

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

Flávio Henrique Martins

**Governança de Dados na Administração Pública Federal Brasileira:
Integração e Qualidade de Dados no Apoio a Tomada de Decisão e Elaboração
de Políticas Públicas**

Brasília
2019

FLÁVIO HENRIQUE MARTINS

**Governança de Dados na Administração Pública Federal Brasileira:
Integração e Qualidade de Dados no Apoio a Tomada de Decisão e Elaboração
de Políticas Públicas**

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: D.ra. Mirella Moura Moro

Brasília
2019

© Flávio Henrique Martins
Todos os direitos reservados

Ficha catalográfica elaborada pela Biblioteca do ICEx – UFMG

Martins, Flávio Henrique

M386g Governança de dados na administração pública federal brasileira:
integração e qualidade de dados no apoio a tomada de decisão e elaboração
de políticas públicas 2019.
xi, 47 f., il.

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

Orientador: D.ra Mirella Moura Moro

1. Computação – Monografias. 2. Governança de dados. 3.
Recuperação da informação. 4. Integração de dados. 5. Qualidade de dados.
6. Extração, transformação e carga de dados – ETL. 7. Suporte a tomada de
decisão.

I. Orientador: Moro, Mirella Moura. II. Título: Governança de Dados na
Administração Pública Federal Brasileira: Integração e Qualidade de Dados
no Apoio a Tomada de Decisão e Elaboração de Políticas Públicas.

CDU 519.6*

FOLHA DE APROVAÇÃO




UNIVERSIDADE FEDERAL DE MINAS GERAIS

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

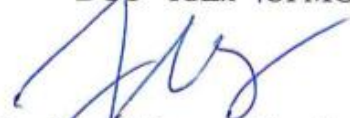
Governança de Dados na Administração Pública Federal Brasileira: Integração e
Qualidade dos Dados no Apoio a Tomada de Decisão e Elaboração de Políticas
Públicas

FLÁVIO HENRIQUE MARTINS

Monografia apresentada aos Senhores:


Prof. Mirella Moura Moro
Orientadora
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

O presente trabalho de conclusão de curso visa demonstrar conceitos de integração e qualidade de dados para elaboração de informação de alto valor na geração de conhecimento para a tomada de decisão no âmbito da Administração Pública Federal. Cada órgão pode possuir sua própria infraestrutura para análise de dados, porém existem alguns métodos gerais para a melhoria da qualidade da informação demonstrada ao usuário final.

Palavras chaves: Integração de dados, qualidade de dados, banco de dados, suporte a tomada de decisão, extração, transformação e carga de dados – ETL.

ABSTRACT

The present work of conclusion of course aims to demonstrate concepts of integration and quality of data to elaborate information of high value in the generation of knowledge for the decision making in the scope of Federal Public Administration.

Each organization may have its own infrastructure for data analysis, but there are some general methods for improving the quality of information demonstrated to the end user.

Keywords: Data integration, data quality, database, decision making support, extract, transform and load – ETL.

LISTA DE FIGURAS

Figura 1 - Funções da Gestão de Dados DMBOK	21
Figura 2 - Funções da Gestão de Dados DMBOK v2.....	22
Figura 3 - Etapas da construção do conhecimento	28
Figura 4 - Tabelas primárias PAA	32
Figura 5 - ETL de integração das tabelas do PAA	33
Figura 6 - ETL de integração e qualidade	33
Figura 7 - Visualização de dados com Data Discovery	34
Figura 8 - Categorias de agricultores sem qualidade	37
Figura 9 - Categoria de agricultores após qualidade	38
Figura 10 - Categorias únicas de agricultores	39
Figura 11 - Sexo agricultores sem qualidade	39
Figura 12 - Sexo agricultores após qualidade	40
Figura 13 - Sexo único agricultores	40

