensões indicam o nível de inteligência de uma cidade: governança, administração pública, planejamento urbano, tecnologia, o meio-ambiente, conexões internacionais, coesão social, capital humano e a economia.

Essa mesma publicação demonstra o envolvimento e a necessidade de que todas as partes interessadas se engajem no desenvolvimento de um ecossistema e seja possível aproveitar os benefícios das cidades inteligentes:

Cidades inteligentes geram inúmeras oportunidades de negócios e possibilidades de colaboração entre os setores público e privado. Todas as partes interessadas podem contribuir, por isso, deve ser desenvolvida uma rede de ecossistemas que envolva todos eles: membros do público, organizações, instituições, governo, universidades, especialistas, centros de pesquisa, etc.

Percebe-se o papel essencial do governo como ator na governança de cidades inteligentes, e discorreremos como se dá o envolvimento do governo no contexto de transparência e de governo aberto, então as iniciativas de cidades inteligentes, que têm o governo como um dos stakeholders e como ator na prestação de serviços, indubitavelmente este deverá atender aos ditames de abertura de dados, seus princípios, diretrizes e objetivos.

3.10.2 Big Data

“In God we trust, all others must bring data”

W. Edwards Deming

O termo Big Data é usado para descrever um grande volume de dados estruturados e não estruturados, por ser tão grande e complexo, torna-se difícil gerenciar e processar usando softwares e ferramentas tradicionais de banco de dados.

A expressão Big Data é frequentemente utilizada em sentido lato para designar quaisquer grandes bases de dados, cuja análise escapa aos métodos tradicionais em banco de dados relacionais e em estatística descritiva ou que requeiram grande capacidade computacional distribuída em numerosas máquinas para ser executada. A estatística descritiva é a parte que procura descrever e avaliar características de uma amostra da população sem generalizar conclusões ou inferências sobre a população como um todo.

De acordo com Kumar and Prakash (2013) Big Data são ativos de informações de grande volume, velocidade e variedade que exigem formas inovadoras e econômicas de processamento de informações para permitir melhor insight e tomada de decisões. Eles têm as seguintes características:

Volume: Volume refere-se à quantidade de dados.

Variedade: Variedade refere-se a origem e formato dos dados: estruturado e não estruturado. Fonte de dados é diversificada. O Gartner mencionou isso como dados obscuros. Os dados obscuros não podem ser vistos diretamente, mas são a maior parte do universo de dados das organizações.

Velocidade: a velocidade lida com a taxa na qual os dados fluem de várias fontes. O fluxo de Big Data é massivo e contínuo.

Veracidade: a veracidade se refere à incerteza dos dados. Ele lida com qualidade, confiabilidade e precisão dos dados.

O Tribunal de Contas da União o TCU (BRASIL, TCU, 2014) realizou fiscalização com objetivo de conhecer iniciativas de publicação de dados abertos na Administração Pública Federal (APF), nesse trabalho verificou a existência de grande volume de dados públicos de interesse coletivo ou geral (dados passíveis de serem abertos) e de tecnologias que podem oportunizar a melhoria da coleta, armazenamento e processamento de dados em grandes volumes, velocidade e variedade, a exemplo do chamado Big Data, com potencial de geração de conhecimento e insumos para a tomada de decisão.

Benefícios do uso de soluções de Big Data podem ser percebidas em diversos segmentos, entretanto, muitos deles sem que o usuário saiba que está acessando um serviço de Big Data. O trabalho realizado pelo TCU (BRASIL, TCU, 2014) catalogou alguns desses casos, por exemplo:

a) personalização das respostas geradas por sistemas de informação, como nas recomendações de compra da Amazon (amazon.com); que sugerem automaticamente novos produtos em função de comparações de históricos de compras de sua base de clientes;

b) captação de informações dispersas em redes de sensores ou por meio de contribuições individuais voluntárias, como a iniciativa da cidade de Boston, nos Estados Unidos, que lançou um aplicativo para celulares que detecta a vibração causada pela passagem de um carro sobre um buraco na estrada, registra a posição geográfica em que isso ocorreu com o GPS e transmite esses dados pela internet para os servidores do departamento de estradas do município, que pode então alocar seus recursos em resposta à incidência de buracos e seu impacto sobre os motoristas que usam o aplicativo (cityofboston.gov/doi/apps/streetbump.asp);

c) detecção de fraudes: controle aduaneiro do trânsito de mercadorias (fcw.com/Articles/2013/09/25/big-data-transform-government.aspx?p=1);

Há também benefícios no uso de soluções de Big Data em conjunto com dados abertos, como: as avaliações de qualidade e disponibilidade de hospitais e escolas, o que permitiria maior empoderamento da população e eventuais ganhos em participação social.

Outro exemplo Big Data com dados abertos, embora ainda não disponível ao público em geral, é o caso que está em desenvolvimento no Governo Brasileiro, o GovData. Entretanto, a solução ainda é restrito a gestores públicos, e os serviços devem ser contratados no âmbito intra-órgãos do Governo (com o Serpro); os dados que estão no DataLake e que incorporam o Big Data são em essência dados abertos (<http://ck.govdata.gov.br/>), e a tendência natural é que haja uma vertente da solução do GovData para a sociedade.

Entende-se, assim, que as tecnologias de Big Data podem e devem fazer uso de dados abertos no âmbito governamental, seja para informar ao cidadão, gerar conhecimento, promover o controle social ou para qualquer outro fim que favoreça o fortalecimento democrático, o progresso da gestão do Estado e dos serviços que por ele são prestados.

3.10.3 Data Lake

Um Data Lake é uma coleção de instâncias de armazenamento de vários ativos de dados adicionais aos silos de origem das fontes de dados. Esses ativos são armazenados em uma cópia quase exata ou exata do formato de origem. O objetivo de um Data Lake é apresentar uma visão não refinada dos dados destinado a atividades exploratórias e de refinamento e análise de dados. Ele tem como público alvo principal analistas mais qualificados, isso porque ele não incorpora (por conceito) soluções de analytics da mesma forma como concebido em Big Data.

Uma alternativa ao desafio da complexidade do Big Data é a criação de Data Lake, embora ele possa também fazer parte de soluções de Big Data. Data Lake é um repositório que armazena um grande e variado volume de dados, estruturado e não estruturado, com o objetivo de extrair insights e análises inteligentes. Uma definição pode ser encontrada no wikipedia:

Um repositório de dados massivo e de fácil acesso, construído sobre hardware de computador (relativamente) barato para armazenar grandes quantidades de dados. Ao contrário dos data marts, que são otimizados para análise de dados armazenando apenas alguns atributos e descartando dados abaixo da agregação de nível, um data lake é projetado para reter todos os atributos, especialmente quando você ainda não sabe qual será o escopo dos dados ou qual será seu uso.

Note que Data Lake está diretamente relacionado a solução de Big Data. O modelo tradicional é ter um único repositório lógico (embora a estrutura física esteja distribuída) para que os dados brutos estejam disponíveis a qualquer pessoa que precise fazer análise sobre eles.

Comumente, inclusive por ser uma solução do tipo software livre, utiliza-se o Hadoop para trabalhar com Data Lakes. O Hadoop mostra-se versátil, robusto e performático e a base de seu desempenho e uso é em função de sua implementação a partir de conceitos de MapReduce originário da empresa Google (DEAN; SANJAY, 2008). Entretanto os conceitos relacionados a Data Lake são bem mais amplos do que apenas Hadoop.

Os Data Lakes usam uma estrutura de organização simples e sem esquema. Os dados são deixados em sua forma natural, disponibilizando uma coleção de dados em diferentes formatos. Uma prática recomendada comum é adicionar identificadores exclusivos e meta tags para que facilite a localização dos dados. A natureza aberta de um Data Lake permite que os analistas explorem ativamente os dados sem quaisquer requisitos ou restrições predefinidas, dando a eles a capacidade de descobrir respostas de várias formas.

Essa estrutura “anarquizada” de Data Lake, “*além de dar medo*”, é seu principal desafio. Para um Data Lake se tornar realmente utilizável, ele precisa ser “*domado*” por

mecanismos de governança que provejam procedimentos para ações de curadoria dos dados, tratamento de dados das origens, catalogar, obter consistência semântica e controles de acesso. Sem esses elementos, os dados podem não ser encontrados ou não serem confiáveis, resultando em um "pântano de dados".

O uso da conjunção Data Lake e dados abertos é uma realidade e apresenta resultados úteis para a sociedade, como é o caso do Cornell Data Lake: a Earth & Atmospheric Sciences da Cornell University criou um Data Lake públicos de dados climáticos, que originalmente foi concebido para gerar dados para suporte à decisão do setor agrícola. Os dados podem ser consultados usando ferramentas padrão, como o Amazon Athena ou o Apache Spark. (CORNELL EAS DATA LAKE, 2018).

3.10.4 Sistemas de recuperação, recomendação e relevância de informação

Sistema de Recuperação de informação (SRI) é uma área da ciência da computação, entretanto originária da ciência da informação, que em linhas gerais lida com o armazenamento de documentos e a recuperação de forma automática de informações sobre esses documentos.

Monteiro (MONTEIRO, FERNANDES, DECARLI & TREVISAN, 2017 p.161-175) apresenta funcionalidades que um SRI clássico deve desempenhar:

- a) dar informações contidas nos documentos, usualmente através dos processos de indexação e descrição dos documentos;
- b) armazenamento e gestão física e/ou lógica desses documentos e de suas representações;
- c) recuperação das informações representadas e dos próprios documentos armazenados, representação de forma a satisfazer as necessidades de informação dos usuários.

O termo genérico “documento” no contexto de recuperação pode ser detalhado em busca por informações em documentos, busca pelos documentos propriamente ditos, busca por metadados que descrevam documentos e busca em banco de dados, sejam eles relacionais e isolados ou banco de dados interligados em rede na internet, não havendo restrição ao tipo de documento, seja ele estruturados, semiestruturados ou não estruturados.

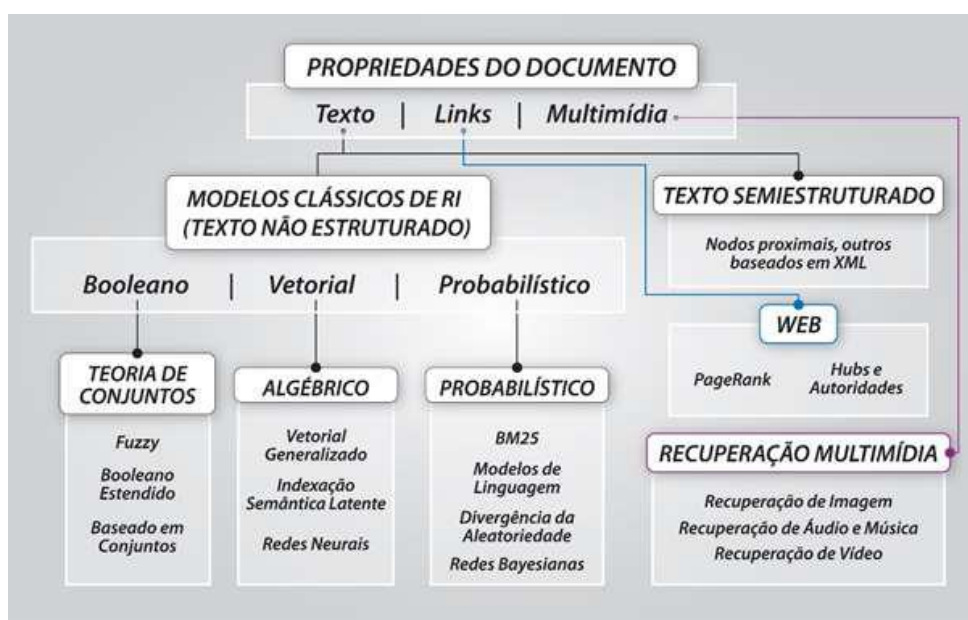
Na ciência da informação há outros conceitos sobre SRI que giram em torno da mesma essência, por exemplo, para Araújo (2012) um SRI deve representar, armazenar, organizar e localizar os itens de informação, e a referida autora aponta que a indexação é a

principal função de um SRI, e seus componentes devem incluir: documentos; necessidades do usuário; consulta formulada e o processo de recuperação propriamente dito.

Já para Ferneda (2012, p.13), têm por função “representar o conteúdo dos documentos do corpus e apresenta-los ao usuário de uma maneira que lhe permita uma rápida seleção dos itens que satisfazem total ou parcialmente a sua necessidade de informação [...].” Os modelos de recuperação de informação vêm sendo desenvolvidos há muito antes até da invenção do computador e dos recursos tecnológicos atuais, porém, muitas ideias relacionadas à recuperação de informação são até hoje utilizadas.

Monteiro (Monteiro, 2017) lista os modelos mais conhecidos e sua recomendação de aplicabilidade como sendo: para textos não estruturado, o modelo Booleano (fuzzy, booleano estendido e baseado em conjuntos), o modelo algébrico (vetorial generalizado, indexação semântica latente e redes neurais) e modelo probabilístico (BM25, modelos de linguagem, divergência da aleatoriedade e redes bayesianas); para textos semiestruturados em que os modelos tratam a estrutura fornecida pelo texto, nodos proximais e outros baseados em XML. Para links, o modelo PageRank, Hubs e autoridades. Para Recuperação Multimídia (em muitos casos é quase que inexistente, modelos de recuperação de imagem, de áudio e música e recuperação de vídeo.

Figura 6 - taxonomia de modelos de recuperação de informação



Fonte: (MONTEIRO, FERNANDES, DECARLI & TREVISAN, 2017)

Em função de escopo do trabalho, esses modelos não serão detalhados neste documento.

Cabe observar que os atuais mecanismos de busca na web são considerados SRI, conforme menciona Monteiro (MONTEIRO, FERNANDES, DECARLI & TREVISAN, 2017). Sua anatomia clássica, em que um mecanismo de busca é um subsistema de um sistema maior, é composta por aplicativos para capturar palavras-chave e frases que identifiquem o conteúdo de uma página web para indexá-la e armazená-la em sua base de dados. No lado do usuário há a inserção de termos de pesquisa que irão acionar um algoritmo de busca para exibir os resultados recuperados da base de dados.

Outra técnica dentro do contexto de um SRI é o sistema de recomendação. Para Cazella, Nunes e Reategui (2015) o objetivo dos sistemas de recomendação é auxiliar no aumento da capacidade e eficácia do processo de indicação. Eles são partes de um Sistema de Recuperação de Informação. A operação tradicional de um sistema de recomendação consiste em direcionar para indivíduos potencialmente interessados nas recomendações, a partir de recomendações fornecidas por outros indivíduos como entrada para o sistema.

Sistemas de recomendação são técnicas computacionais, baseada em análise de dados com técnicas de aprendizagem de máquina e recuperação de informação, para gerar recomendações baseadas em algum tipo de filtragem, as mais comuns são: colaborativa (considera a experiência de todos os usuários), baseada em conteúdo (considera a experiência do usuário alvo) e híbrida (as duas abordagens são consideradas).

Filtragem colaborativa é a técnica de recomendação de itens a um usuário, tendo como base avaliações passadas de todos os usuários. As avaliações são registradas por como avaliações em escala de satisfação com o produto, avaliação binária (gostei, não gostei) ou por meio de descrição de comentários. Filtragem Colaborativa são comumente utilizadas em e-commerce e em Redes Sociais (exemplo, a filtragem colaborativa pode ser encontrada no facebook e no spotify).

A filtragem baseada conteúdo tem como premissa que os usuários tendem a se interessar por itens semelhantes aos escolhidos no passado. Então os sistemas de recomendação selecionam itens com base em manifestações de interesse realizadas no passado pelo usuário. Exemplos anúncios em redes sociais são apresentados aos usuários com base em seu perfil de interesse, histórico de navegação e postagens.

As técnicas para detectar e aprender sobre o perfil do usuário são normalmente baseadas em realimentação de relevância, algoritmos genéticos, redes neurais, e classificadores bayesianos. O modelo de espaço vetorial e indexação semântica podem ser usados para representar documentos.

A filtragem por similaridade consiste em descobrir itens similares aos que o usuário já adquiriu ou que pessoas próximas a ele ou com perfil parecido tenham adquirido. Recomendações por similaridade são calculadas por meio de técnicas como busca booleana, métodos probabilísticos e modelos vetoriais. Alguns exemplos de algoritmos de Classificação mais utilizados são: Vizinho K-mais-próximo, Árvores de Decisão, Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Word2Vec, Classificador de Bayes, SVM e Redes Neurais.

As recomendações podem ter como base sinalizações de outras pessoas com interesses semelhantes ou por meio de opiniões de usuários conhecedores do assunto, por serem especialistas ou por terem experimentado o objeto da pesquisa. Os sistemas de recomendação, incorporados aos SRI, visam facilitar o processo de decisão sobre resultados de buscas de informações.

Um velho conceito que passou a ser amplamente utilizado em conjunto com os SRI: relevância.

Embora o uso do termo “relevância” tenha sido incorporado ao campo da recuperação de informação desde a década de 50, sua importância emergiu com a mudança de paradigma que envolve a mídia em que se encontra a informação (lembrando que em meados de 1950 a informação digital não estava consagrada). Em especial com o advento da internet e com o crescente volume de informações geradas a partir da propagação de serviços pela rede.

Monteiro (MONTEIRO, FERNANDES, DECARLI & TREVISAN, 2017) apresenta um conceito sobre relevância como sendo: “[...] a derivação de uma medida indicando a probabilidade de um documento satisfazer a um determinado pedido [...]”.

Em seu artigo demonstra que o uso do termo foi incorporado ao campo da recuperação da informação, conforme mencionado, na década de 1950, por Mooers, mas a primeira exploração extensa do conceito de relevância apareceu em 1958, durante a ‘International Conference for Scientific Information’ (ICSI) em que apresentou os seguintes aspectos:

- a) Relevância não é, exclusivamente, uma propriedade do documento;
- b) Relevância não é uma propriedade dicotômica (isto é, entre documentos satisfatórios e insatisfatórios ao pedido do usuário);
- c) Existe uma “relevância para o usuário” que deve ser julgada (configurando-se, portanto, em julgamento da relevância).

Relevância resume-se em mostrar os resultados possivelmente mais importantes em forma de ranque, do mais relevante ao menos relevante, com base em recomendações.

Os modelos utilizados nos sistemas de recuperação de informação influenciam o resultado dos processos de relevância, por exemplo: em um modelo booleano, os resultados recuperados são equivalentes para o sistema, pois ao utilizar a teoria dos conjuntos restringe-se apenas aos resultados que pertencem à lógica de busca, sem a possibilidade de ordenar/classificá-los do mais para o menos relevante.

No modelo vetorial, palavras e expressões possuem valores de representação, assim os resultados podem ser ordenados do mais relevante (maior valor) para o menos relevante (menor valor). O probabilístico baseia-se na teoria matemática das probabilidades para apresentar a primeira leva de resultados que o sistema considera relevante, em que relevância é determinada a partir da interação do usuário com o sistema (MONTEIRO, FERNANDES, DECARLI & TREVISAN, 2017)

3.10.5 Conclusão

Nessa seção foram apresentadas, a título de exemplo, algumas tecnologias contemporâneas capazes de transformar dados abertos em soluções uteis, tanto para a sociedade - por meio de desenvolvimento de aplicativos ou para melhoria em serviços públicos, como também para suporte à formulação de políticas públicas.

Essas plataformas têm os vieses tanto de apoiar na busca, processamento e análise dos dados - como é o caso do Data Lake, Big Data e sistemas de recuperação, relevância e recomendação de informação, e ainda de forma integrada entre as plataformas -, como também para se beneficiarem de dados abertos de forma a prestar serviços para a sociedade, como é o caso de cidades inteligentes e Iot; e ainda nesse caso, além de poderem usufruir de dados abertos, podem ser também fornecedoras de dados abertos governamentais, uma vez que é o principal ator em implementações de cidades inteligentes no fornecimento de serviços públicos inteligentes.

3.11 **Experiências de sucesso no uso de dados abertos com tecnologias disruptivas**

Na seção anterior foram apresentadas algumas tecnologias que podem se beneficiar de dados abertos ou agregar valor (no sentido de fornecer inteligência analítica) a dados abertos governamentais.

Um dos objetivos do Decreto 8.777 (BRASIL, 2016) está relacionado à promoção do desenvolvimento tecnológico e a inovação nos setores público e privado e fomento de novos negócios; e algumas soluções desenvolvidas pelo setor de desenvolvimento de software, por organizações não governamentais, por acadêmicos ou por outros setores, utilizando dados abertos governamentais, podem ser encontradas no portal de dados abertos do Governo Federal: www.dados.gov.br/aplicativos.

Nesta seção serão apresentados alguns casos que se destacaram no provimento de soluções inteligentes utilizando dados abertos em conjunto com algumas das soluções disruptivas apresentadas na seção anterior.

3.11.1 O caso Big Data do Governo Federal: GovData

A Plataforma de Análise de Dados do Governo Federal (GovData) tem o objetivo de simplificar o acesso, compartilhamento e avaliação de gestores públicos a diferentes dados governamentais. Essa plataforma é uma das ações resultantes do Decreto nº 8.789 (BRASIL, 2016b), que dispõe sobre o compartilhamento de bases de dados na administração pública federal.

O GovData é uma alternativa para prover funcionalidades de Big Data e Analytics, em conformidade com o normativo que veda a contratação de soluções de Big Data pelos órgãos do Governo:

“Fica vedada a contratação de soluções de armazenamento massivo de dados, tipo Big Data, e de análise de dados, estruturados ou não, tipo Analytics, por órgãos e entidades integrantes do SISP. 1.1. Solicitações de exceção ao disposto no caput deverão ser submetidas pelo órgão ou entidade interessado, com as devidas justificativas, à apreciação da STI”. (BRASIL, PLANEJAMENTO, 2018).

Conforme divulgado pelo gestor da plataforma, o Ministério do Planejamento (BRASIL, PLANEJAMENTO, 2018b), nesse ambiente digital, é possível ter acesso a diferentes bases de dados governamentais e a ferramentas de cruzamento de informações. Nessa fase, as bases de maior interesse do Governo Federal para participarem da iniciativa, denominadas Bases SISP são:

- o Cadastro de Pessoas Físicas (CPF);
 - o Sistema Integrado de Administração de Recursos Humanos (Siape);
 - o Sistema Integrado de Administração Financeira do Governo Federal (Siafi);
- e
- o Registro Nacional de Veículos Automotores (Renavam).

Principais características da plataforma GovData:

- 1- Grande volume de dados (Histórico);
- 2- Atualizações das Bases SISP com periodicidades pré-definidas (mensal, semestral e anual);
- 3- Disponibilização de ferramentas de análise e cruzamento de dados;
- 4- Ambiente seguro com Autenticação e Controle de Acesso;
- 5- Acesso somente para Gestores Públicos;
- 6- Acesso às principais bases do Governo Federal (Bases SISP) mediante autorização de acesso dos órgãos responsáveis;
- 7- Racionalização dos gastos públicos

3.11.2 O caso “Operação Serenata de Amor”

A utilização de dados abertos do governo brasileiro não é uma tarefa trivial. Há até quem utilize dados estrangeiros, para realização de prova de conceito, por não conseguir utilizar dados disponíveis pelo Governo em função da qualidade dos modelos.

Esse foi o empecilho vivenciado por um Analista do Serpro, ao realizar uma prova de conceito com o objetivo de demonstrar aplicação de técnicas em dados governamentais agregando valor ao governo brasileiro e à sociedade. A falta de metadados adequados, e o uso de formatos proprietários dificultam o uso desses dados. (GOIS, 2017).

O projeto “Operação Serenata de Amor” venceu esse desafio por meio de integração de equipes com um mesmo propósito: desenvolver um projeto aberto que usa ciência de dados e inteligência artificial para auditar contas públicas e auxiliar no controle social.

Trata-se de um projeto aberto com a finalidade de fiscalizar gastos públicos e compartilhar as informações de forma acessível a qualquer pessoa.

A ideia surgiu do cientista de dados Irio Musskopf, como forma de participar ativamente do processo democrático, fiscalizando os gastos públicos. Ao compartilhar a ideia com amigos, o projeto ganhou força e hoje conta com um time formado por 10 pessoas trabalhando diretamente e mais de 600 voluntários dispostos a realizá-lo.

Inicialmente o foco era confrontar os dados de gastos realizados pelos parlamentares com os limites legais estabelecidos pela CEAP – Cota para Exercício da Atividade Parlamentar; era usar a tecnologia para melhorar a política.

O trabalho em equipe fez com que um dos parceiros focasse em outro projeto com a finalidade de transformar dados abertos com baixa qualidade em dados utilizáveis, é o projeto “Brasil.IO”. Um exemplo dessa ação foi a disponibilização dos dados dos sócios das

empresas brasileiras para acesso livre a partir das bases de dados disponibilizadas pela Receita Federal.

Na verdade, a Receita Federal tinha feito isso, mas utilizando um formato que quase ninguém compreendia, e que, por isso, era grande a dificuldade para criar aplicações e casos de uso com esses dados. (Brasil.IO)

Outro parceiro trabalhou em outro projeto, o “Perfil Político”: um projeto que visa cruzar praticamente vários bancos de dados públicos e oferecer informação sobre o comportamento, intenções e histórico dos políticos; é também uma plataforma para comparar e conhecer o histórico dos milhares de candidatos às eleições. A ferramenta ajuda as pessoas a selecionarem perfis de candidatos. O Perfil Político trabalha com dados públicos e oficiais de diversas fontes, como o TSE e a Câmara dos Deputados.

Atualmente a arquitetura tecnológica do projeto é baseada em:

Utiliza: Jupyter, Python

Toobox: ferramentas genéricas com a função de buscar todos os dados utilizados pela Serenata de Amor a partir de suas fontes. Mantém atualizados os dados.

Rosie: robô com inteligência artificial, é uma implementação do Scikit-learn com classificadores para analisar e identificar suspeitas em gastos de deputados federais em exercício de sua função, a partir dos dados recuperados pelo toolbox, relacionados às verbas de CEAP – Cota para Exercício da Atividade Parlamentar

Processamento de linguagem natural, ou NLP;

Jarbas: robô para prover os dados, apresenta resultados em dashboard com os achados suspeitos.

Pandas: biblioteca para ciência de dados que fornece estruturas de dados de alto desempenho e fáceis de usar e ferramentas de análise de dados para a linguagem de programação Python (<https://pandas.pydata.org/>).

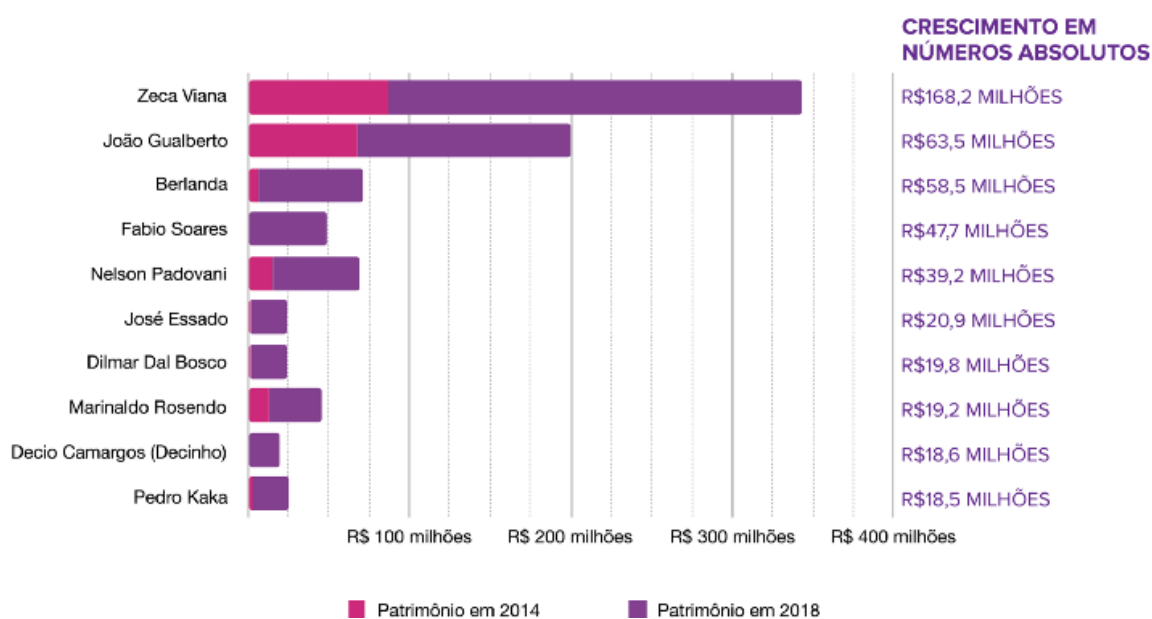
SciKit-learn - Ferramentas simples e eficiente para mineração e análise de dados (<https://scikit-learn.org/stable>);

O Projeto Serenata de Amor divulgou números em seu sítio (SERENATA, s.d.) obtidos com dados abertos:

- Uso da transparência: o Brasil é o 8º mais transparente, mas é o 69º mais corrupto de 180;
- A partir dos achados da Rosie, 8.276 reembolsos suspeitos encontrados;

- 735 Deputados diferentes encontrados com gastos suspeitos;
- R\$ 3,6 milhões encontrados em reembolsos suspeitos;
- R\$ 6.205,00 é o valor de uma nota fiscal, referente a uma refeição, que foi reembolsada;
- Um Deputado costuma gastar R\$ 6.000,00 mensais em gasolina; em média equivale a 30 tanques de gasolina em um mês;
- Dois Deputados já pediram reembolso de 13 refeições consumidas por cada um, em um único dia;
- Um Deputado já foi reembolsado com gastos em bebidas alcoólicas em Las Vegas;
- Média dos 100 candidatos com maior enriquecimento é de R\$ 10 milhões em evolução patrimonial nos últimos 4 anos:

Figura 7 - evolução do patrimônio de candidatos



Fonte: (OPERAÇÃO SERENATA DE AMOR, 2018)

3.11.3 O caso “Portal Tesouro Transparente”

O Portal Tesouro Transparente (BRASIL, STN, 2015) foi lançado pela Secretaria do Tesouro Nacional com o objetivo de criar uma plataforma de transparência que torne disponíveis à sociedade brasileira dados e informações que o Tesouro produz ou consolida.

A plataforma disponibiliza dados abertos publicados pelo Tesouro Nacional e vai além: disponibiliza análise com base nesses dados, apresentados em forma de gráficos e painéis, além de vasta explicação sobre os dados – histórias dos dados. O cidadão conta ainda com acesso às principais publicações e boletins gerados com dados disponibilizados no mesmo ambiente, e também com vídeos, animações didáticas, séries explicativas, tutoriais e eventos relacionados ao Tesouro Nacional.

O Portal passou a ser um dos principais canais de comunicação com a sociedade.

Figura 8 - imagens na página principal do portal Tesouro Transparente



Fonte: (BRASIL, STN, 2015)

O Portal Tesouro Transparente está dividido em seis temas: “Estatísticas”, “Fiscais e Transparência”, “Planejamento Fiscal”, “Dívida Pública Federal” e Ativos da União”, “Gastos da União”, “Estados e Municípios”, “Contabilidade e Custos”, e a área de dados abertos.

Cada tema apresenta dados de diferentes maneiras para serem explorados pelo usuário. Ao navegar pelo site, é possível realizar recortes de dados, fazer seleções múltiplas para análises comparativas entre gráficos, realizar filtros com variações de dados, exportar informações para programa de planilha eletrônica, entre outras possibilidades para o enriquecimento da análise.

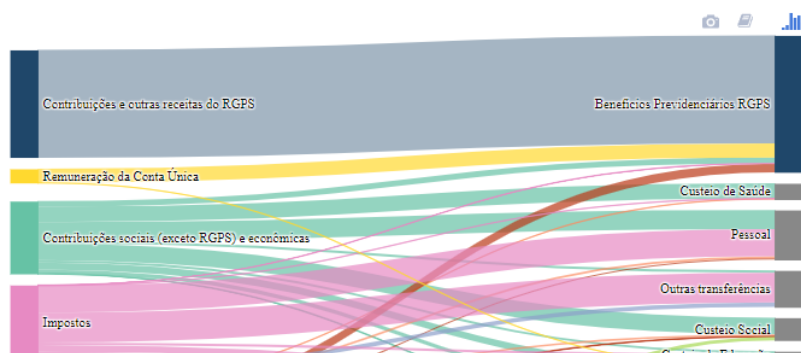
Os gráficos e relatórios apresentados contam também com histórias a partir dos dados e contemplam profunda descrição sobre os painéis apresentados.

Figura 9 - diagrama: “como são aplicadas as receitas federais?”

O diagrama: como são aplicadas as receitas federais?

Os diagramas estão exibidos logo abaixo. Observe que é possível movimentar os “nós” (as extremidades dos fluxos) clicando e arrastando com o mouse. Além disso, passar o mouse sobre os elementos do diagrama faz com que sejam exibidos os valores estimados.

A versão completa



Fonte: (BRASIL, STN, 2015a)

A área de dados abertos disponibiliza os diversos conjuntos de dados, em atendimento ao Decreto 8.777, inclusive os utilizados para apresentação dos painéis de informações. Porém, ainda traz uma inovação ao permitir acesso aos dados por meio de grid:

Figura 10 - Grid de visualização dos dados de despesas com o sistema integrado de Administração.

Despesas com o Sistema Integrado de Administração ...

URL: <https://www.tesourotransparente.gov.br/ckan/dataset/743cbd86-97dc-4e7b-89cc-d1769c396948/resource/4e>

Do resumo do conjunto de dados

Este conjunto de dados apresenta as despesas utilizadas para custeio do ambiente de pr Financeiro.

Fonte: Despesas com o Sistema Integrado de Administração Financeira - SIAFI - Ambiente de Produção

Data Explorer

Exercici...	Mes	Valor po...	Despes...	Despes...	Franqui...
2015	JANEIRO	0,089	7.930.33...	7.930.33...	85.765.583
2015	FEVERE...	0,089	7628212...	1555854...	85.765.583
2015	MARCO	0,089	7.633.13...	23.191.6...	85.765.583
2015	ABRIL	0,089	7.568.72...	30.760.4...	85.765.583
2015	MAIO	0,09073...	7.026.04...	37.786.4...	83.478.5...
2015	JUNHO	0,0955	8.745.75...	46.532.2...	77.189.0...

Fonte: (BRASIL, STN, 2015b)

Tecnologia:

O portal foi desenvolvido pelas equipes da Superintendência de Suporte e Dados do Serpro em conjunto com o Núcleo de Sistemas de Informações de Negócio da STN (Nusin).

Para a construção da página de apresentação do portal e as chamadas das notícias, o Serpro utilizou a ferramenta Joomla, um dos principais sistemas de gestão de conteúdo para desenvolver sites com design customizado. O CKAN - Comprehensive Knowledge Archive Network, que é uma aplicação web de catalogação de dados, foi usado para fazer a página de dados e organizar as informações por temas no site.

A equipe do Tesouro utilizou o software de Business Intelligence (BI) Qlikview (<https://www.qlik.com/pt-br>) para a construção de painéis e gráficos da página. A ferramenta agrupa informações e exibe os dados de uma forma visual mais atraente. Além disso, o Tesouro usou o D3JS (<https://d3js.org/>), solução que produz visualizações dinâmicas e interativas de dados, para fazer a biblioteca de dados da página de painéis de séries temporais do portal.

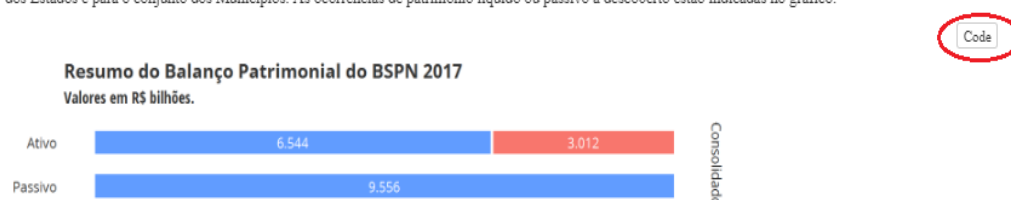
O portal disponibiliza ainda diversos códigos nas páginas responsáveis pela geração de seus gráficos e painéis e que podem ser copiados e reaproveitados pelos interessados:

Figura 11 - Imagem do Balanço Patrimonial do BSPN 2017, exemplificativa de disponibilização de acesso ao código de geração do gráfico

O balanço patrimonial

O balanço patrimonial evidencia a situação patrimonial da entidade pública, elencando seus **ativos** (bens e direitos) e seus **passivos** (obrigações). Quando o ativo é maior do que o passivo, temos um *patrimônio líquido*. No entanto, quando o passivo supera o ativo, temos uma situação de *passivo a descoberto*.

O gráfico a seguir resume o Balanço Patrimonial, demonstrando os valores dos ativos e passivos para a Federação ("consolidado"), para a União, para o conjunto dos Estados e para o conjunto dos Municípios. As ocorrências de patrimônio líquido ou passivo a descoberto estão indicadas no gráfico.



Fonte: (BRASIL, STN, 2015c)

4 MECANISMOS PARA VIABILIZAR TRANSPARÊNCIA ATIVA E ABERTURA DE DADOS NA ANAC

Embora sejam perceptíveis os avanços do Governo rumo às iniciativas para atendimento aos princípios da publicidade e da transparência com dados e informações digitais, constatável ao acessar qualquer sítio do Governo Federal, onde normalmente podemos encontrar uma seção intitulada de “Acesso a informação” e outra de “Dados abertos”, ainda tem sido pouco efetivo o alcance dos resultados propostos pelos compromissos que constituem o significado de governo aberto (ver seção “Governo aberto, dados abertos governamentais e os compromissos com a OGP).

Vale relembrar que a seção “Reflexão sobre dados abertos” referenciou o termo “Prontidão para a abertura de dados”, por Machado Junior (2015, p.30). Esse termo é conveniente para este trabalho, uma vez que seu objetivo é apresentar mecanismos, dentro de um contexto de governança de informações digitais, que promova ações de transparência e abertura de dados, de forma ativa e que auxilie na diminuição da assimetria de informação entre a Agência, a firma regulada e o cidadão.

Ocorre que as ações de abertura de dados no Governo Federal, e não poderia ser diferente na ANAC, são tratadas de forma passiva, ou seja, por provocação. Somente são consideradas a partir da existência dos dados. Não há mecanismos que tratem essas questões de forma ativa, em momento de concepção de solução, ou durante o levantamento de necessidades, e o pior: não há uma avaliação prévia sobre a eficiência da disponibilização desses dados e muito menos uma avaliação posterior sobre o resultado para a sociedade.