LISTA DE SIGLAS

APF – Administração Pública Federal

BI – *Business Intelligence*

DAMA - *Data Management Association*

DAMA-DMBOK – *DAMA - Guide to the Data Management Body of Knowledge*

DW – *Data Warehouse*

EDW – *Enterprise Data Warehouse*

GD – Governança de Dados

IBGC – Instituto Brasileiro de Governança Corporativa

IFAC – *International Federation of Accountants*

TCU – Tribunal de Contas da União

TI – Tecnologia de Informação

SUMÁRIO

1. INTRODUÇÃO	11
2. OBJETIVOS	14
2.1. Objetivo Geral	14
2.2. Objetivo Específico	14
3. REFERENCIAL TEÓRICO	15
3.1. Governança	15
3.2. Governança Corporativa	16
3.3. Governança Pública	17
3.4. Governança de Dados	19
3.4.1. Dado	24
3.4.2. Informação	25
3.4.3. Conhecimento	26
3.4.4. Inteligência	27
3.4.5. Integração de Dados	28
3.4.6. Qualidade de Dados	35
3.4.7. Banco de Dados Corporativo	41
4. INICIATIVAS NA ADMINISTRAÇÃO PÚBLICA FEDERAL	42
5. CONCLUSÃO	45
6. REFERÊNCIAS BIBLIOGRÁFICAS	47

1. INTRODUÇÃO

Atualmente, com a evolução tecnológica e a grande capacidade de geração e armazenagem de conteúdo, faz-se necessário métodos e ferramentas para análise dos dados coletados no suporte a alta gestão. MOSLEY *et al.* (2012, p. 9), em sua publicação DAMA-DMBOK, que será apresentada em um momento oportuno neste trabalho, afirma que os “dados e informações são a força motriz da economia no século 21. Na era da informação, os dados são reconhecidos como vitais para uma organização”.

As bases de dados, smartphones, Internet das Coisas, Big Data e a própria Internet são grandes fontes a serem utilizadas na elaboração de políticas públicas e tomada de decisão. A necessidade da informação digital de qualidade e imediata é uma tendência inevitável e as instituições públicas precisam se adaptar a essa nova realidade. Segundo MOSLEY *et al.* (2012, p. 9), “nenhuma organização pode ser eficaz sem dados de alta qualidade”.

O cenário atual com grande volume de dados e informações se torna cada vez mais complexo para os gestores que tomam decisões estratégicas em seu cotidiano. Muitas vezes as informações estão desatualizadas, imprecisas ou até incorretas, dificultando assim, qualquer tomada de decisão e elaboração de políticas públicas assertivas.

Neste momento, para que as instituições governamentais alcancem seus objetivos estratégicos, seus dados precisam ser integrados entre os órgãos federais e tratados para a melhoria e enriquecimento da qualidade na disponibilização da informação, e desta forma, transformá-los em ativos de alto valor para cada instituição e para a sociedade.

Segundo MOSLEY *et al.* (2012, p. 9), “agências governamentais, instituições educacionais e organizações sem fins lucrativos também necessitam de dados de alta qualidade para orientar suas atividades operacionais, táticas e estratégicas”.

Nesse sentido, os órgãos da Administração Pública Federal - APF, tem por objetivo alavancarem seus projetos, formular políticas públicas que atendam a necessidade da sociedade e melhorar a qualidade da tomada de decisão. Ou seja,

estarem prontamente disponíveis com informações que irão embasar as decisões tomadas pela alta gestão de cada instituição.

Em geral, cada órgão investe individualmente em sua própria infraestrutura de hardware e em sistemas de apoio a tomada de decisão, baseando-se somente na utilização das ferramentas analíticas e não na construção de um repositório analítico único contendo dados tratados para responder às perguntas de negócio de forma confiável.

Nesse contexto, é necessário entender que os dados dificilmente atingirão um estado perfeito, mas que algumas ações podem ser tomadas para que se possa alcançar o maior nível de qualidade possível. É essencial que as instituições públicas brasileiras possuam uma iniciativa ou uma área relacionada à Governança de Dados (GD) para o planejamento, monitoramento, controle e execução das atividades relacionadas aos ativos de dados e informações.