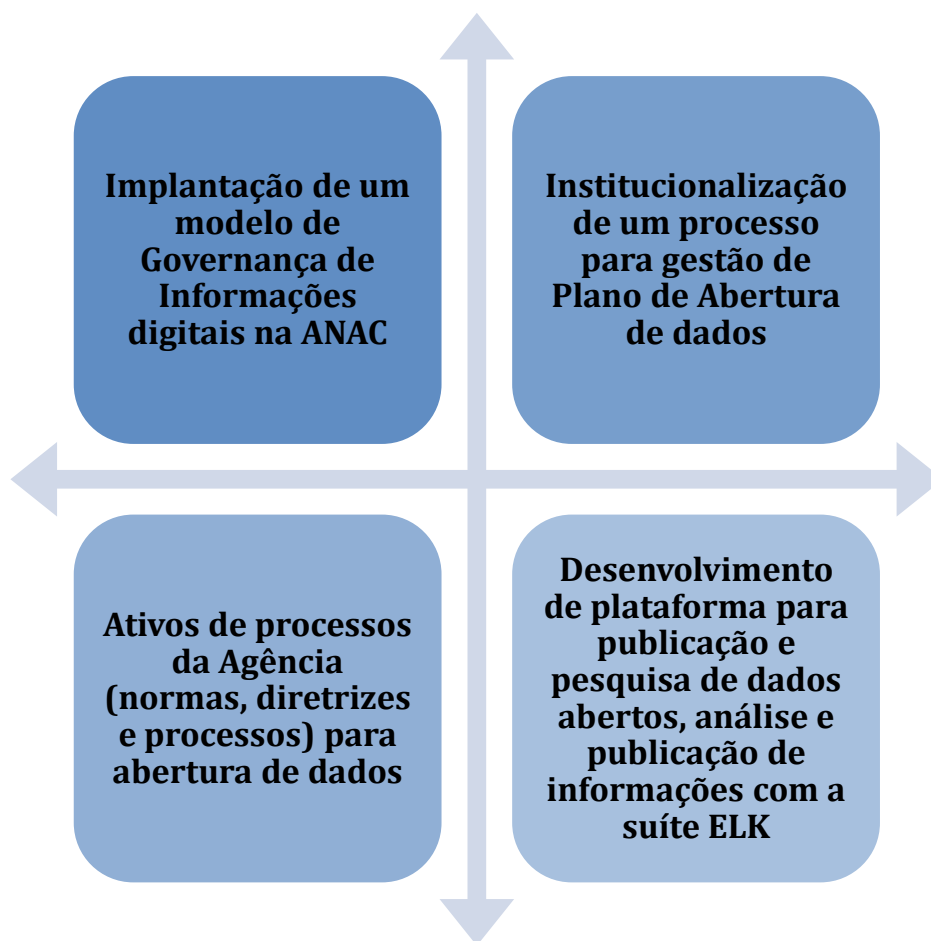
O objetivo desses mecanismos é viabilizar o atendimento aos princípios da transparência e publicidade; e para auxiliar na reflexão sobre o tema proposto. A seguir são listadas algumas frases registradas neste trabalho que nos levam a uma reflexão sobre a importância do tema. Embora haja diversas outras frases, essas foram selecionadas em função de seu apelo intrínseco à necessidade de tratar as questões de transparência e abertura de dados como elementos vitais na garantia de direitos fundamentais da sociedade:

- A Declaração Universal dos Direitos Humanos assegura ao ser humano, além de outros, o direito fundamental de acesso à informação, e ao Estado o dever da transparência;
- O acesso aberto à ciência contribui para a inovação e o crescimento econômico.
- Os avanços científicos, tanto substantivos quanto metodológicos, são agora intensivos em dados e requerem acesso aberto a dados científicos;
- O custo da pesquisa é reduzido [...];

- Eles (dados públicos) são muito semelhantes aos bens públicos, porque o uso deles por uma pessoa não impede seu uso pelos outros. No entanto, estes são bens públicos de um novo tipo: não apenas o seu uso não diminui o estoque comum, mas o enriquece;
- princípios do Open Government Data: completo, primários, tempestivo, acessibilidade, processável por máquinas, não-discriminatório, não-proprietário, licença livre;
- outros princípios recomendados pela Open Government Data: gratuito e online, permanente, confiável, presunção de abertura, documentado, seguro para abrir, projetado com participação da sociedade;
- os quatro compromissos assumidos pelo Governo brasileiro que o torna Governo Aberto;
- Ao disponibilizar os dados o Governo estimula a criação de novos negócios e o desenvolvimento de soluções inovadoras;
- O seu efeito positivo garante a saúde da democracia e mantém a confiança nas instituições, uma vez que possibilita a qualquer cidadão ser um fiscal dos atos Administração;
- abertura de dados é mais do que uma questão ou uma perspectiva técnica e deve ser tratada como uma plataforma completa, um ecossistema que inclui componentes políticos, legais, avanços tecnológicos e mecanismos de classificação dos dados segundo diversos critérios;
- Para abrir os seus dados, uma instituição deve estar preparada para disponibilizá-los a um público amplo e diverso, de maneira a atender às demandas não só dos cidadãos, mas também dos diversos intermediários de dados;
- O processo de preparação para a abertura dos dados envolve mudanças institucionais, em termos de cultura, metodologia, capacitação (interna à instituição e dos usuários), envolvimento dos cidadãos, infraestrutura tecnológica e financiamento da inovação.

Os mecanismos que serão abordados neste trabalho e que foram considerados como fatores chave de sucesso para a criação de um ecossistema em que as partes interessadas interajam entre si, mas ao mesmo tempo haja clara fronteira de papéis e responsabilidades, são:

Figura 12 - imagem representativa dos mecanismos para viabilizar transparência ativa e abertura de dados na ANAC

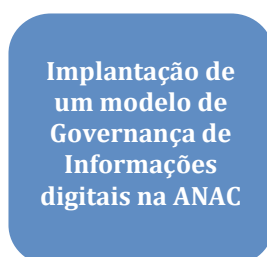


Fonte: produzida pelo autor

O detalhamento de cada um desses mecanismos será descrito nas próximas seções.

4.1 Mecanismo 1 - Implantação de um modelo de Governança de Informações digitais na ANAC

Figura 13 - imagem representativa do mecanismo 1 para viabilizar transparência ativa e abertura de dados na ANAC



Fonte: produzida pelo autor

Conforme consta em processo administrativo da Agência Nacional de Aviação Civil (BRASIL, ANAC, 2017), no decorrer de seus doze anos (ela passou a operar em 20 de março de 2006, então hoje já deve ter mais que doze anos) e ainda com um legado herdado do antigo Departamento de Aviação Civil (DAC) a ANAC, para que fosse possível realizar suas atividades regimentais, acumulou sob sua tutela um expressivo acervo de dados estruturados e não estruturados, além de diversos sistemas de gestão, os quais se desenvolviam em um cenário de documentação insuficiente dos sistemas e estruturas de dados, falta de integração dos dados, limitação dos processos para aferição da qualidade dos dados existentes, indefinição de estrutura corporativa para governança de informações, falta de estrutura para geração de processos de inteligência de negócio (BI) e baixa capacidade de prover transparência dos atos da Agência.

Além dos dados estruturados (os que são gerenciados por sistemas de bancos de dados e são mantidos com controles de TI – backup, monitoramento, suporte, etc.), havia também informações digitais que eram armazenadas em máquinas de usuários, tratava-se de arquivos de planilhas, textos, pdf e outros tipos de arquivos que faziam parte de uma outra categoria de dados: os não estruturados. Grande parte deles armazenavam informações digitais relevantes, mas não eram controlados como informações digitais importantes, e pouco se sabia sobre eles: quantos eram? onde estavam? havia backup? eram redundantes? eram confiáveis?

Esse acervo de informações digitais (estruturado e não estruturado) são ativos organizacionais de valor intangível. É por meio da análise dessas informações digitais que são definidas, por exemplo, tarifas de serviços públicos: tarifas aéreas, de transporte de passageiros, de cargas, de habilitação de pilotos, de aeronaves, de embarcações, etc. Além das tarifas dos serviços públicos, todas as licitações para concessão de serviços públicos, que envolvem bilhões de reais, são sustentadas por essas análises informacionais. A título de exemplo são listados alguns desses contratos, conforme divulgado pela ANAC (BRASIL, ANAC, 2017):

- Contrato de Concessão de Aeroportos nº 001/ANAC/2012 no valor de R\$ 5.334.640.000,00,
- Contrato de Concessão de Aeroportos nº 002/ANAC/2012 – SBGR no valor de R\$ 17.697.076.000,00,
- Contrato de Concessão de Aeroportos nº 003/ANAC/2012 no valor de R\$ 12.983.951.000,00),
- Concessão do Aeroporto Hercílio Luz no valor de R\$ 2.269.594.665,00,

- Concessão do aeroporto de Fortaleza no valor de R\$ 3.503.536.466,00,
- Concessão do aeroporto de Natal no valor de R\$ 650.000.000,00,
- Concessão do aeroporto de Porto Alegre no valor de R\$ 4.239.073.572,

A situação fática da falta de controle aliada ao expressivo tamanho do acervo de informações digitais necessário para a regulação do setor de aviação civil pela ANAC exigiu a adoção de modelo de governança que possibilitasse o controle, delegação de competência, responsabilização e gestão das informações digitais. A ausência de uma boa governança poderia acarretar em desinformação de tal ordem que sua consequência recairia, em última instância, sobre os custos dos serviços pagos pelos usuários, que são autorizados pela Agência.

4.1.1 Projeto “Governança de Informações Digitais”

Para solucionar esses problemas, foi acrescentado à carteira de projetos prioritários da ANAC, por meio da Portaria nº 2013, de 5 de agosto de 2016, o projeto “Remodelagem das Bases de Dados”, e um dos produtos previstos no projeto era a elaboração de um modelo de governança de informações digitais.

Em 20 de dezembro de 2017, por meio da Portaria nº 4.227, o projeto teve seu nome alterado para “Governança de Informações Digitais”. Essa mudança tinha como objetivo alteração de seu sentido conotativo, uma vez que buscava mudar o entendimento da essência do projeto, tirando-o de uma tendência tecnicista (em função do termo “remodelagem”) para uma visão estratégica, direcionada pelo termo “Governança”.

Cabe destacar que o produto principal do projeto era a “Instituição da política de governança de informações digitais e estruturação, através da elaboração de diretrizes, dos processos de curadoria, modelagem e qualidade de dados”.

Falar sobre a questão dos dados não estruturados

4.1.1.1 Diagnósticos sobre gestão de dados

A primeira fase do projeto prioritário previu a realização de diagnóstico sobre a forma de gestão dos dados estruturados da Agência (BRASIL, ANAC, 2017), e a forma de controle desses ativos e uma análise de maturidade dessas atividades.

O diagnóstico sobre os dados constatou que esse acervo de informações digitais estruturadas da ANAC alcançava 7,5 terabytes e com tendência de crescimento contínuo, que eram manipulados por 127 sistemas e, de acordo com a documentação apresentada sobre os sistemas e bancos de dados, com pouca maturidade em gestão e controle de dados.

Com relação à maturidade no controle e gestão desse acervo de dados e informações digitais da Agência, a equipe do projeto realizou duas avaliações de maturidade com o objetivo de compreender as práticas para alcançar um nível de maturidade adequado.

As metodologias consideradas para avaliação do nível de maturidade foram:

a. Data and Information Governance Maturity Framework, que é uma plataforma de avaliação de maturidade contemplada no modelo de governança de dados e informações digitais da Nova Zelândia, que aqui denominaremos como DIGM.

b. DMM (Data Management Maturity) Model é um modelo de melhoria de maturidade e capacidade de processo para a gestão de ativos de dados e atividades relacionadas, de uma organização. Ele contém melhores práticas para o estabelecimento, a construção, a manutenção e a otimização efetiva da gestão de dados através do seu ciclo de vida, desde a criação até a entrega, manutenção, arquivamento e descarte.

A classificação do DMM Model é baseada em 5 níveis conforme a tabela abaixo:

Tabela 1: Níveis de Maturidade de processos SEI DMM Model.

Nível	Descrição	Perspectiva
1 - Performed (Inicial)	Os processos são realizados de forma improvisada ou conforme a necessidade dentro de projetos. Os mesmos processos não são aplicados em todas as áreas de negócio. As atividades são geralmente reativas e não preventivas. Melhorias fundamentais podem existir, mas estas melhorias não são levadas para toda a organização ou mantidas.	Dados são tratados como requisitos para implementação de projetos de sistemas.
2 – Managed (Gerenciado)	Os processos são planejados e executados de acordo com políticas; emprega pessoas com as habilidades e recursos adequados para produzir artefatos/resultados controlados; envolve as partes interessadas relevantes; são monitorados, controlados e avaliados para ter aderência aos processos definidos.	Existe a consciência da importância do gerenciamento dos dados como um ativo corporativo crítico para o negócio.
3 – Defined (Definido)	Um conjunto de processos padronizados são empregados e seguidos de forma consistente. Processos para tratar necessidades específicas são adaptados a partir de um conjunto de processos	Dados são tratados no nível organizacional como essenciais para o sucesso no

	padrões de acordo com as diretrizes da organização.	cumprimento da missão.
4 – Measured (Gerenciado quantitativamente)	Métricas dos processos são definidas e usadas para gestão de dados. Isto inclui técnicas de análises estatísticas e quantitativas. O desempenho é gerenciado através de todo ciclo de vida do processo.	Dados são tratados como fonte de vantagem competitiva.
5 – Optimized (Otimizado)	O desempenho do processo é otimizado através da aplicação das análises do nível 4 para identificar oportunidades de melhorias. Melhores práticas são compartilhadas entre pares e segmento do mercado.	Dados são vistos como críticos para sobrevivência no mercado competitivo e dinâmico.

Fonte: (DMM, s.d.)

Esta metodologia foi aplicada, por meio de entrevistas nas áreas de negócios da ANAC, Superintendências e Assessorias que dão suporte à Diretoria, com as pessoas designadas como “Ponto Focal” responsáveis pelos sistemas de suas áreas, para analisar as práticas relacionadas aos grupos de processos Estratégia da Gestão de Dados e Gestão da Governança de Dados.

Cabe salientar que o DMM Model prevê mais outros 4 grupos de processos: Qualidade de Dados, Operações de Dados, Plataforma e Arquitetura e Processos de Suporte. Estes grupos de processos não fazem parte do escopo do diagnóstico de maturidade, pois eles são totalmente dependentes das práticas de Gestão de Dados e Governança, e pelos conceitos do DMM Model se você não tem pelo menos o nível 1 (Inicial) nestes processos de planejamento e controle as outras práticas funcionais não devem atender nem o nível Inicial.

O objetivo das entrevistas foi coletar de cada área de negócio a percepção de quais são os ativos de dados mais importantes da ANAC, como eles são tratados atualmente, quais os riscos e oportunidades, qual o nível de consciência sobre gestão de dados, governança de dados e curadoria de dados, obedecendo às práticas funcionais abaixo:

Tabela 2: Práticas funcionais, do nível inicial, dos grupos de processos DMM Model.

Categorias	Grupos de Processos DMM Model	Práticas Funcionais Nível Inicial (1 – Performed)
Estratégia de gestão de dados	Estratégia	EGD1.1 Objetivos, prioridades e escopo da Gestão de Dados refletem os objetivos de negócio dentro dos projetos.

	Comunicação	CGD1.1 Comunicações são gerenciadas localmente.
	Função	FGD1.1 O gerenciamento de recursos de dados e a sua supervisão são orientados por eventos.
	Justificativa de projetos	JPGD1.1 A justificativa é desenvolvida para cada iniciativa de projeto. JPGD1.2 Os benefícios e custos da Gestão de Dados são documentados e usados para o financiamento do projeto.
	Financiamento de projetos	FPGD1.1 Projetos de Gestão de Dados são financiados baseados em análises de custos e benefícios.
Governança de dados	Gestão da Governança	GD1.1 As funções de Governança de Dados são realizadas. GD1.2 Propriedade, curadoria e responsabilidade para conjuntos de dados estão definidas por projeto.
	Glossário de Negócio	GN1.1 Termos de negócio estão definidos para um propósito específico. GN1.2 Modelos de dados lógicos são criados usando termos de negócios definidos e aprovados.
	Gestão de Metadados	GM1.1 Documentação de metadados é desenvolvida, armazenados e acessíveis.
(*)	Curadoria de Dados	CD1.1 Identificação dos Curadores de dados, por assunto de negócio e conjuntos de dados, de acordo com a estrutura organizacional.

Fonte: (DMM, s.d.) ajustado pelo autor

(*) - O DMBok define o papel do “Data Steward” ou Curador de Dados, e estes colaboradores executam atividades fundamentais para a Gestão de Dados de acordo com as regras, políticas e supervisão da Governança de Dados.

No DMM Model não existe uma avaliação da maturidade da Curadoria de dados, portanto utilizou-se uma classificação adaptada do livro de referência “Advanced MDM and Data Governance in Practice” de Mark Allen e Dalton Cervo.

Com relação ao nível de maturidade, aplicando-se a média dos 9 itens avaliados, referente aos grupos de processos, concluiu-se que a ANAC atende a 3, portanto **0,33 do nível inicial** de acordo com a metodologia DMM Model.

O DIGM é um modelo de governança de dados e informações digitais, desenvolvido pelo Governo Neozelandês, que contempla uma plataforma para avaliação de maturidade. Essa avaliação considera 5 blocos de questões relacionadas aos momentos evolutivos de controle das informações digitais e dados, são eles:

Tabela 3 – Descrição dos grupos de processos para avaliação de maturidade em governança de dados do ict.govt.nz

Documentamos e Controlamos nossos Dados e Processos de Informação,	Os processos de negócio são complexos, mal compreendidos e fragmentados, múltiplas versões da verdade: Simplificar os processos de negócio simplificará os processos de dados e informação Isto nos dá uma visão clara dos nossos fluxos de dados e informação, maior qualidade e maior confiança
Incorporamos as nossas responsabilidades de dados e informações digitais	Dados e informações digitais são vistos como um ativo de negócios crítico, e não apenas uma preocupação com as TIC Clareza sobre direitos e responsabilidades Clareza na custódia de dados e informações Uma melhor formação assegurará que a governança de dados seja utilizada em toda a agência
Compartilhamos uma linguagem de dados e informações digitais	Os termos têm significados diferentes dentro e entre agências Compreensão compartilhada com uma linguagem de dados e informações digitais comum A terminologia comum de definição de dados e informações digitais dá uma maior confiança em nossos dados e informações digitais
Asseguramos a qualidade de nossos dados e informações digitais	A qualidade é difícil de medir: • Regras padronizadas são necessárias para avaliar a integridade dos dados e informações digitais, exatidão e moeda. • Tomar decisões-chave com dados e informações digitais de alta qualidade
Usamos nossos dados e informações digitais com sabedoria	Dados e informações digitais não confiáveis aumentam os riscos nas divulgações e decisões de negócios: • Confiança nos dados e informações digitais utilizados para tomar decisões importantes • Os relatórios corretos para as pessoas certas -> relatórios consistentes e confiáveis

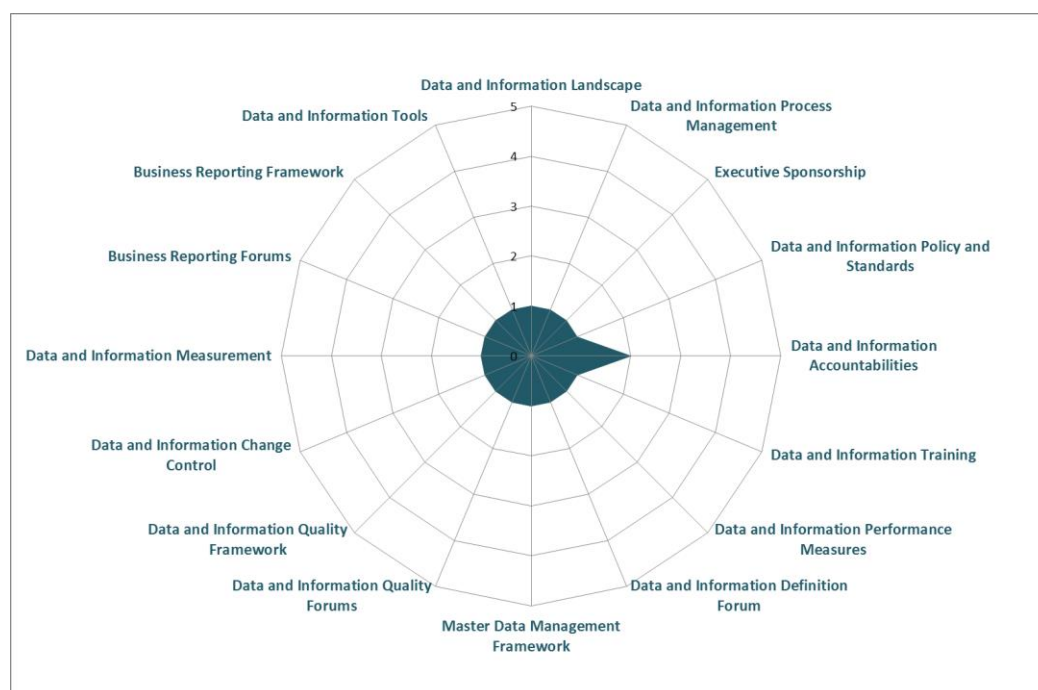
	<ul style="list-style-type: none"> • As práticas de gestão de dados e informações digitais reduzem o risco de divulgação inadequada • Explorar o valor dos dados e informações digitais para melhorar a eficácia dos serviços
--	---

Fonte: (NEW ZEALAND GOVERNMENT, 2018) ajustado pelo autor.

A metodologia prevê a avaliação de 16 processos e a cada um deles é atribuído um nível de 1 a 5, sendo que o nível 1 equivale a processos não executados ou executados precariamente, e o nível 5 é o que representa o “estado da arte” do referido processo. Além disso, a metodologia avalia cada um dos processos de forma isolada, não atribuindo uma pontuação geral, e sim por cada um dos processos.

O gráfico abaixo apresenta a situação da avaliação realizada nos processos da ANAC com relação a governança de dados e informações digitais; observa-se que dos 16 processos previstos na avaliação, 15 estão no nível 1 e apenas 1 encontra-se no nível 2:

Figura 14 - avaliação de maturidade dos processos da ANAC com relação a governança de dados e informações



Fonte: (BRASIL, ANAC, 2017)

4.1.1.2 Conceitos da ANAC sobre governança de informações digitais

A proposição de um sistema de governança no âmbito do projeto prioritário “Governança de informações digitais” alinha-se às boas práticas, conceitos e casos apresentados neste documento.

Um dos principais norteadores, impulsionadores da boa governança, é o parquet de contas, Tribunal de Contas da União, no que se refere ao tema amplo sobre governança pública.

Entretanto, previamente à apresentação do modelo governança para informações digitais da ANAC, faz-se necessário compreender alguns pilares da boa governança de informação, conceitos que sobre informação, informação digital e gestão da informação digital.

Sobre a fundamentação para elaboração de política de governança de informações digitais:

O Decreto nº 7.724, de 16 de maio de 2012 define informação como sendo: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato;

Informação, segundo a Resolução-TCU nº 254, de 10 de abril de 2013: conjunto de dados, textos, imagens, métodos, sistemas ou quaisquer formas de representação dotadas de significado em determinado contexto, independentemente do suporte em que resida ou da forma pela qual seja veiculado.

Informação, segundo Undaneta (1992): consiste nos dados processados para serem exibidos de forma inteligível às pessoas que irão utilizá-los.

De acordo com a ISO/IEC 2382-1:1993(en) - *Information technology, vocabulary*, informação é conhecimento sobre objetos, tais como fatos, eventos, coisas, processos ou idéias, incluindo conceitos, que dentro de um determinado contexto tem um significado particular (*Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning*).

Já o DAMA/DMBOK - Data Management Body of Knowledge instituição internacional que editou uma coleção de melhores práticas para gerenciamento de dados, considera informação com um elemento que possui valor para o negócio, deve ser tratado como ativo estratégico e com valor equivalente ao negócio:

Dados e informações digitais são a força vital da economia do século XXI. Na Idade da Informação, os dados são reconhecidos como um bem vital da empresa (*Data and information are the lifeblood of the 21 century economy. In the information Age, data is recognize as a vital enterprise asset*).

“Organizations that do not understand her overwhelming importance of managin data and information as tangible assets in the new economy will not survive” (Organizações que não entendem sua poderosa importância de gerenciar dados e informações como ativos tangíveis na nova economia não sobreviverão) Tom Peters, 2001.

A equipe do projeto adicionou a ação de processamento eletrônico ao objeto informação (visto haver uma delimitação relacionada a dados digitais no escopo do projeto), além de tratar como um “bem” (embora ainda intangível) com características indispensáveis no suporte a processos de negócio da ANAC e concluiu o seguinte conceito abarcando a devida importância desse patrimônio para a Agência (ativo de informação digital):

Ativos de Informação Digital: dado ou informação, manipulados por mecanismos computacionais, na forma de registro eletrônico, que sustentam ou agregam valor a um ou mais processos de negócio da Agência.

Ainda na concepção do DAMA/DMBOK, de igual forma aos demais ativos organizacionais, dados são criados ou adquiridos, armazenados ou mantidos, usados, descartados, compartilhados, e todas outras operações que sofrem qualquer tipo de ativo. Eles participam de processos e têm um ciclo de vida, assim, necessitam ser gerenciados como ativos estratégicos:

“Como qualquer recurso, os dados têm um ciclo de vida e, para gerenciar os ativos de dados, as organizações gerenciam o ciclo de vida dos dados. Os dados são criados ou adquiridos, armazenados e mantidos, usados e destruídos. *(Like any asset, data has a lifecycle, and to manage data assets, organizations manage the data lifecycle. Data is created or acquired, stored and maintained, used, and eventually destroyed).*” (DAMA/DMBOK)

Dessa forma, a equipe do projeto definiu o conceito de ciclo de vida de informações digitais digital como sendo:

Estágios pelos quais informações digitais atravessam desde sua criação, uso, transformação, disponibilização para acesso, classificação, até seu descarte, armazenamento ou disponibilização para outros processos, como por exemplo, gestão do conhecimento.

Em coerência com definições de um ativo intangível, faz-se necessária a previsão de processos de gestão para seu ciclo de vida. Devido a sua importância institucional é fundamental a existência de processos e normas de controle para garantia de sua relevância, integridade, segurança, confidencialidade e disponibilidade.

Davenport (1998, p. 173) define o gerenciamento da informação como processos, isto é, “um conjunto estruturado de atividades que incluem o modo como as empresas obtêm, distribuem e usam a informação e o conhecimento”.

Considerando o conceito de informação digital para a ANAC (como um elemento essencial aos seus processos de negócio), o tamanho de seu acervo e a transversalidade que informações digitais são utilizadas na Agência, constatamos que o seu ciclo de vida exige um tratamento que vai além de gerenciamento de informações digitais.

É necessário criar um sistema de governança que busque garantir que os ativos de informação digital estejam alinhados à estratégia organizacional inclusive acompanhando suas periódicas mudanças, que exista uma estrutura decisória e mitigadora de conflitos para informações digitais, que haja delegação de competências para gestores e curadores, definição de responsabilidade sobre uso de informação digital, segregação de atividades, monitoramento da qualidade, e que seja essencial às decisões organizacional, enfim, que tenha funções de governança e funções de gestão à luz do sistema de governança no setor público, conforme apresentado anteriormente nesta mesma seção.

Com base nessa constatação de necessidade, a equipe do projeto definiu o conceito sobre governança de informações digitais da ANAC como sendo:

Governança de informações digitais: estrutura decisória no âmbito de diretrizes, gestão e controle, na qual são definidos os papéis e as responsabilidades dos agentes envolvidos com o ciclo de vida dos ativos de informação digital da ANAC, buscando apoiar o alcance dos objetivos estratégicos da Agência;

4.1.2 Política de Governança de Informações Digitais da ANAC

Esse modelo de governança foi materializado por meio da publicação da Instrução Normativa nº 115, de 14 de agosto de 2017 que instituiu a Política de Governança de Informações Digitais - PGID da ANAC. Essa instrução normativa define o ecossistema que envolve a governança e a gestão de informações digitais na Agência, com o seguinte objetivo:

”A PGID tem como objetivo estabelecer uma estrutura de governança de informações digitais, definindo competências e responsabilidades que nortearão as atividades de gestão de informações digitais da ANAC, para o atingimento dos princípios e diretrizes nela estabelecidos, bem como para assegurar que os ativos de informação digital provejam valor e sejam consistentes, íntegros e relevantes para subsidiar decisões dos públicos de interesse”.

Os atores que compõem a política de governança de informações digitais são:

I - Comitê de Tecnologia de Informação, responsável pela aprovação de estratégias, diretrizes, processos, por estabelecer prioridades, indicadores e metas e decidir acerca de conflitos que envolvam informações digitais, além de ser o ponto de contato da Diretoria com relação à política;

II – Curadores, responsáveis gerenciar ativos de informação, em especial pelo ciclo de vida de informações digitais sob sua responsabilidade;

III - Conselho de Curadores, responsável por mediar e decidir acerca de questões técnicas de informações digitais, além de promover ações integradas entre os curadores;

IV – Área de Governança de Informações Digitais – AGID, responsável por implementar as atividades de planejamento, monitoramento, melhoria contínua e inovação relativas a gestão de informação digital;

V - Superintendência de Administração e Finanças – SAF, área responsável pelas regras de classificação de informações sob a égide da Lei de Acesso à Informação;

VI - Superintendência de Tecnologia da Informação – STI, responsável por criar e manter diretrizes operacionais relacionadas à implementação operacional da PGID.

Além desses, há ainda outros atores que agem como instâncias externas de apoio, não figuram formalmente da política, entretanto há outros instrumentos normativos que os designam e os autorizam a atuar em conjunto com a Agência, por exemplo:

a atuação da Ouvidoria da Agência está prevista no Decreto 8.777, Art. 5º, § 40º: “A autoridade designada nos termos do art. 40 da Lei nº 12.527, de 2011, será responsável por assegurar a publicação e a atualização do Plano de Dados Abertos, e exercerá as seguintes atribuições”

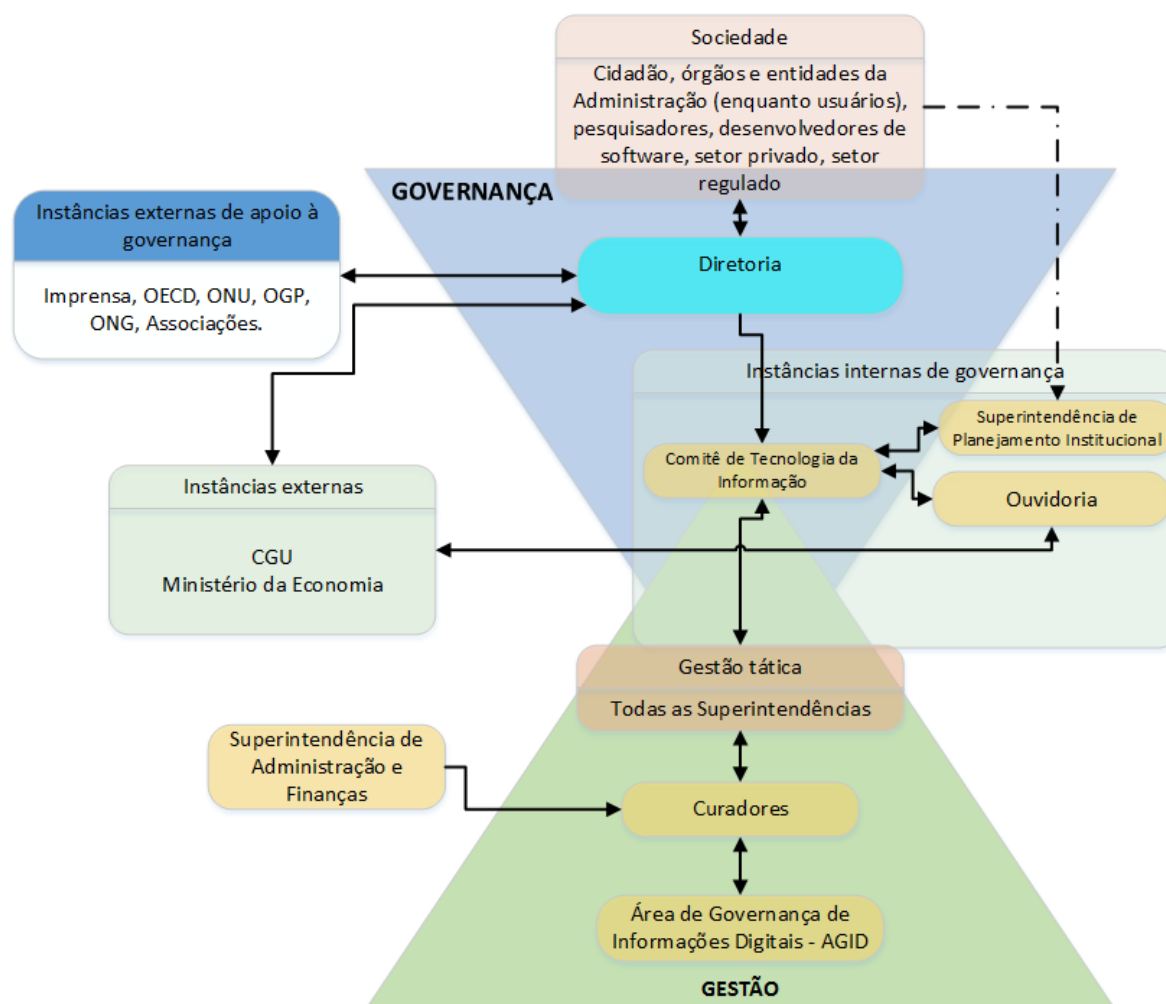
a atuação da sociedade, da imprensa e de outras organizações está prevista no mesmo Decreto, Art. 6º “Às solicitações de abertura de bases de dados da administração pública federal aplicam-se os prazos e os procedimentos previstos para o processamento de pedidos de acesso à informação, nos termos da Lei nº 12.527, de 2011”;

a atuação de outros órgãos da administração direta e indireta atuam na política com autorização do Decreto 8.789, Art. 1º: “Os órgãos e as entidades da administração pública federal direta e indireta e as demais entidades controladas direta ou indiretamente pela União que forem detentoras ou responsáveis pela gestão de bases de dados oficiais disponibilizarão aos órgãos e às entidades da administração pública federal direta, autárquica e fundacional interessados o acesso aos dados sob a sua gestão,”

A seguir é apresentado o modelo esquemático, visual, da PGID:

Figura 15 - modelo esquemático visual da Política de Governança de Informações Digitais

Figura 13 – modelo esquemático visual da Política de Governança de Informações Digitais:



Fonte: (BRASIL, TCU, 2014, p. 28) ajustado pelo autor

Embora a PGID da ANAC tenha sido publicada em uma data posterior à publicação do Decreto 8.777, que institui a política de dados abertos do Governo Federal, ela não tratou esse tema com alinhamento suficiente e explícito com o Decreto. Sua relação com o tema dados abertos ficou restrita a um inciso dentre os princípios da PGID:

“CAPITULO III

DOS PRINCÍPIOS

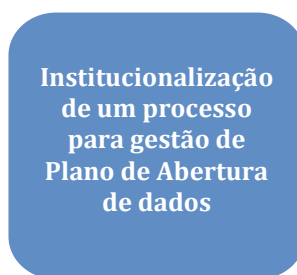
Art. 4º Constituem princípios desta PGID:

VIII - observar as normas relacionadas a informações digitais expedidas pelo Ministério do Planejamento, Orçamento e Gestão, em especial a Estratégia de Governança Digital - EGD e o Plano de Dados Abertos - PDA.” (BRASIL, 2016)

Isso leva a crer que naquele momento (da instituição da PGID) ou o tema não era prioritário ou não havia maturidade para alinhar o tema à política. Entretanto, a abertura de dados faz parte de um comando legal (Decreto 8.777), e como tal esse assunto deve ser incorporado na agenda normativa do órgão, visto que as ações de abertura de dados são imposições do Decreto e o ato de realizar as atividade carecem de delegação do respectivo órgão, sob pena de descumprimento do princípio da Legalidade (vide a menção sobre o Princípio da Legalidade na seção Governança de dados e informações).

4.2 Mecanismo 2 - Institucionalização de um processo para gestão de Plano de Abertura de dados

Figura 16 - imagem representativa do mecanismo 2 para viabilizar transparência ativa e abertura de dados na ANAC



Fonte: produzida pelo autor

Em novembro de 2018 a ANAC publicou seu Plano de Dados Abertos para o biênio 2018-2020 (BRASIL, ANAC, 2018), em atendimento ao Decreto 8.777. Os objetivos desse plano são:

1) divulgar à sociedade o cronograma e ações que nortearão as atividades de abertura de dados no âmbito da Agência Nacional de Aviação Civil – ANAC, os quais passaram a ser introduzidos no portfólio de processos e procedimentos da Agência de forma institucionalizada e sistematizada; e

2) é ampliar e aprimorar a abertura de dados no âmbito da ANAC, com vistas a aumentar a transparência e a participação social, bem como a melhoria da qualidade dos dados disponibilizados.

Nesse Plano de Dados Abertos a Agência inovou ao apresentar um processo em ciclo PDCA adotado e que será institucionalizado para elaboração e revisão do plano, a forma de participação da sociedade para auxiliar na priorização dos dados a serem abertos, os critérios utilizados para priorização dos dados, a estratégia definida para abertura dos dados com

respectivo plano de ação, o modelo de sustentação de modo a perenizar o fluxo de atualização e manutenção dos dados, a estrutura de governança e seu processo decisório, além disso, e como ponto mais importante, o plano da Agência para abertura de 40 conjuntos de dados durante o biênio 2018 – 2020.

Enfatiza-se um comando fundamentador contido no Decreto nº 8.638, de 2016 (BRASIL, 2016a), que é diretriz de perseguição obrigatória pela Administração pública, e fator de existência das iniciativas de dados abertos:

Art. 3º A Política de Governança Digital observará os seguintes princípios:

I - foco nas necessidades da sociedade;

II - abertura e transparência;

[...]

VII - participação e controle social. (BRASIL, 2016a).

4.2.1 O Processo para abertura de dados

Nesta seção, são enumerados e apresentadas as atividades e metodologia adotada para efetivar a abertura de processos em um ciclo PDCA (do inglês: plan – planejar, do – executar, check – checar, avaliar, act – agir, corrigir, melhorar).

Essa forma busca viabilizar um processo baseado em monitoramento e melhoria contínua das ações, desde o levantamento dos conjuntos de dados candidatos a serem abertos até a execução das atividades necessárias para abertura dos dados.