É fato que os dados pertencentes a uma organização podem ser mais valiosos do que ela própria. No caso das instituições governamentais, quando estes são tratados e enriquecidos com dados de outras fontes, são o diferencial para proporcionar o suporte confiável à alta gestão e na elaboração de políticas públicas eficientes e eficazes. Conforme MOSLEY *et al.* (2012, p. 9), na publicação DAMA-DMBOK, afirma que “os dados, e as informações criadas a partir dos dados, atualmente são amplamente reconhecidos como ativos da organização”.

Nesse sentido, a GD permeia áreas como gestão dos processos, de ações políticas, dos patrocinadores de projetos de dados, de pessoas envolvidas, de tecnologias utilizadas, de cultura organizacional, de segurança da informação, de arquitetura dos dados e outras áreas de conhecimento que implicam no uso e manipulação dos dados e informações.

Com esse entendimento, o Governo Federal, através de seus Ministérios, possui uma grande massa de dados sobre a população brasileira, programas sociais, territórios, identificadores cadastrais, segurança pública e diversos outros temas, onde existe pouca ou nenhuma integração entre essas bases de dados de suas entidades.

Sendo assim, é possível notar que além da integração entre repositórios de dados governamentais, a qualidade do dado contido nestas é muito importante.

Decisões tomadas baseadas em informações de má qualidade podem levar a decisões errada e com graves consequências. Da mesma forma, a elaboração de uma política pública com informações distorcidas, podem não ser efetivas e gerando gastos excessivos e desnecessários.

No que tange este trabalho, ele não terá a finalidade de descrever todas as áreas da GD e nem explicar todas as etapas e papéis decisórios para a implantação de um programa de governança voltado para os dados corporativos que não estiverem alinhados com o objetivo proposto.

2. OBJETIVOS

2.1. Objetivo Geral

O presente trabalho visa orientar as áreas dos órgãos Administração Pública Federal (APF) que promovem a disseminação do conhecimento, através da análise de dados e informações, com princípios e fundamentos utilizados na governança de dados para promover a integração e qualidade de dados no suporte a tomada de decisão e elaboração de políticas públicas.

2.2. Objetivo Específico

Abordar à aplicação da Governança de Dados nos temas de integração em bases de dados de instituições públicas federais, a melhoria da qualidade da informação gerada e seus metadados para compor uma base de dados analítica no contexto da Administração Pública Federal.

3. REFERENCIAL TEÓRICO

Para que seja entendida a finalidade da Governança de Dados e as áreas de conhecimento de integração e qualidade de dados, será necessário compreender conceitos importantes na disponibilização de informação de alto valor para o suporte a tomada de decisão e elaboração de políticas públicas.

Pode-se afirmar que existem algumas vertentes de Governança, na perspectiva abordada no trabalho, faz-se necessário conceituar a Governança, Governança Corporativa, Governança Pública e a Governança de Dados que serão discutidas em um tópico específico.

3.1. Governança

De forma geral, a **Governança** foi conceituada pelo Banco Mundial em sua publicação *Governance and Development* (1992, p. 3), como sendo “o exercício da autoridade, controle, administração, poder de governo”. De forma mais específica para a finalidade da instituição, foi conceituada como sendo “a maneira pela qual o poder é exercido na administração dos recursos sociais e econômicos de um país visando o desenvolvimento”.

Nesse sentido, percebe-se que a governança está alinhada com todos os objetivos estratégicos de uma organização, definindo como estas serão direcionadas, monitoradas e fiscalizadas, no alinhamento institucional de como seus dados serão utilizados no apoio às diretrizes corporativas estratégicas, bem como na formulação de políticas públicas.

Para o Tribunal de Contas da União - TCU (2014, p.11), o conceito de Governança está associado ao momento em que o proprietário deixa de gerir sua organização e delega à administração de seus recursos a um terceiro. Em alguns casos, a falta de proporção entre informação, poder e autoridade pode levar a um conflito de interesse.

A Governança em sentido amplo é um termo que pode ser utilizado em diversos contextos. As definições mais utilizadas são a Corporativa, a Pública e a mais recente, de Dados.