A divulgação de bases de dados em formato aberto se dará a partir de página específica na internet, com URL estável, em um portal de dados da ANAC. Esse portal contará com um catálogo de informações sobre as bases de dados divulgadas, de forma a permitir o entendimento necessário à utilização dos dados. Em paralelo, os dados serão publicados também no portal de Dados Abertos do Governo Federal (www.dados.gov.br).

A divulgação dos dados deverá observar as regras de proteção de informações nos termos dos art. 7º, § 3º, art. 22, art. 23 e art. 31 da Lei nº 12.527, de 2011.

4.2.1.1 Premissas

Além de estar alinhado aos objetivos mencionados anteriormente no PDA, o processo de abertura dos conjuntos de dados deve considerar as seguintes premissas:

- Publicar os dados considerados relevantes para a sociedade o mais rápido possível, no formato disponível e informando as eventuais limitações de qualidade dos dados;

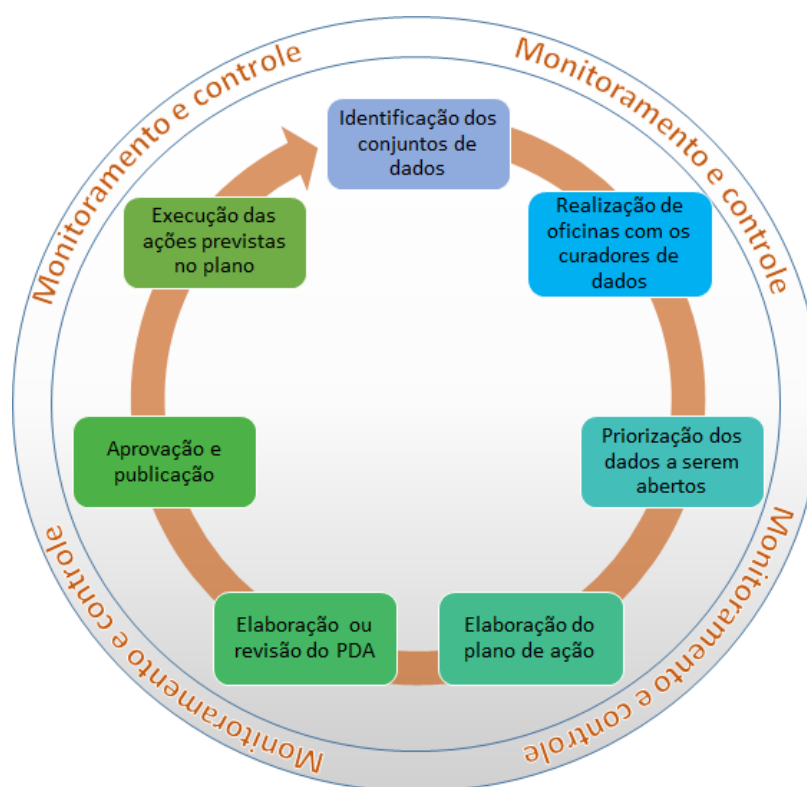
- Sempre que possível, publicar dados e seus metadados conforme estabelecido no Plano de Ação da INDA;
- Publicar os dados da Agência seguindo os padrões definidos pela e-PING, pela Infraestrutura Nacional de Dados Abertos - INDA e pela Infraestrutura Nacional de Dados Espaciais – INDE e Governo Eletrônico;
- Catalogar os dados abertos da Agência no Portal Brasileiro de Dados Abertos, ponto central de acesso aos dados do Governo Federal;
- Catalogar os dados geoespacializados, caso existam, na INDE; e
- Promover a integração entre os catálogos de metadados INDA e INDE.

4.2.1.2 Ciclo do processo de abertura de dados:

O ciclo do processo para abertura dos dados da ANAC foi elaborado de forma a sistematizar ações que possibilitem um melhor acompanhamento e controle das ações, delegação de responsabilidade aos curadores de dados, realização de atividades padronizadas para dimensionar esforços, priorização e elaboração de cronograma para que os dados sejam abertos e a criação de um ambiente favorável à melhoria contínua das atividades. O ciclo do processo é composto pelas seguintes ações:

- Identificação dos conjuntos de dados;
- Realização de oficinas com os curadores de dados;
- Priorização dos dados a serem abertos;
- Elaboração do plano de ação;
- Elaboração ou revisão do PDA;
- Aprovação e publicação;
- Execução das ações previstas no plano;
- Monitoramento e controle.

Figura 17 - imagem representativa do ciclo do processo para abertura dos dados da ANAC



Fonte: produzida pelo autor

4.2.1.2.1 Identificação dos conjuntos de dados

O objetivo dessa atividade é identificar os conjuntos de dados produzidos ou acumulados pela ANAC e que deverão ser franqueados aos cidadãos o acesso, de forma aberta, e que não recaia vedação expressa de acesso, além de identificar os gestores desses conjuntos de dados e providenciar a designação dos curadores desses conjuntos de dados com amparo do art. 11 da Instrução Normativa nº 115, de 2017: “Art. 11. Compete a cada Unidade Organizacional definir curadores de informações digitais, de acordo com regras estabelecidas pelo Comitê de Tecnologia de Informação.” (BRASIL, ANAC, 2017)

No levantamento deverão ser considerados tanto os dados estruturados como também os semiestruturados e os não estruturados. Os dados já publicados deverão também ser considerados na atividade para verificar a existência de casos em que possam ser implementadas evoluções em tecnologias visando fornecer disponibilização de dados de forma automática e, quando aplicável, conforme a periodicidade de atualização na origem, com interfaces de aplicações web amigáveis para facilitar o consumo de dados em tempo real.

4.2.1.2.2 Realização de oficinas com os curadores de dados

Conforme previsto na Política de Governança de Informações Digitais da ANAC, o curador é o servidor com competência para definir melhorias necessárias nos conjuntos de dados, tanto no que tange a procedimentos de gestão como também a questões técnicas.

A realização de oficinas com os curadores de dados possibilita a elaboração do plano de forma colaborativa, conscientizar e envolver os curadores sobre a necessidade da abertura dos dados, e esclarece seus papéis e responsabilidades, os quais derivam do Decreto nº 8.777, art. 5º, § 2º (BRASIL, 2016).

As principais atividades desenvolvidas durante as oficinas são:

- Apresentar aos curadores o tema de dados abertos;
- Explicitar o contexto em que o tema se insere e os compromissos assumidos pelo governo;
- Apresentar um levantamento preliminar de dados que a Agência se dispõe a abrir para ratificação, inclusão de novos ou exclusão pelo curador;
- Estimar esforços para melhorias de procedimentos e técnicas para abertura dos conjuntos de dados;
- Apresentar o formato do Plano de ação e do Plano de Dados Abertos;

4.2.1.2.3 Priorização dos dados a serem abertos

A abertura de dados deve, sobretudo, garantir os princípios da publicidade e da transparência da administração pública. Para se estabelecer as metas de abertura de dados e etapas de priorização, levamos em consideração o critério grau de relevância para o cidadão, alinhado com a Resolução CGINDA nº 3, de 2017, do Ministério do Planejamento

Critério grau de relevância para o cidadão:

Para que o processo de abertura dos dados da ANAC alcance seu principal objetivo, que é ser útil ao cidadão, deve-se avaliar a priorização de cada conjunto de dados a partir da avaliação pela principal parte interessada, a sociedade.

Participação social para priorização dos conjuntos de dados

Para elaboração do plano, a forma de participação do cidadão para estimar a relevância dos serviços e conjuntos de dados sob essa ótica (visão do cidadão), deve-se adotar mecanismo de participação social por meio do portal da Agência, conforme descrito abaixo:

Divulgação do projeto de dados abertos no portal da ANAC (www.anac.gov.br) em área de maior destaque com notícia explicativa e convite para que a sociedade participe da construção (priorização) do plano;

Disponibilização de formulário de enquete na página de dados abertos da ANAC (www.anac.gov.br/dadosabertos), para que a sociedade possa registrar quais conjuntos de dados considera mais importantes, os quais serão considerados como critério de priorização da ordem de abertura dos dados.

Para fins de cálculo, os dados coletados e processados durante esse período serão intitulados de PRIORIZAÇÃO_SOCIEDADE.

Outro importante critério considerado para abertura de dados diz respeito ao grau de maturidade da organização das informações e dados existentes, ou seja, o quão os dados estão em condições de serem disponibilizados de acordo com os requisitos do Decreto 8.777, de 2016, e quanto preparados estão os mecanismos técnicos que sustentam esses dados (sistemas, infraestrutura, modelo de banco de dados).

Trata-se da avaliação sobre o nível de maturidade da organização das informações e dados existentes, necessidades de ajustes em procedimentos internos e questões relacionadas a conformação e qualidade dos dados.

Na realização das oficinas com os curadores foram estimados os esforços para adequação desses elementos. Essa estimativa de esforço demonstra o quão preparado está o ambiente do conjunto de dados para que ele esteja em condições de ser publicado como dado aberto. Para fins de cálculo esse fator foi intitulado de LACUNA_DE_MATURIDADE; quanto menor o valor estimado para a LACUNA_DE_MATURIDADE, mais adequada está o conjunto de dados para ser aberto.

Essa avaliação influencia na priorização da abertura de conjuntos de dados, uma vez que aqueles dados com menor previsão de esforço para serem abertos poderão ser priorizados, pois poderão estar disponibilizados num prazo menor.

Equação para priorização dos conjuntos de dados

A priorização dos conjuntos de dados a serem abertos considera os dois fatores apresentados: PRIORIZAÇÃO_SOCIEDADE e LACUNA_DE_MATURIDADE, salientando

que quanto menor o valor estimado para a LACUNA_DE_MATURIDADE, mais adequado está o conjunto de dados para ser aberto.

Conforme apresentado no exemplo abaixo, a equação do produto entre o primeiro e o inverso do segundo resulta no índice de prioridade do conjunto de dados.

Exemplo:

Para o conjunto de dados “Dados Estatísticos do Transporte Aéreo”, com PRIORIZAÇÃO_SOCIEDADE = 10,75 e LACUNA_DE_MATURIDADE = 0,333, então:

A priorização = (PRIORIZAÇÃO_SOCIEDADE * (1 / LACUNA_DE_MATURIDADE)) * 100,

$$A \text{ priorização} = (11,08\% * (1/0,333))*100$$

A priorização para “Dados Estatísticos do Transporte Aéreo” final é = 33,23

Essa equação deve ser aplicada em cada um dos conjuntos de dados para termos a lista completa de conjuntos de dados priorizada. Salienta-se que os conjuntos de dados com as priorizações de maior valor serão os primeiros a serem executados para abertura dos dados.

4.2.1.2.4 Elaboração do plano de ação

A elaboração do plano de ação é a etapa em que o curador, em conjunto com equipe técnica de TI, estima o esforço necessário (em dias) para realização de ações para que cada conjunto de dados atenda aos critérios de “dado aberto” e seja publicado como dado aberto.

O plano de ação prevê a avaliação e estimativa das seguintes atividades:

- Ajustar procedimentos internos – para a assunção de papéis e responsabilidades pelo curador, podem ser necessários alguns ajustes em sua rotina de trabalho, por exemplo: criação de e-mail para divulgação do canal de contato para a sociedade, alocação e capacitação de equipe para responder questionamentos da sociedade, definição de etapa de validação dos dados precedendo à etapa de publicação, ou outras atividades que possam subsidiar a gestão da publicação dos dados;
- Estruturar os dados para abertura – caso os conjuntos de dados não estejam em formato adequado para ser publicado (em formato aberto), podem ser necessárias ações para conformação de conjuntos de dados, por exemplo, dados não estruturados ou semiestruturados talvez tenham que ser remodelados para um formato estruturado de forma a possibilitar sua publicação no formato de dado aberto;

- Descrever os metadados - cada conjunto de dado deve possuir um conjunto de informações associados ele, os metadados. O preenchimento dessas informações sobre os dados deve também ser estimada no plano de ação.
- Higienizar dados – algumas atividades necessárias para abertura de dados podem impactar na qualidade dos dados, um exemplo é a atividade estruturação de dados. Uma vez alterada a formatação de dados (de não estruturado ou semiestrutura para estruturado) poderão ser necessárias atividades para ajustes dos dados na nova estrutura. Para isso, deve-se estimar o esforço dessa adequação;
- Automatizar rotinas para publicação – conforme definido pelo Comitê Gestor da Infraestrutura Nacional de Dados Abertos (Resolução CGINDA nº 3, de 2017, do Ministério do Planejamento, art. 1º, § 3º): “Deverá ser priorizada a disponibilização de forma automática e, quando aplicável, conforme a periodicidade de atualização na origem, com interfaces de aplicações web amigáveis para facilitar o consumo dos mesmos em tempo real” (BRASIL, PLANEJAMENTO, 2017). Esse comando normativo remete à necessidade de estimar esforços para elaboração de rotinas automatizadas que facilitem o processo de geração, publicação e consumo dos dados abertos.
- Publicar dados no portal da ANAC – refere-se à estimativa de esforço para a atividade de publicação do dado a ser aberto no portal de dados abertos da ANAC (www.anac.gov.br/dadosabertos). Além dos conjuntos de dados abertos, esse ambiente hospeda também seus respectivos arquivos de metadados, os quais formam o catálogo de dados abertos da ANAC em formato eletrônico;
- Cadastramento no Portal Brasileiro de Dados Abertos – similar à atividade de catalogação no portal da ANAC, deve-se estimar o esforço necessário (em dias) para que as informações referentes ao dado aberto estejam disponíveis no portal de dados abertos do Governo Federal. Essa é a última atividade que completa o ciclo de abertura de dados.

4.2.1.2.5 Elaboração ou revisão do PDA

As atividades anteriores subsidiarão a elaboração ou revisão do plano, que é o principal produto desse processo. Dentre os elementos que devem integrar o documento, o plano precisa prever: os canais de comunicação, as formas de interação com a sociedade, cronograma com prazos e responsabilidades, bem como obediência às metodologias referenciadas para catalogação. Em suma, trata-se da atividade de materialização do plano com atenção especial à adequação aos requisitos previstos para sua elaboração.

A elaboração ou revisão do PDA poderá ensejar ajustes no Plano Diretor de Tecnologia da Informação – PDTI, em consonância com a Instrução Normativa nº 04, de 11 de setembro de 2014, com vistas a autorizar ações ou investimentos ainda não previstos.

4.2.1.2.6 Aprovação e publicação

Conforme definido no art. 6º da Resolução CGINDA nº 3, de 2017, do Comitê Gestor da Infraestrutura de Dados Abertos do Ministério do Planejamento (2017), o processo de aprovação e publicação é o compromisso da Direção da Agência com os termos descritos no plano, é o marco da finalização da elaboração ou revisão do plano e autoriza o início da próxima etapa que é a materialização das atividades do plano de ação para cada conjunto de dados de acordo com o cronograma do plano.

“Art. 6º - Os Planos de Dados Abertos deverão ser aprovados e instituídos pelo dirigente máximo do órgão ou entidade e publicados em transparência ativa, na seção "Acesso à Informação" do sítio eletrônico de cada órgão, nos termos do 'Guia de publicação ativa nos sítios eletrônicos dos órgãos e entidades do Poder Executivo Federal', disponível no Portal de Acesso à Informação (<http://www.acessoainformacao.gov.br/lai-para-sic/sic-apoio-orientacoes/guias-e-orientacoes>).” (BRASIL, PLANEJAMENTO, 2017)

4.2.1.2.7 Execução das ações previstas no plano

A etapa de execução das ações do PDA é o momento em que são realizados os trabalhos previstos no plano de ação referente a cada conjunto de dados, são as atividades de ajuste de procedimentos internos, estruturação dos dados para abertura, descrever os metadados, higienizar dados, automação de rotinas para publicação e finalizando nas atividades de publicação dos dados.

Cada conjunto de dados demandará um gerenciamento específico para sua realização, será necessária a alocação de recursos técnicos, envolvimento do curador designado e atividades de gestão das ações.

A Superintendência de Tecnologia da Informação gerenciará as ações técnicas do plano enquanto o curador conduzirá as atividades relacionadas a adequação de procedimentos internos previstas no plano de ação, além de avaliar o impacto das ações previstas no PDA sobre o PDTI e a necessidade de revisão deste.

4.2.1.2.8 Monitoramento

O processo de monitoramento consiste de atividades necessários para acompanhar, analisar e organizar o progresso e o desempenho das atividades do plano;

identificar mudanças de cenários ou necessidades de ajustes ou de elaboração de revisão do plano.

Essas atividades são executadas durante todo o ciclo do processo de abertura de dados e devem ser registradas lições aprendidas, necessidades de mudanças, de novas revisões do plano e de desempenho das atividades.

Um ponto de extrema importância dentro dessa atividade é a criação de um canal de comunicação entre a sociedade e a Agência que seja capaz de resolver problemas e melhorar o processo. Cabe destacar que no portal de dados abertos do Governo Federal (www.dados.gov.br) existe a previsão de, ao cadastrar o conjunto de dados a ser publicado, informar qual o canal de contato com o cidadão por meio da divulgação de conta de e-mail do autor e do mantenedor do conjunto de dados, além de haver a possibilidade de registro booleano sobre a satisfação com os dados, do tipo “Like” e “unlike”, conforme abaixo:

Estes dados estão disponíveis como o esperado?



Ocorre que ambas as formas não são efetivas para a sociedade, em especial ao público-alvo a que se destinam os dados abertos: pesquisadores, estatísticos, desenvolvedores de software, estudantes, enfim, um público especializado que suas dúvidas não são triviais, não convém que dependam de procedimentos lentos de resposta ou que se limitem a registrar algo do tipo “foi útil” ou “não foi útil”.

Para que esse mecanismo seja efetivo, deve existir um processo eletrônico do tipo *workflow* e que o cidadão possa acompanhar e interagir com quem está lhe respondendo dentro da Agência, além da existência de acordos de tempos mínimos de resposta, aos moldes dos conhecidos SLA (do inglês: acordo de nível de serviço).

Uma lista dessas questões com as respectivas respostas deverá ser disponibilizada no próprio portal de dados abertos da Agência para auxiliar outros usuários com questões similares.

A execução do monitoramento é de responsabilidade de duas áreas na Agência, cada uma com papéis definidos:

A Superintendência de Tecnologia da Informação é responsável pelas questões técnico/operacionais para abertura de dados e tem suas competências descritas na Política de Governança de Informações Digitais (BRASIL, ANAC, 2017), art. 8º.

A Ouvidoria, que é a autoridade designada pelo art. 40 da Lei nº 12.527, de 2011 (Lei de Acesso a Informação), tem suas competências definidas no art. 5º, § 4º, do Decreto nº 8.777, de 2016 (Política de Dados Abertos do Poder Executivo Federal).

4.2.1.3 Governança

O quadro a seguir representa a estrutura de governança do PDA. Algumas dessas funções são:

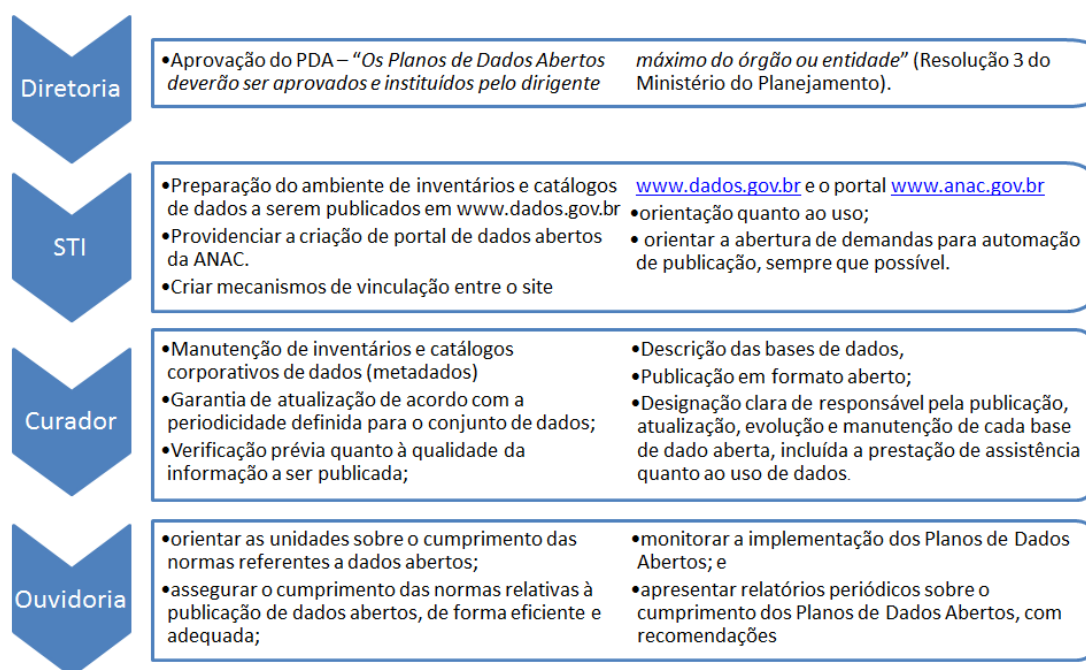
Diretoria da ANAC, como dirigente máximo, responsável pela aprovação do PDA no nível estratégico.

Curador, responsável pela avaliação da qualidade, descrição de metadados e persistência dos dados publicados.

Superintendência de Tecnologia da Informação, responsável pelas questões técnico/operacionais, viabilizar o desenvolvimento de soluções para automação das publicações e preparação dos ambientes.

Ouvidoria, responsável por orientar e assegurar o cumprimento das normas, monitorar as ações do plano e elaborar relatórios de cumprimento ao plano:

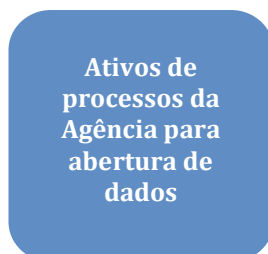
Figura 18 - Estrutura de governança do Plano de Dados Abertos da ANAC



Fonte: produzida pelo autor

4.3 Mecanismo 3 - Ativos de processos da Agência (normas, diretrizes e processos) para abertura de dados

Figura 19 - imagem representativa do mecanismo 3 para viabilizar transparência ativa e abertura de dados na ANAC



Fonte: produzida pelo autor

Esse mecanismo visa atribuir um engajamento cívico aos processos de transparência de dados, de forma que haja proatividade e antecipação de atitudes internas (de servidores) para envolver a sociedade na definição dos dados, prover canais de diálogo e fornecer gráficos e relatórios tendo como origem os dados disponibilizados e a necessidade manifestada da sociedade.

Entretanto, como dito em seções anteriores, a Administração Pública somente pode exercer e executar o que a lei autoriza e determina, não pode se afastar dos limites dos comandos legais sob pena de cometer ato inválido, desvio de poder ou excesso de poder.

Se por um lado esse princípio é uma garantia aos administrados (cidadão) contra excessos da administração, por outro lado ele pode ser um fator de inibição de iniciativas bem-intencionadas e com potencial para bons resultados.

Esse mecanismo tem a intenção de tão somente materializar o que foi previsto em atos normativos, não que seu objetivo seja de fornecer conformidade normativa ou uma meta burocrática, mas que sejam alcançados, e o mais importante: e mantidos os objetivos e motivações da existência desses atos.

Ele é focado em atividades que impactam no gerenciamento de dados (coleta de dados, desenvolvimento ou manutenção de sistemas, definição de dados a serem publicados), no ambiente de publicação dos dados abertos da Agência, na organização e forma de apresentação desses dados, no mecanismo de interação com a sociedade usuária desses dados e na necessidade de agregação ou geração de gráficos e relatórios com base em dados abertos. Novamente, esses paradigmas suportam o atingimento aos objetivos normativos.

Dessa forma, ativos de processos da Agência (normas, diretrizes e processos) necessitam refletir esses paradigmas, os quais são detalhados abaixo:

4.3.1 Captação e recebimento de dados

As atividades da ANAC estão relacionadas aos seus cinco macroprocessos:

I - Regulamentação: compreende os processos relacionados ao estabelecimento de requisitos a entidades do Sistema de Aviação Civil por meio da elaboração e atualização de atos normativos de competência da ANAC;

II - Certificação e Outorga: compreende os processos relacionados a verificação do atendimento a requisitos estabelecidos em atos normativos para que produto, empresa, processo, serviço ou pessoa possa prestar serviços, executar atividades ou ser operado dentro do Sistema de Aviação Civil;

III - Fiscalização: compreende os processos relacionados a verificação da conformidade de produtos, empresas, processos, serviços ou pessoas que atuam, de forma lícita ou ilícita dentro do Sistema de Aviação Civil, e a respectiva ação da Agência em caso de não conformidade;

IV - Relações Institucionais: compreende os processos de relacionamento da ANAC com entes externos, dentro e fora do Sistema de Aviação Civil, a exceção dos processos já relacionados a outros macroprocessos; e

V - Gestão Interna: compreende processos de suporte ou de gestão cujos clientes são servidores e áreas internas da Agência, de forma a manter ou melhorar processos internos, competências, estrutura e infraestrutura administrativa.

Esses macroprocessos, com exceção do V – Gestão Interna, são dependentes de dados e informações originadas das empresas do setor de aviação civil brasileiro, que são as entidades e organizações reguladas pela ANAC. Dessa forma, as empresas, de acordo com seu ramo de atividade, têm obrigação legal de enviar seus dados à ANAC. Como exemplo desses dados, cito os seguintes:

Dado	Descrição
Dados a serem enviados à ANAC pelas Concessionárias de Serviço Público de Infraestrutura Aeroportuária (BRASIL, ANAC, 2019):	
Programação anual de manutenção	A programação anual de manutenção que contempla o período de abril de 2019 a março de 2020, em planilha eletrônica editável (.xls), conforme padronização constante no Anexo I da Portaria SRA nº 449/2019, contendo a quantidade de manutenções

	planejadas, com a indicação da frequência e do tempo estimado de indisponibilidade do equipamento para execução da manutenção.
IQS Tempo na fila de inspeção de segurança - horários de medição	Mede os tempos de espera na fila de inspeção de segurança, diariamente, conforme os critérios definidos no Anexo III da Portaria SRA nº 449/2019.
PSP - resultado das entrevistas e áudio das entrevistas diretas	Encaminha até o 15º (décimo quinto) dia do mês subsequente ao mês da aferição, os resultados das entrevistas da PSP e áudio das entrevistas diretas, conforme Anexos X e XI da Portaria nº 449/2019.
PSP – lista de voos extras	Encaminha, mediante solicitação da ANAC, a lista dos voos extras utilizados
informações relativas à movimentação aeroportuária junto à ANAC.	Os administradores de aeroportos deverão encaminhar à ANAC, até o dia 25 de janeiro de cada ano, o Compilado de Movimentação Aeroportuária – CMA.
Operadores aéreos nacionais que tenham emissões de CO2 acima de 10.000 (dez mil) toneladas anuais pelo uso de aeronaves com peso de decolagem certificado acima de 5.700 kg, deverão enviar:	
Relatório Anual de Emissões	Até o último dia útil do mês de abril, a partir de 2020, o Relatório Anual de Emissões Verificado referente ao ano anterior, juntamente com o respectivo Parecer de Verificação.
Empresas que exploram os serviços de transporte aéreo doméstico regular de passageiros deverão enviar:	
Registro dos dados das tarifas aéreas comercializadas correspondentes aos serviços de transporte aéreo doméstico e internacional regular de passageiros.	Registrar na ANAC, até o último dia útil do mês subsequente, os dados das tarifas comercializadas no mês anterior correspondentes aos serviços de transporte aéreo doméstico e internacional regular de passageiros, de acordo com instruções, ou comunicar à ANAC caso não tenha emitido, no mês anterior, bilhetes de passagem correspondentes aos serviços de transporte aéreo regular doméstico e internacional de passageiros.
Percentuais de atrasos e cancelamentos de voos regulares, domésticos e internacionais de passageiros no Brasil	Os percentuais de atrasos e cancelamentos de voos serão mensalmente divulgados pela ANAC, por meio de 2 (dois) tipos de arquivos eletrônicos, em seu site.
Dados estatísticos do transporte aéreo do Brasil	Os dados são mensalmente fornecidos à ANAC, até o dia 10 do mês subsequente ao de referência, pelas empresas brasileiras e

	estrangeiras que exploram os serviços de transporte aéreo público regular e não regular no Brasil.
--	--

Esses são alguns exemplos de dados entrantes na Agência e processados por diversos sistemas de informação. Além disso, eles são fontes de trabalho da ANAC para diminuição da assimetria de informação entre o setor regulado, a Agência e a Sociedade.

Esses são também o principal foco das iniciativas de transparência e publicação de dados abertos. É por meio deles que a sociedade especializada (imprensa, pesquisadores, acadêmicos, mercado de software) divulgam informações, relatórios e gráficos mais aprimorados que a própria agência se propõe.

Cabe ainda esclarecer que qualquer determinação de obrigação às empresas reguladas é precedida de dois atos administrativos para que haja validade e eficácia: a) consulta ou audiência pública para que conhecimento prévio e oportunidade de manifestação pelas empresas, e b) norma, derivada da consulta ou audiência, contendo todos os detalhes da obrigação.

Algumas das normas que determinam o envio de dados à ANAC são:

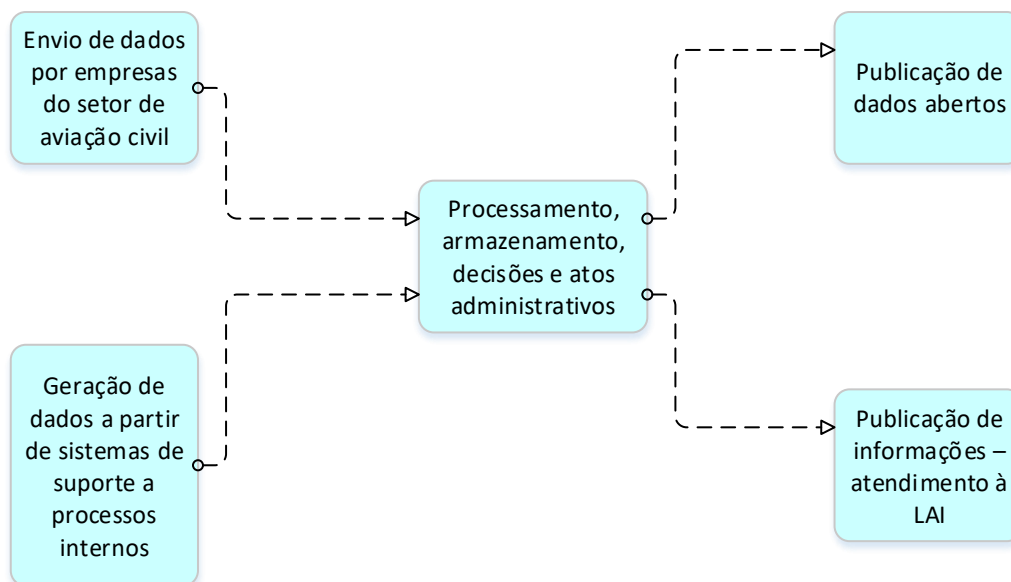
- RESOLUÇÃO Nº 464, 22/02/2018 - Regulamenta a apresentação de informações relativas à movimentação aeroportuária e aprova a Emenda nº 02 ao RBAC nº 153;
- RESOLUÇÃO Nº 496, 28/11/2018 - Regulamenta o monitoramento, o reporte e a verificação de dados de emissão de CO2 relativos ao transporte aéreo internacional.
- Resolução nº 191 de 16/06/2011 - Regulamenta o fornecimento de dados estatísticos relativos aos serviços de transporte aéreo público.
- PORTARIA Nº 3.542/SRA, DE 19 DE NOVEMBRO DE 2018 - Dispõe sobre a apresentação das informações estabelecidas na Resolução nº 372, de 15 de dezembro de 2015; na Decisão nº 205, de 20 de dezembro de 2017 e na Resolução nº 453, de 20 de dezembro de 2017.

Conforme dito anteriormente, a edição e publicação desses atos normativos foi precedida de consulta pública ou audiência pública com participação de diversos setores da sociedade.

A partir do envio desses dados enseja-se o tratamento, processamento, armazenamento e decisões internas e externas, conforme o caso – por exemplo: multa por atraso de envio -, com suporte de recursos de tecnologia da informação.

Diagrama de alto nível de fluxo de atividades para publicação de dados e transparência:

Figura 20 - imagem representativa do fluxo de atividades para publicação de dados e transparência



Fonte: produzida pelo autor

Os dados enviados pelas empresas e os dados de processamento interno são tratados por meio de sistemas desenvolvidos, pela área de tecnologia da Agência, ou adquiridos por meio de licitação.

4.3.2 Dados gerados a partir de sistemas de gestão

Outra canal de geração de dados na Agência é por meio de seus sistemas internos de gestão, que são mecanismos para controle de seus procedimentos, por exemplo: processos licitatórios, registros de procedimentos para concessão de aeroportos, deliberação pela Diretoria de processos de sanção administrativa, julgamento (administrativo) de autos de infração, processo para concurso público, registros de servidores, registros de receitas, e diversos outros atos administrativos internos.

Esses dados são mantidos por sistemas desenvolvidos internamente ou adquiridos, seja por compra (licitação), seja por adoção de software de outro órgão da Administração Pública. As aquisições para desenvolvimento, manutenções, correções ou evoluções de soluções de TI são de competência da área de tecnologia da Agência, a Superintendência de Tecnologia da Informação – STI.

A forma como a STI deve atuar é baseada, principalmente, em duas Instruções Normativas: a 115/2017 e a 120/2018. Já para aquisições de soluções e bens ela é submetida

aos ditames da Instrução Normativa nº 4 do Ministério do Planejamento (atualmente, Ministério da Economia), que disciplina o processo de planejamento, aquisição e fiscalização da contratação de bens e serviços de TI.

As atividades de desenvolvimento, manutenção, correção ou evolução de sistemas devem ser executadas conforme previsões na Metodologia de Desenvolvimento e Manutenção de Sistemas da STI.

O Mecanismo 3 visa proceder a ajustes nos três artefatos abaixo:

- Instrução Normativa nº115 de 14 de agosto de 2017, que Institui a Política de Governança de Informações Digitais - PGID da ANAC;
- Instrução Normativa nº 120, de 22 de fevereiro de 2018, que Institui a Política de Governança de Tecnologia da Informação e Comunicação - PGTIC da ANAC.
- Metodologia de Desenvolvimento e Manutenção de Sistemas da ANAC

A Instrução Normativa 115 tem como objetivo a criação de um arcabouço para governança e gestão de dados.

Este mecanismo prevê ajustes nessa Instrução Normativa de forma a exigir participação do curador de dados em demandas de desenvolvimento e manutenção de sistemas. De acordo com aquele normativo, curador: servidor público, designado pelo superintendente ou chefe de unidade, com a responsabilidade de gerenciar ativos de informação específicos.

Este mecanismo prevê a participação antecipada do curador de dados em avaliação de necessidade de planejamento de publicação de dados abertos, em casos decorrentes do processo de desenvolvimento ou manutenção de software, ou seja: caso ocorra qualquer demanda de desenvolvimento ou manutenção de sistema que impacte em mudança de estrutura de dados, o curador do dado impactado deverá ser envolvido para que possa avaliar a necessidade de iniciar um planejamento de transparência e abertura de dado referente à mudança na estrutura de dados afetada, conforme abaixo:

Art. 12 Os curadores exercerão as seguintes atividades em consonância com os padrões, políticas e diretrizes aprovados:

- I - identificar e resolver eventuais problemas de informações digitais;
- II - definir e manter requisitos, regras de negócio, indicadores e metas para a qualidade de informações digitais;
- III - executar as ações necessárias visando garantir a qualidade de informações digitais;
- IV - garantir que os requerimentos de negócio sejam refletidos em requerimentos de informações digitais e regras de dados;
- V - definir regras de acesso aos dados;

VI - manter atualizada a documentação sobre informações digitais no Catálogo de Informações Digitais;

VII - garantir que demandas que impactem estruturas de dados sob sua curadoria sejam encaminhadas a processo para publicação de dados abertos

VIII - definir requisitos para armazenamento, guarda e descarte de informações digitais;

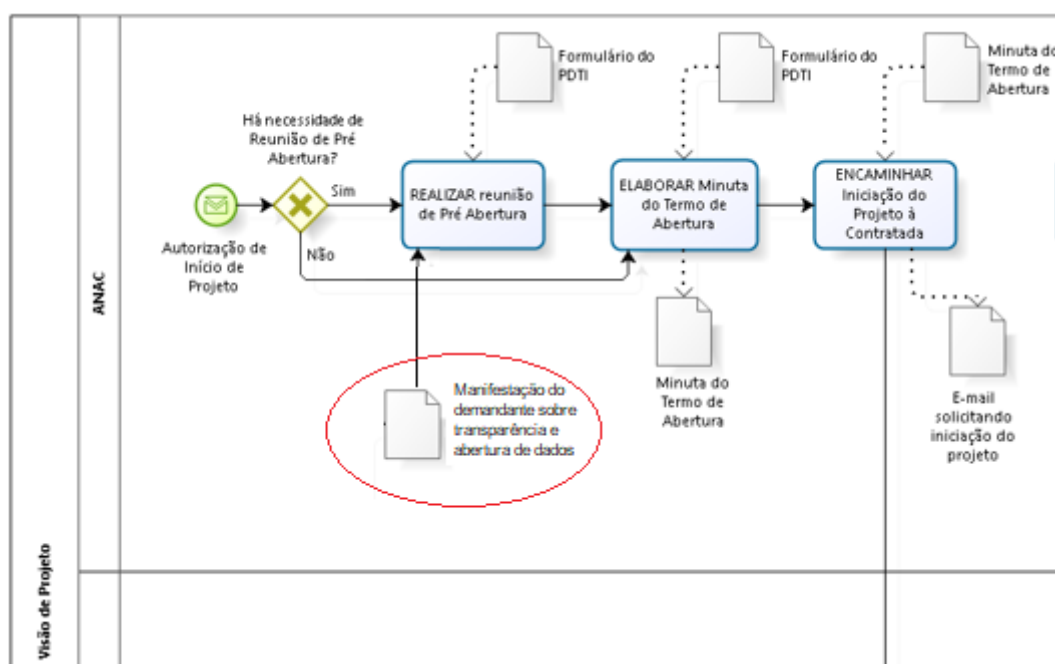
VII - IX - exercer outras competências delegadas pelo Comitê de Tecnologia de Informação relativas a gestão de informação.

Parágrafo único. Toda ação que envolva informações digitais deve ser previamente aprovada pelo curador responsável. (BRASIL, 2017).

Outra mudança é na Metodologia de Desenvolvimento e Manutenção de sistemas. Essa metodologia tem como finalidade padronizar os processos de trabalho, as funções desempenhadas e as tarefas a serem realizadas por todos os envolvidos no desenvolvimento e manutenção de software da ANAC.