Assim, para este trabalho trataremos a Governança como o sistema em que as organizações são administradas, controladas e avaliadas, no alcance de seus objetivos estratégicos através da utilização de informação de alto valor.

Corroborando com esse conceito, o IFAC (2013), nos diz que “a governança compreende os compromissos estabelecidos na garantia de que os resultados pretendidos pelas partes interessadas sejam definidos e alcançados”.

3.2. Governança Corporativa

Com essa finalidade, surge o conceito de **Governança Corporativa** que é definida pelo IFAC, 2013 como sendo “o sistema pelo qual as organizações são dirigidas e controladas sob três princípios fundamentais: transparência, integridade e prestação de contas”. Ainda segundo o IFAC, 2013, “esses princípios são relevantes para entidades do setor público, como para entidades do setor privado”.

Para o TCU (2014, p. 12), “no Brasil, o crescente interesse pelo tema não é diferente. Tanto no setor privado quanto no público, existem iniciativas de melhoria da governança, as quais se relacionam e se complementam”.

No caso do Brasil, observou-se a partir dos anos 2000 um aumento da necessidade que foi proposta no segmento privado para a melhoria da gestão de investimento e gestão de empresas nacionais. Assim, surge o Instituto Brasileiro de Governança Corporativa (IBGC), que tem como propósito contribuir para o desempenho sustentável das organizações e influenciando os agentes de nossa sociedade no sentido de maior transparência, justiça e responsabilidade.

Pode-se afirmar que a Governança Corporativa “é o sistema pelo qual as empresas e demais organizações são dirigidas, monitoradas e incentivadas”, seguindo um código de boas práticas que seriam “princípios básicos, alinhando interesses com a finalidade de preservar e otimizar o valor econômico de longo prazo

da organização, facilitando seu acesso a recursos e contribuindo para a qualidade da gestão da organização”¹.

O Código de Governança Corporativa do IBGC (2015) traz como princípios básicos:

Transparência: Consiste no desejo de disponibilizar para as partes interessadas as informações que sejam de seu interesse e não apenas aquelas impostas por disposições de leis ou regulamentos. Para o IFAC, 2013, a transparência visa “garantir que as partes interessadas possam ter confiança nos processos de tomada de decisão e nas ações das entidades do setor público, na gestão de suas atividades e nos indivíduos dentro delas”.

Equidade: Caracteriza-se pelo tratamento justo e isonômico de todos os sócios e demais partes interessadas (*stakeholders*), levando em consideração seus direitos, deveres, necessidades, interesses e expectativas.

Prestação de contas: Os agentes de governança devem prestar contas de sua atuação de modo claro, conciso, compreensível e tempestivo, assumindo integralmente as consequências de seus atos e omissões e atuando com diligência e responsabilidade no âmbito dos seus papéis.

Responsabilidade Corporativa: Os agentes de governança devem zelar pela viabilidade econômico-financeira das organizações, reduzir as externalidades negativas de seus negócios e suas operações e aumentar as positivas.

Mesmo o IBGC tendo desenvolvido esse Código para organizações empresariais, tornou-se um ponto de direção para cooperativas, empresas públicas e órgãos governamentais (BRASIL, TCU, 2014, p. 16).

3.3. Governança Pública

A **Governança Pública** é percebida como um sistema que estabelece a harmonia do poder entre sujeitos (Estado, Administração e cidadão) com o objetivo de possibilitar que o interesse público predomine sobre desejos particulares (MATIAS-PEREIRA, 2010).

Desta maneira, o TCU, órgão que exerce fiscalização da União e entidades da Administração Pública, complementa o conceito discorrendo que seria um sistema que envolve vários agentes, entre sociedade, alta administração, servidores ou colaboradores e os órgãos controladores. O objetivo é de proporcionar uma confiança

¹ IBGC. **Governança Corporativa**. Disponível em: <<<https://www.ibgc.org.br/governanca/governanca-corporativa>>>. Acesso em: 02/10/2018

para a sociedade das ações estatais, utilizando-se da supremacia do interesse público (BRASIL, TCU, 2014, p. 21).