A mudança prevê a participação do curador de dados juntamente com o demandante da solução, logo na iniciação do projeto de desenvolvimento ou manutenção de sistema para avaliar impactos da mudança na estrutura de dados sob sua curadoria. Nesse momento deverá ser gerado um artefato explicitando a mudança e ensejando a ação para o processo de transparência e abertura de dados, o qual entrará na ação de planejamento de abertura de dados. Isso significa que a partir de mudança na estrutura de base de dados transacional, inicia-se o planejamento para publicação de dados de interesse da sociedade.

Figura 21 - Fluxo de Projeto de Desenvolvimento e Manutenção ajustado



Fonte: (BRASIL, ANAC, 2010)

4.3.3 Divulgação de dados e informações

Todos os dados recebidos das empresas do setor de aviação civil ou gerados e mantidos pelos sistemas internos -, desde que não ofendam direitos individuais, não sejam sigilosos ou coloquem em risco a segurança nacional, devem ser divulgados de acordo com a Lei de Acesso à Informação e o com Decreto de dados abertos.

Para os dados divulgados com base no Decreto 8.777 há critérios específicos como formato aberto, livre de licença, garantia da periodicidade de atualização, disponibilização de canal de contato com a sociedade, permissão irrestrita de uso. Já a Lei de Acesso à informação explicita um rol de informações que devem ser divulgadas: dados da estrutura organizacional, despesas, repasses ou transferência de recursos financeiros, procedimentos licitatórios, e outros detalhes técnicos:

§ 3o Os sítios de que trata o § 2o deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão;

II - possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;

III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;

IV - divulgar em detalhes os formatos utilizados para estruturação da informação;

V - garantir a autenticidade e a integridade das informações disponíveis para acesso;

VI - manter atualizadas as informações disponíveis para acesso;

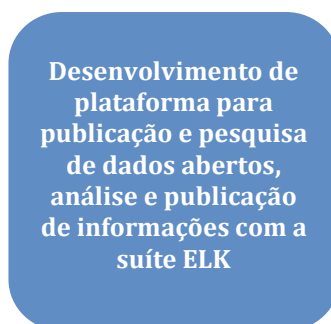
VII - indicar local e instruções que permitam ao interessado comunicar-se, por via eletrônica ou telefônica, com o órgão ou entidade detentora do sítio;(Lei de acesso à Informação)

Além da publicação dos dados e informações conforme previstos nos dois normativos, deverão ser realizadas tarefas para transformar os dados em informação, identificar padrões e aplicar técnicas de visualização de dados.

Essa atividade será melhor detalhada no mecanismo 4 indexação e recuperação de dados abertos e metadados de dados abertos com Elasticsearch divulgação de informação, na próxima seção.

4.4 Mecanismo 4 – Desenvolvimento de plataforma para publicação e pesquisa de dados abertos, análise e publicação de informações com a suíte ELK

Figura 22 - imagem representativa do mecanismo 4 para viabilizar transparência ativa e abertura de dados na ANAC



Fonte: produzida pelo autor

O Mecanismo 4 trata do desenvolvimento de uma plataforma para prover recursos de transparência e de comunicação com a sociedade. O objetivo é fornecer um ambiente que contemple acesso aos dados abertos publicados, pesquisa, gráficos, painéis e relatórios, com a aplicação de técnicas de visualização de dados e descoberta de padrões com suíte ELK Stack. (ELASTIC, s.d.).

"ELK" é o acrônimo para três projetos de código aberto: Elasticsearch, Logstash e Kibana. A função de cada componente do stack é assim definida:

ElasticSearch

O Elasticsearch (ELASTIC, s.d.) tem a função de fazer a indexação de dados e consulta a estes índices e dados, provendo real-time analytics. Possui outras características como Escalabilidade, Alta-disponibilidade, multi tenant (quando um sistema multiusuário tem funcionamento específico e independente para cada usuário), Full text search.

A forma de trabalho do ElasticSearch com relação a índices leva a uma confusão quando raciocinamos de forma análoga a um índice em banco de dados relacional, entretanto os conceitos são diferentes: um documento é a unidade de pesquisa e índice. Um índice consiste em um ou mais documentos e um documento consiste em um ou mais campos.

Na terminologia do banco de dados relacional, um Documento corresponde a uma linha da tabela e um Campo corresponde a uma coluna da tabela. Na plataforma ES pode-se criar mais de um índice para finalidades diferentes.

Outra característica do ES é o mecanismo de tokenizadores ou “tokenizer”. Os tokenizadores são usados para dividir o conteúdo dos documentos em um fluxo de termos ou tokens. Um tokenizer simples pode dividir a string em termos onde encontrar espaço em branco ou pontuação. O ES tem diversos tipos tokenizers embutidos que podem ser usados para construir analisadores personalizados. Além disso, é possível também implementar um novo tokenizer customizado, de acordo com suas necessidades e características dos dados.

Exemplos de tokenizers:

a) Letter Tokenizer - Divide os tokens sempre que encontra um caractere que não é letra:

Comando:

```
POST _analyze
{
  "tokenizer": "letter",
  "text": "The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."
}
```

Resultado:

```
[ The, QUICK, Brown, Foxes, jumped, over, the, lazy, dog, s, bone ]
```

b) Lowercase Tokenizer - Também divide os tokens sempre que encontra um caractere que não é letra, mas coloca todos tokens em lowercase:

Comando:

```
POST _analyze
{
  "tokenizer": "lowercase",
  "text": "The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."
}
```

Resultado:

```
[ the, quick, brown, foxes, jumped, over, the, lazy, dog, s, bone ]
```

c) Whitespace Tokenizer - Divide os tokens sempre que encontra um espaço em branco

Comando:

```
POST _analyze
```



```
{
  "tokenizer": "whitespace",
  "text": "The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."
}
```

Resultado:

[The, 2, QUICK, Brown-Foxes, jumped, over, the, lazy, dog's, bone.]

d) Standard Tokenizer - É o tokenizer padrão do ES, que funciona bem na maioria dos idiomas:

Comando:

POST _analyze

```
{
  "tokenizer": "standard",
  "text": "The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."
}
```

Resultado:

[The, 2, QUICK, Brown, Foxes, jumped, over, the, lazy, dog's, bone]

e) UAX URL Email Tokenizer - Funciona como o tokenizer padrão, mas trata emails e URLs como um token único:

Comando:

POST _analyze

```
{
  "tokenizer": "uax_url_email",
  "text": "Email me at john.smith@global-international.com"
}
```

Resultado:

Não há quebra no nome da conta de e-mail: [Email, me, at, john.smith@global-international.com]

f) Classic Tokenizer - É uma heurística que funciona muito bem no idioma Inglês. Ele considera e-mails e URLs como um só token, separa tokens em hifens sempre que não existe números no token e separa os tokens em quase todos caracteres de pontuação.

Comando:

```
POST _analyze
{
  "tokenizer": "classic",
  "text": "The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."
}
```

Resultado:

```
[The, 2, QUICK, Brown, Foxes, jumped, over, the, lazy, dog's, bone ]
```

g) NGRAM Tokenizer - Divide o texto em n-gramas (Default: até 2 caracteres) e trata cada um deles como um token diferente:

Comando:

```
POST _analyze
{
  "tokenizer": "ngram",
  "text": "Quick Fox"
}
```

Resultado:

```
[ Q, Qu, u, ui, i, ic, c, ck, k, "k ", " ", " F", F, Fo, o, ox, x ]
```

h) Edge NGRAM Tokenizer - Recomendado para buscas em tempo de digitação (search-as-you-type). Ele divide os primeiros n caracteres de cada palavra do texto em diferentes tokens.

Logstash

Similar a um ETL, é a ferramenta que faz captura das informações em arquivos, processa e gera resultados para o Elasticsearch ou outro formato de dados.

O Logstash (ELASTIC, s.d.) é um mecanismo de coleta de dados baseado em código aberto com recursos de “pipelining” (ou enfileiramento de instruções) de execução em tempo real. O Logstash pode unificar dinamicamente dados de fontes diferentes e normalizar os dados em destinos de sua escolha.

Embora o Logstash originalmente tivesse como foco a coleta de logs, seus recursos se estendem muito além desse caso de uso. Qualquer tipo de origem de dado pode ser

enriquecido e transformado com uma ampla variedade de plugins de entrada, filtro e saída, com muitos codecs nativos simplificando ainda mais o processo de ingestão.

Como funciona o Logstash

O pipeline (uma pilha de comandos) de processamento de eventos do Logstash possui três estágios: entradas → filtros → saídas. As entradas geram eventos, os filtros as modificam e as saídas as enviam para outro lugar. Entradas e saídas suportam codecs que permitem codificar ou decodificar os dados à medida que eles entram ou saem do pipeline sem ter que usar um filtro separado.

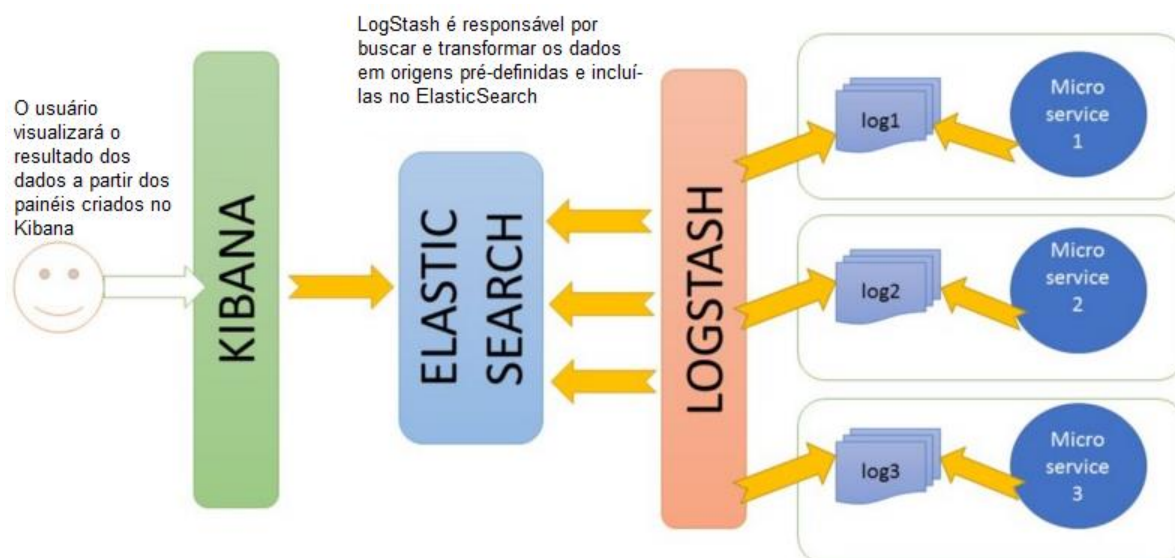
Um caso de uso é a utilização de origens de dados abertos para incorporação a índices no ES usando o Logstash.

Kibana

O Kibana (ELASTIC, s.d.) é uma ferramenta baseada em código aberto para exploração e visualização de dados. É especializada em grandes volumes de dados em tempo real e streaming. O software facilita a compreensão e a interpretação de grandes e complexos fluxos de dados referenciados a partir do Elasticsearch.

A suíte ELK trabalha integrada de forma a automatizar todo o processo que tem como início a origem pré-definida dos dados, até a apresentação da informação tratada em formato específico, como painéis e gráficos:

Figura 23 – interação entre produtos da suíte ELK



Fonte: (KHANNA, 2019) com adaptações

4.4.1 Realização de piloto com a suíte ELK

Para concepção do mecanismo 4 com o uso da suíte ELK foi realizado um teste piloto contemplando instalação, importação de dados a partir de dado aberto, preparação e tratamento de dado e publicação de gráfico experimental.

Instalação:

Foi realizada a instalação dos produtos Elasticsearch (elastic stack) versão 6.6 e o Kibana 6.6.2

Os produtos foram instalados em um notebook Intel Core i3 CPU 2Ghz com 8mb de memória Ram, barramento de 64 bits, com sistema operacional Windows 10 home single language.

Importação de dados a partir de conjuntos de dados abertos:

Os dados selecionados foram seis arquivos com extensão CSV contendo dados estatísticos relativos aos serviços de transporte aéreo público. Cada arquivo contém 7 colunas, sendo elas: ano de realização, mês de realização código ICAO da empresa, código ICAO do aeródromo, Tarifa-N e assentos comercializados. Todos os arquivos estão disponíveis na página de dados abertos da ANAC, www.anac.gov.br/dadosabertos.

Cada arquivo tem aproximadamente 500.000 linhas.

Os arquivos foram importados utilizando recurso do Kibana. Um outro produto da suíte ELK (logstash) é especializado em tratamento de dados antes da carga, entretanto, não foi necessário utiliza-lo em função do escopo reduzido do teste piloto. Porém, é recomendável usar uma solução como o logstash que possibilite agilidade nos momentos de carga de dados.

Preparação e tratamento de dado:

Os arquivos escolhidos para participarem do teste piloto são disponibilizados no portal de dados abertos da ANAC (www.anac.gov.br/dadosabertos), na área temática voos e operações aéreas, link Dados Estatísticos do Transporte Aérea. Nessa página são disponibilizados os arquivos dos dados estatísticos em formato CSV e um arquivo de metadados que tem a explicação sobre o conjunto de dados.

Para que fosse possível a importação pelo Kibana, apenas um ajuste foi necessário em cada um dos arquivos a serem utilizados no teste piloto. O formato de casa decimal da coluna Tarifa-N não estava adequado ao exigido pelo Kibana, o separador “.” foi substituída pelo separado “,”.

Publicação de gráfico experimental:

A construção do gráfico experimental foi realizada no Kibana. Foi construído usando o item “Visualize”, que é um módulo para desenvolvimento de gráficos e relatórios do produto.

Foi construído um gráfico do tipo “heat map”, que são células dentro de uma matriz. Para a composição do gráfico foram usados os campos empresa e origem, o gráfico foi filtrado para apresentar três empresas aéreas: GLO, TAM E PTB.

O resultado do gráfico é apresentado abaixo:

Figura 24 - gráfico do tipo “heat map” como resultado do piloto



Fonte: produzido pelo autor

Esse mesmo gráfico foi incorporado a um outro recurso do Kibana para apresentação de dashboards. O teste de criação de dashboards considerou somente esse gráfico, entretanto a proposta da solução é a incorporação de diversos gráficos em um mesmo ambiente. O resultado é apresentado abaixo:

Figura 25 - gráfico "heat map" incorporado em dashboard construído no piloto



Fonte: produzido pelo autor

4.4.2 Conclusão

Atualmente diversos gráficos, painéis e relatórios são gerados, na agência por diversas áreas, entretanto de forma descentralizada, desintegrada (vários deles até poderiam ser analisados em conjunto para gerar resultados mais inteligentes) e em espalhados em diversos endereços no portal da Agência.

O objetivo desse piloto dentro do escopo dos mecanismos apresentados é demonstrar a viabilidade da criação de um ambiente integrado que proveja inteligência aos dados mantidos pela Agência, tendo um ecossistema de governança de informações digitais que proveja continuidade, controle, delegação, competência e responsabilização aos envolvidos. Esse ambiente integrado será um portal de inteligência de informações da ANAC.

5 CONCLUSÃO E TRABALHOS FUTUROS

Esse trabalho propôs atingir um objetivo geral de elaboração de mecanismos, dentro de um contexto de governança de informações digitais, que promovam ações de transparência e abertura de dados, de forma ativa, que auxilie na diminuição da assimetria de informação entre a Agência, empresas do setor de aviação civil reguladas pela ANAC e a sociedade.

Os estudos realizados e registrados no referencial teórico auxiliaram na fundamentação e elaboração desses mecanismos. Foram apresentadas as evoluções na temática de transparência e abertura de dados desde meados do século XVIII em diversos países; a importância do tema para a sociedade de pesquisadores e acadêmica, os quais foram precursores no uso e compartilhamento de informações de pesquisa.

Outros pontos que auxiliaram na definição desses mecanismos foram estudos que apontam dificuldades com relação a abertura de dados, utilização pela sociedade e a manutenção do processo de publicação de dados pelos órgãos da Administração Pública; questões como complexidade técnica das bases de dados disponibilizadas, que pode inibir seu uso, a qualidade dos dados disponibilizados, como dados incompletos e desatualizados, os quais demonstram precariedade de governança; e a exigência da publicação de dados brutos, desagregados, conforme preconiza o Decreto de Dados Abertos.

Nesse ponto há uma vertente de pensamento, apresentada pela Association of Government Accountants, que acredita que os dados deveriam ser agregados e apresentados de forma amigável e compreensível, além de serem distribuídos por diversos meios e mídias, inclusive a internet.

Esse pensamento sobre publicação de dados agregados leva a outras questões: quem deve processar esses dados para que eles façam sentido? O próprio governo? seria a sociedade organizada, por meio dos seus ativistas, militantes, hackers e suas associações não governamentais? Além dessas, há uma outra importante questão a ser considerada: a interpretação por conveniência por parte de quem processa os dados. A falta de isenção a depender do interesse do ator pode levar a risco de interpretações.

A partir desses estudos, foram elaborados 4 mecanismos que tendem a diminuir esses problemas: os mecanismos 1 e 2 (respectivamente: implantação de um modelo de governança de informações digitais na ANAC e institucionalização de um processo de gestão para o plano de abertura de dados) visam criar um ecossistema de governança e gestão para

abertura de dados dentro de um contexto de informações digitais na Agência de forma a monitorar a qualidade, a completude e a atualidade dos dados, além de criar um compromisso entre o ator publicador e os atores geradores dos dados.

Já o mecanismo 3 - Ativos de processos da Agência (normas, diretrizes e processos) para abertura de dados -, define papéis, responsabilidades e competências dos atores. Isso, dentro de um contexto de Administração Pública, apesar de aparente burocracia, é um fator poderoso para ação dos envolvidos.

Sua função vai além de legitimar a execução desse ecossistema, uma vez que é exigência legal o atendimento ao Princípio da Legalidade na Administração Pública,

O mecanismo 4 – Desenvolvimento de plataforma para publicação e pesquisa de dados abertos, análise e publicação de informações com a suíte ELK -, é a materialização dos outros três mecanismo dentro de um ambiente de comunicação com a sociedade, por meio de um portal de dados abertos, de forma a disponibilizar ao cidadão dados, informações e um canal de contato com a Agência para questões sobre as informações disponibilizadas.

A escolha da suíte ELK se deu em função de suas características de software livre e seu potencial de recursos que contemplam desde a busca, com recursos de “tokenizers” para cada tipo de dado a ser pesquisado, serviços de extração, tratamento e carga de dados e geração de diversos formatos de gráficos e relatórios.

A definição de uma solução para os problemas apontados, dividida em quatro mecanismos, tende a aumentar a disponibilidade, qualidade e uso dos dados publicados pela Agência, entretanto, essa abordagem se torna minúscula se enxergada de forma restrita ao setor de aviação civil. Se for explorada sob o viés de regulação do setor de transporte, de forma ampla: aviação civil, transporte terrestre e aquaviário, o valor da informação para a sociedade especializada e para os tomadores de decisão será mais valiosa; mais ainda: se for enxergada não somente como uma estrutura disponibilizada às atividades de regulação do setor de transporte, mas como uma estrutura de suporte às políticas públicas para o setor de transporte, o potencial de exploração desses dados e informações será em nível de resultados para a sociedade de forma ampla.

Assim, recomenda-se a continuidade deste trabalho com duas vertentes complementares entre si e complementares a este trabalho:

- 1) desenvolvimento de mecanismos que promovam ações de transparência e abertura de dados, aos moldes deste trabalho, para o setor de transporte de forma ampla; e
- 2) desenvolvimento de mecanis

mos que promovam ações de transparência e abertura de dados para suporte às políticas públicas do setor de transporte.