Seu ponto central é a relação entre o Estado e o cidadão, trazendo-o a participar do processo constitutivo das políticas públicas, ou seja, como um agente ativo na administração pública. Nesse modelo, não se pretende somente focar nos resultados das ações políticas para a sociedade, mas também na maneira como esses são atingidos.

Para o IFAC (2013), a função da governança no setor público é a garantia que as ações das instituições públicas estejam alinhadas com o interesse da sociedade.

Nesse sentido, o documento Referencial Básico de Governança, publicado pelo TCU (2014, p. 9), elucida a definição de Governança Aplicada ao Setor Público como sendo um conjunto de instrumentos de liderança, estratégia e controle que são utilizados na avaliação, direcionamento e monitoramento da gestão, para o desenvolvimento de políticas públicas e serviços que atendam a população.

Seguindo essa linha de princípios citados, a Governança no Setor Público é mais complexa por precisar harmonizar a distribuição de poder entre a sociedade, a fim de atingir objetivos sociais, econômicos e políticos em um escopo muito maior do que o setor privado, o qual tem o foco em como as organizações são governadas e administradas.

Percebe-se que não são apenas mecanismos de análise, direção e monitoramento, mas também envolvem as relações entre estruturas, processos e costumes, que delimitam como os agentes sociais serão atendidos, como serão tomadas as decisões e como serão executados os poderes e as atribuições (GRAHN; AMOS; PLUMPTRE, 2003). Observa-se que o Setor Público deve ser capaz de gerir seus sistemas políticos e administrativos, para que sejam solucionadas todas as demandas possíveis.

As palavras do Dr. Jorge Gerdau Johannpeter, à época coordenador da Câmara de Gestão e Planejamento do Governo Federal, ressaltam a importância de se ter um Referencial como o que foi proposto pelo TCU e a sua aplicabilidade em vários setores e organizações:

[...] o conteúdo, que é realmente de grande dimensão e profundidade, e, embora o material seja voltado para o setor público, os principais conceitos mencionados também se aplicam ao setor privado.

O que também me encantou é que venho reforçando que o conceito de Governança do setor público está muito pouco presente em nosso país, e essa é uma das partes mais importantes do processo de gestão (BRASIL, TCU, 2014, p. 11).

Este aporte traz a orientação de práticas adequadas para a Governança Pública, embasando-se em ações desenvolvidas no âmbito do próprio TCU. Desta forma, outras entidades podem utilizar-se para aplicar dentro da realidade administrativa e gerencial.

3.4. Governança de Dados

No contexto da APF, que é o objetivo deste trabalho, o Ministério do Planejamento, em sua Portaria nº 58/2016, que dispõe sobre procedimentos complementares para o compartilhamento de bases de dados oficiais entre órgãos e entidades da administração pública federal direta e indireta e as demais entidades controladas direta ou indiretamente pela União, em seu Art. 2º, Inciso VI, define a GD como:

Governança de dados: conjunto de políticas, processos, pessoas e tecnologias que visam a estruturar e administrar os ativos de informação, com o objetivo de aprimorar a eficiência dos processos de gestão e da qualidade dos dados, a fim de promover eficiência operacional, bem como garantir a confiabilidade das informações que suportam a tomada de decisão. (BRASIL, 2016a)

Como literatura auxiliar, sendo reconhecida internacionalmente e amplamente utilizada, o guia *DAMA - Guide to the Data Management Body of Knowledge* (DAMA-DMBOK) que traz um conjunto de melhores práticas para GD. A DAMA é uma organização internacional sem fins lucrativos, composta por profissionais da área de gestão de dados com a missão de promover a educação e desenvolvimento no gerenciamento de dados e informações no suporte às estratégias de negócio.

Nesse sentido, MOSLEY *et al.* (2012, p. 47), define que:

Governança de Dados é o exercício de autoridade e controle (planejamento, monitoramento e execução) sobre a gestão de ativos de dados. A função de governança de dados guia como todas as outras funções de gestão de dados são realizadas.

Para LADLEY (2012, p. 11), em sua publicação *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*, define:

Governança de dados é a organização e implementação de políticas, procedimentos, estrutura, papéis e responsabilidades que delinham e reforçam regras de comprometimento, direitos decisórios e prestação de contas para garantir o gerenciamento apropriado dos ativos de dados.

Em relação aos objetivos da GD, MOSLEY *et al.* (2012, p. 47) elenca os seguintes pontos:

1. Definir, aprovar, e comunicar estratégias de dados, políticas, padrões, arquitetura, procedimentos e métricas;
2. Acompanhar e forçar o cumprimento de regulatórios e conformidades com políticas de dados, padrões, arquitetura e procedimentos;
3. Patrocinar, acompanhar e supervisionar as entregas de projetos e serviços de gestão de dados;
4. Gerenciar e resolver questões relacionadas a dados; e
5. Entender e promover o valor dos ativos de dados.

A GD não pode ser vista apenas como uma responsabilidade apenas da área de Tecnologia de Informação (TI) por estar altamente envolvida com tecnologia, ela deve ser compartilhada com as áreas de negócio da organização, com a área de governança corporativa para o devido alinhamento com o plano estratégico e também, melhorar qualidade da tomada de decisão com os objetivos institucionais.

DUMONT *et al.* (2006, p. 22) nos diz que:

Os sistemas de informação, geralmente utilizados pelas administrações, foram concebidos do ponto de vista de quem gera a informação e distantes do foco da necessidade de quem precisa da informação para uma tomada de decisão assertiva.

Nesse mesmo sentido, MOSLEY *et al.* (2012, p. 13) nos diz que a responsabilidade sobre a gestão dos dados é compartilhada:

A gestão de dados é uma responsabilidade compartilhada entre os profissionais de gestão de dados da TI e os gestores de dados das áreas de negócio (*Data Stewards*) que representam os interesses coletivos dos produtos de dados e dos consumidores de informação.

Essa responsabilidade é muito importante para que exista o alinhamento correto entre as áreas requisitantes da informação e a área de TI. De forma geral, a TI guarda os dados e os deixa disponíveis para o consumo das áreas de negócio, porém quem define qual dado é relevante e importante é a própria área que se utiliza deles na construção da informação e geração do conhecimento.

“Enquanto gestores de dados cuidam dos ativos de dados, os profissionais de gestão de dados desempenham funções técnicas para proteger e permitir a utilização eficaz dos ativos de dados da organização” MOSLEY *et al.* (2012, p. 13).

O guia DAMA-DMBOK, em sua primeira versão em 2012, nos traz 10 áreas de conhecimento, sendo a Governança de Dados o guia todas as outras. Essas áreas também podem ser chamadas de funções da gestão de dados e podem ser visualizadas conforme figura 1.