6 REFERÊNCIAS

- AGUNE, R. M., GREGÓRIO FILHO, A. S., & BOLLIGER, S. P. (2010). Governo aberto: disponibilização de base de dados e informações em formato aberto. In: *Congresso Consad de Gestão Pública*, 3. CONSAD: III
- ANP. (2015). Relatório de Análise de Impacto Regulatório. Acesso em 14 de Fevereiro de 2019, Disponível em http://www.anp.gov.br/images/Consultas_publicas/Concluidas/2015/n10/Relatorio_Analise_Impacto_Regulatorio.pdf
- ARAÚJO, E. N. (2006). *Curso de Direito Administrativo*. São Paulo: Saraiva.
- ASSEMBLEIA GERAL DAS NAÇÕES UNIDAS. (1948). *Direitos Humanos - Declaração Universal dos Direitos Humanos*. Fonte: UNICEF: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>
- BALSTAD, R. (2012). *Overview of Scientific Data Policies*. Columbia University, Estados Unidos: The National Academies Press. doi:10.17226/17019
- BNDES. (2017). Produto 2 Roadmap tecnológico.
- BOER, D. (1998). *Steamy Windows: Transparency and Openness in Justice and Home Affairs*. In *Openness and Transparency in the European Union*. (V. Deckmyn, & I. Thomson, Eds.) Maastricht, Netherlands: European Institute of Public Administration.
- BOVENS, M. B. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism, *West European Politics*. doi:10.1080/01402382.2010.486119
- BRAINER, L. (14 de Abril de 2015). *Quando surgiram os 8 princípios dos dados abertos?* Acesso em 26 de Janeiro de 2019, disponível em OPEN KNOWLEDGE BRASIL: <https://br.okfn.org/2015/04/14/quando-surgiram-os-oito-principios-dos-dados-abertos/>
- BRASIL. (1988). Constituição da República Federativa do Brasil de 1988.
- BRASIL. (1994). *Decreto nº 1.171: Aprova o Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal*.
- BRASIL. (1995). Lei nº 8.987, de 13 de fevereiro de 1995: Dispõe sobre o regime de concessão e permissão da prestação de serviços públicos previsto no art. 175 da Constituição Federal, e dá outras providências. Acesso em 11 de Fevereiro de 2019, disponível em http://www.planalto.gov.br/ccivil_03/LEIS/L8987cons.htm
- BRASIL. (2000). *Lei Complementar nº 101: Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências*.

BRASIL. (23 de Fevereiro de 2005). Decreto nº 5.378: Institui o Programa Nacional de Gestão Pública e Desburocratização - GESPÚBLICA e o Comitê Gestor do Programa Nacional de Gestão Pública e Desburocratização, e dá outras providências.

BRASIL. (2011). LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011.

BRASIL. (16 de maio de 2012). Decreto 7.724 de 16 de maio de 2012, que Regulamenta a Lei no 12.527, de 18 de novembro de 2011, que dispõe sobre o acesso a informações;. Acesso em 15 de Janeiro de 2019, disponível em http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7724.htm

BRASIL. (16 de Maio de 2013). Lei nº 12.813: Dispõe sobre o conflito de interesses no exercício de cargo ou emprego do Poder Executivo federal e impedimentos posteriores ao exercício do cargo ou emprego.

BRASIL. (11 de maio de 2016). Decreto 8.777 de 2016, Institui a Política de Dados Abertos do Poder Executivo federal. Brasil. Acesso em 1 de Fevereiro de 2019, disponível em http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8777.htm

BRASIL. (15 de Janeiro de 2016a). Decreto nº 8.638: Institui a Política de Governança Digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. Acesso em 28 de Novembro de 2018, disponível em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8638.htm

BRASIL. (29 de Junho de 2016b). DECRETO Nº 8.789: Dispõe sobre o compartilhamento de bases de dados na administração pública federal.

BRASIL. (2018). LEI 13.709: Lei Geral de Proteção de Dados Pessoais. Acesso em 2 de Janeiro de 2019, disponível em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm

BRASIL, ANAC. (2010). MDMS: Metodologia de Desenvolvimento de Sistemas.

BRASIL, ANAC. (14 de Agosto de 2017). Instrução Normativa nº 115, 14/08/2017: Institui a Política de Governança de Informações Digitais - PGID da ANAC.

BRASIL, ANAC. (2017). Nota técnica 0692564 no Processo nº 00058.515309/2017-91.

BRASIL, ANAC. (2018). *Plano de Dados Abertos*. Acesso em 3 de Janeiro de 2019, disponível em ANAC: <https://www.anac.gov.br/dadosabertos>

BRASIL, ANAC. (2019). Portaria nº 449 SRA: Aprova o Compêndio de Elementos de Fiscalização - CEF da Portaria nº 449/SRA, de 11 de fevereiro de 2019.

BRASIL, CGU. (2016). *Governo Aberto*. Fonte: O que é Governo Aberto: <http://governoaberto.cgu.gov.br/a-ogp/o-que-e-governo-aberto>

BRASIL, Decreto. (Setembro de 2011). DECRETO DE 15 DE SETEMBRO DE 2011: Institui o Plano de Ação Nacional sobre Governo Aberto e dá outras providências.

BRASIL, Lei. (18 de Novembro de 2011a). Lei 12.527 de 2011, Regula o acesso a informações. Brasil. Acesso em 1 de Fevereiro de 2019, disponível em http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

BRASIL, PLANEJAMENTO. (13 de Outubro de 2017). CGINDA: Resolução nº 3: Aprova as normas sobre elaboração e publicação de Planos de Dados Abertos.

BRASIL, PLANEJAMENTO. (2018). Boas práticas, vedações e orientações para contratação de software e de serviços. Acesso em 11 de Abril de 2019, disponível em <https://www.governodigital.gov.br/documentos-e-arquivos/Contratacao%20de%20Softwares%20e%20de%20Servicos%20de%20Desenvolvimento%20e%20Manutencao%20de%20Sistemas%20-Fabrica%20de%20Software.pdf>

BRASIL, PLANEJAMENTO. (2018a). Plataforma de Análise de Dados. Acesso em 11 de Abril de 2019, disponível em <http://www.planejamento.gov.br/govdata-privado/plataforma-de-analise-de-dados-govdata/govdata>

BRASIL, STN. (2015). *Tesouro Nacional Transparente*. Acesso em 29 de Janeiro de 2019, disponível em Tesouro Nacional Transparente: <https://www.tesourotransparente.gov.br/>

BRASIL, STN. (2015a). *Fluxo de Receitas e Despesas*. Acesso em 1 de Abril de 2019, disponível em Tesouro Nacional Transparente: <https://www.tesourotransparente.gov.br/historias/fluxo-de-receitas-e-despesas>

BRASIL, STN. (2015b). *Despesas com o Sistema Integrado de Administração*. Acesso em 11 de Abril de 2019, disponível em Tesouro Nacional Transparente: <https://www.tesourotransparente.gov.br/ckan/dataset/siafi-despesas-siafi-operacional/resource/4e68e2ca-d182-4fad-87d5-7e0873ae5393>

BRASIL, STN. (2015c). *BSPN em foco*. Acesso em 9 de Abril de 2019, disponível em Tesouro Nacional Transparente: <https://www.tesourotransparente.gov.br/historias/bspn?searchterm=BSPN>

BRASIL, TCU. (2014). *Referencial Básico de Governança, 2ª versão*.

BRASIL, TCU. (2014). Relatório de levantamento. Publicação de dados abertos na Administração Pública Federal. Acesso em 23 de Dezembro de 2018, disponível em https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKewjYscadrbnhAhWEJLkGHWuVAfMQFjAAegQIBRAC&url=http%3A%2F%2Fwww.tcu.gov.br%2FConsultas%2FJuris%2FDocs%2Fjudoc%2FAcord%2F20141006%2FAC_2569_38_14_P.doc&usg=AOvVaw0w1EIUh_MhUifmFauyOR

BRASIL, TCU. (2016). O porquê de governança de dados em organizações de controle. *Revista do TCU*.

CAZELLA, C. S., NUNES, M. A., & REATEGUI, E. B. (2015). A Ciência da Opinião: Estado da arte em Sistemas de Recomendação. Acesso em 25 de Fevereiro de 2019, disponível em https://www.academia.edu/26470946/A_Ci%C3%A7%C3%A2ncia_da_Opini%C3%A3o_Estado_da_arte_em_Sistemas_de_Recomenda%C3%A7%C3%A3o

CHIGNARD, S. (2013). *A brief history of Open Data*. Acesso em 15 de Dezembro de 2018, disponível em Paris Innovation Review: <http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>

CORNELL EAS DATA LAKE. (2018). *Cornell EAS Data Lake*. Acesso em 10 de Abril de 2019, disponível em Cornell Earth & Atmospheric Sciences Data Lake: <https://datalake.eas.cornell.edu/>

DAHL, R. A. (1998). *On Democracy*. Yale University Press.

DAVENPORT, T. H. (1998). *Ecologia da INFORMAÇÃO*. Futura.

DAVIS, J. (1998). *Access to and Transmission of Information: Position of the Media*. Maastricht: European Institute of Public Administration.

DGI, D. G. (2007). Defining Data Governance. Acesso em 2 de Fevereiro de 2019, disponível em <http://www.datagovernance.com/defining-data-governance>

DI PIETRO, M. S. (2002). *Direito Administrativo*. São Paulo: Atlas.

DI PIETRO, M. S. (2006). *Direito Administrativo*. São Paulo: Atlas.

DIMAIO, A. (2012). Open Data and The New Divide. *Gartner Blog Network*. Acesso em 9 de Fevereiro de 2019, disponível em https://blogs.gartner.com/andrea_dimaio/2012/05/14/open-data-and-the-new-divide/

DMM. (s.d.). *Data Management Maturity (DMM)SM*. Acesso em 11 de Janeiro de 2019, disponível em CMM Institute: <https://cmminstitute.com/data-management-maturity>

ELASTIC. (s.d.). *The Elastic Stack 7.0*. Acesso em 11 de Março de 2019, disponível em Elastic: <https://www.elastic.co/pt/products/>

FGV. (2017). Índice de Dados Abertos para o Brasil. Acesso em 4 de Janeiro de 2019, disponível em <http://dapp.fgv.br/wp-content/uploads/2017/04/IndiceDadosAbertosBrasil2017.pdf>

FGV. (2018). Índice de Dados Abertos para as Cidades.

FRANÇA. (1789). Déclaration des Droits de l'Homme et du Citoyen de 1789. Acesso em 1 de Janeiro de 2019, disponível em <https://www.conseil-constitutionnel.fr/le-bloc-de-constitutionnalite/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789>

GEERE, D. (13 de Dezembro de 2011). *The history of Creative Commons*. Fonte: Wired: <https://www.wired.co.uk/article/history-of-creative-commons>

GOIS, N. B. (2017). *Open Analysis*. Fonte: Serpro.

HESS, C., & Ostrom, E. (2007). *Understanding Knowledge as a Commons: From Theory to Practice*. (C. Hess, & E. Ostrom, Eds.) London, ENGLAND: The MIT Press.

IESE. (2018). Cities in Motion Index. (P. BERRONE, & J. E. RICART, Eds.) Espanha. doi:10.15581/018.ST-471

IFAC, I. F. (2013). *Good governance in the public sector: consultation draft for an international framework*.

ITU International Telecommunication Union. (2005). ITU Internet Reports 2005: The Internet of Things. Geneva, Suíça. Acesso em 22 de Dezembro de 2018, disponível em <https://www.itu.int/net4/ITU-T/search/api/redirection?dest=http%3A%2F%2Fhandle.itu.int%2F11.1002%2Fwebsites%2F010932e168d0b7e8a74c9c5cdfbc05f0d4c3f2b4&position=9&page=1>

JUNIOR, F. T. (2015). Avaliação da prontidão de dados para a abertura de dados das instituições públicas brasileiras: caso de uma instituição financeira pública brasileira. Brasília, DF. Acesso em 14 de Fevereiro de 2019, disponível em <https://bdtd.ucb.br:8443/jspui/bitstream/123456789/1457/1/Francisco%20Tanajura%20Machado%20Junior.pdf>

KHANNA, R. (29 de Janeiro de 2019). *How to use Elasticsearch, Logstash and Kibana to visualise logs in Python in realtime*. Acesso em 12 de Abril de 2019, disponível em Medium: <https://medium.freecodecamp.org/how-to-use-elasticsearch-logstash-and-kibana-to-visualise-logs-in-python-in-realtime-acaab281c9de>

KOLLOCK, P., & SMITH, M. (1996). *Managing the Virtual Commons: Cooperation and Conflict in Computer Communities*. Los Angeles, Estados Unidos.

KUMAR, S., & PRAKASH, A. (2016). Role of Big Data and Analytics in Smart Cities. *International Journal of Science and Research (IJSR)*, 5.

LAJARA, T. T., & MAÇADA, A. C. (2013). GOVERNANÇA DA INFORMAÇÃO – Estudo de caso em uma empresa de defesa. *EnAnpad*.

LESSIG, L. (2005). *CULTURA LIVRE: Como a Grande Mídia Usa a Tecnologia e a Lei Para Bloquear a Cultura e Controlar a Criatividade*. São Paulo, SP: Editora Francis .

MEIJER, A. (Janeiro de 2015). Government Transparency in Historical Perspective: From the Ancient Regime to Open Data in The Netherlands. doi:10.1080/01900692.2014.934837

MELLO, C. A. (2008). *Curso de Direito Administrativo*. São Paulo: Malheiros.

MONTEIRO, S. D., FERNANDES, R. P., DECARLI, G. C., & TREVISAN, G. L. (2017). Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 22(n.50). doi:10.5007/1518-2924.2017v22n50p161

MOSER, C. (2001). How open is 'open as possible'? Three different approaches to transparency and openness in regulating access to EU documents. Acesso em 13 de Março de 2019, disponível em http://irihs.ihs.ac.at/1389/1/pw_80.pdf

MOVIDESK (Ed.). (2018). *Como estruturar uma cultura voltada ao sucesso do cliente*. Acesso em 10 de Abril de 2019, disponível em Blog do Movidesk: <https://conteudo.movidesk.com/inteligencia-artificial-no-atendimento/>

NAGPAL, N. (2017). Analyzing Role of Big Data and IoT in Smart Cities. India. Acesso em 14 de Março de 2019, disponível em https://ijaems.com/upload_images/issue_files/29%20IJAEMS-MAY-2017-45-Analyzing%20Role%20of%20Big%20Data%20and%20IoT%20in%20Smart%20Cities.pdf

NATIONAL RESEARCH COUNCIL. (1995). *On the Full and Open Exchange of Scientific Data*. Washington, DC: The National Academies Press. doi:10.17226/18769

NEW ZELAND GOVERNMENT. (15 de Setembro de 2018). *Guidance and Resources*. Fonte: Data and Information Governance and Maturity: <https://www.ict.govt.nz/guidance-and-resources/architecture/data-and-information-governance-and-maturity/>

OLIVER, R. W. (2004). *What is Transparency?* McGraw-Hill. doi:10.1036/0071457224

OPEN GOVERNMENT DATA. (8 de Dezembro de 2007). *The annotated 8 Principles of Open Government Data*. Fonte: opengovdata.org: <https://opengovdata.org>

OPEN GOVERNMENT PARTNERSHIP. (2011). *Declaração de Governo Aberto*. Operação Serenata de Amor. (2018). *Candidatos chegam a evoluir patrimônio em 49.000% desde 2014*. Acesso em 14 de Janeiro de 2019, disponível em Medium: <https://medium.com/serenata/candidatos-chegam-a-evoluir-patrim%C3%B4nio-em-49-000-desde-2014-55907b3ff423>

PRAT, A. (2005). The More Closely we are Watched, the Better we Behave? Acesso em 24 de Janeiro de 2019, disponível em <http://www.columbia.edu/~ap3116/papers/behave.pdf>

REINALDO, D. R. (1997). A PUBLICIDADE DOS ATOS E DECISÕES ADMINISTRATIVOS. *Revista do Tribunal Regional Federal 1ª Região*, 4. Fonte: https://bdjur.stj.jus.br/jspui/bitstream/2011/21676/publicidade_atos_decisoes_administrativos.pdf

SERENATA. (s.d.). *Serenata de amor: Inteligência Artificial para controle social da Administração Pública*. Acesso em 28 de Novembro de 2018, disponível em Operação Serenata de amor: <https://serenata.ai/>

SINGH, V., SRIVASTAVA, I., & JOHRI, V. (2014). Big Data and the Opportunities and Challenges for Government Agencies. *International Journal of Computer Science and Information Technologies*, Vol. 5.

THE ROYAL IRISH ACADEMY. (2002). *The Cathach / The Psalter of St Columba*. Fonte: THE ROYAL IRISH ACADEMY IS IRELAND'S LEADING BODY OF EXPERTS IN THE SCIENCES AND HUMANITIES.

U.S. GOVERNMENT. (14 de Dezembro de 2007). S.2488 - *OPEN Government Act of 2007*. Fonte: Congresso.gov: <https://www.congress.gov/bill/110th-congress/senate-bill/2488/text>

UNESCO. (2015). Concepts of Openness and Open Access. Acesso em 25 de Janeiro de 2019, disponível em <http://www.unesco.org/open-access/terms-use-ccbysa-en>

UNITED NATION. (2018). UNITED NATIONS E-GOVERNMENT SURVEY 2018: GEARING E-GOVERNMENT TO SUPPORT TRANSFORMATION TOWARDS SUSTAINABLE AND RESILIENT SOCIETIES. Nova York. Acesso em 14 de Dezembro de 2018, disponível em publicadministration.un.org

VAZ, J. C., RIBEIRO, M. M., & MATHEUS, R. (2010). Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil. *Democracia e Interfaces Digitais para a Participação Pública*, 9(edição especial (2010)). Fonte: <https://portalseer.ufba.br/index.php/ppgau/issue/view/556>

WIKIPEDIA. (2013). *Categoria:Licenças de direitos autorais*. Fonte: Wikipédia: A enciclopédia livre: https://pt.wikipedia.org/wiki/Categoria:Lic%C3%A7%C3%A3o_de_Dados_Pessoais

WIKIPEDIA. (2018). Lei Geral de Proteção de Dados Pessoais. Acesso em 2 de Fevereiro de 2019, disponível em https://pt.wikipedia.org/wiki/Lei_Geral_de_Prote%C3%A7%C3%A3o_de_Dados_Pessoais

WIKIPEDIA. (2019). *Copyright Act of 1909*. Acesso em 3 de Janeiro de 2019, disponível em Wikipedia: The Free Encyclopedia: https://en.wikipedia.org/wiki/Copyright_Act_of_1909

WIKIPEDIA, PETABYTE. (s.d.). Petabyte. Acesso em 7 de Janeiro de 2019, disponível em <https://en.wikipedia.org/wiki/Petabyte>

WORLD BANK. (2013). The International Bank for Reconstruction and Development. Worldwide Governance Indicators. Acesso em 1 de Outubro de 2013, disponível em <http://info.worldbank.org/governance/wgi/index.aspx#faq>

ZANINI, L. E. (2010). O Estatuto da Rainha Ana: estudos em comemoração dos 300 anos da primeira lei de copyright. Porto Alegre, RS, Brasil. Acesso em 4 de Janeiro de 2019, disponível em http://www.revistadoutrina.trf4.jus.br/artigos/edicao039/leonardo_zanini.html