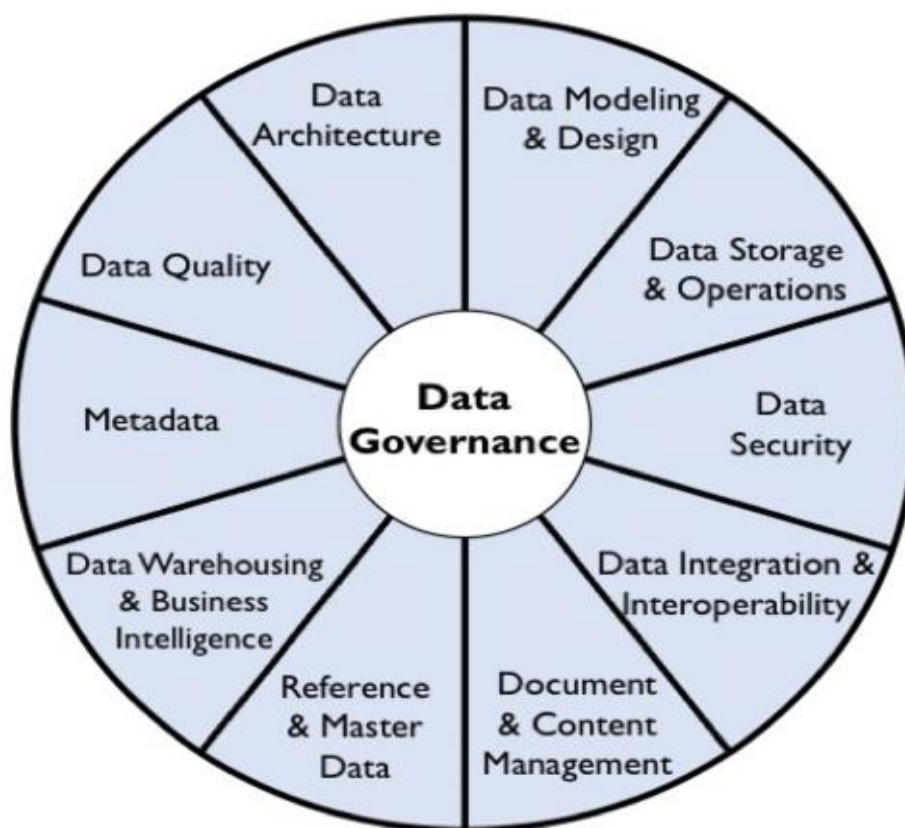
Figura 1 - Funções da Gestão de Dados DMBOK



Fonte: MOSLEY *et al.* (2012)

Já em sua segunda versão, o guia DAMA-DMBOK publicado no ano de 2017, possui 11 áreas de conhecimento, sendo acrescida da área de integração e interoperabilidade de dados, que serão demonstradas na figura 2.

Figura 2 - Funções da Gestão de Dados DMBOK v2



Copyright© 2017 DAMA International

Fonte: DAMA-DMBOK v2 (2017)

As áreas de conhecimento descritas no DAMA-DMBOK, tanto em sua primeira versão como na segunda, se fazem necessárias para o desenvolvimento de um programa de GD, pois este irá alinhar as necessidades e estratégias de negócio, com a governança corporativa e a área de TI.

Nesse contexto, a área de TI deve fornecer toda a tecnologia no que tange a coleta, processamento, transformação, uso e reuso, disseminação de conteúdo e atualização de qualquer tipo de dado e informação sob a sua guarda e gestão.

É importante ressaltar que o programa de implantação da GD deve ter um apoio *Top-Down*, ou seja, a alta gestão deve patrocinar essa iniciativa para o devido alinhamento e implantação com as diversas áreas envolvidas dentro do órgão. É interessante que o ponto de partida para a implantação da GD seja primeiramente com as pessoas envolvidas, com a política que será elaborada e a mudança na cultura

organizacional e depois avançar para os processos de governança, administração de dados e finalmente para a tecnologia.

Apesar da mudança organizacional não estar no estágio primário da implantação de um programa de GD, ela é muito necessária. Para MOSLEY *et al.* (2012, p. 62), “uma mudança organizacional significativa é muitas vezes necessária para implementar uma gestão mais eficaz dos dados”.

Todas essas conceituações das diferentes vertentes de Governança e da Governança de Dados são necessárias para entender que os dados de uma instituição são de extrema importância e precisam ser administrados e controlados sob os cuidados de uma área específica. Estes dados, quando bem administrados, servem para gerar informação de alto valor e embasar as ações da alta gestão na tomada de decisão.

Diferentes necessidades de governança de dados exigem atividades específicas de controle, bem como, requisitos específicos. Praticar a GD é encontrar a quantidade e o nível correto de controle. Adotar uma abordagem iterativa mitigará os riscos de implementação e ajudará uma organização a se concentrar na elaboração de políticas para sua necessidade, sendo eficaz e bem-sucedida no gerenciamento de seus ativos mais importantes.

Conforme as organizações federais conhecem seus dados, melhores são identificados os requisitos da informação desejada. Segundo CHOO (2013, p. 20), “à medida que o papel da informação dentro das empresas ficar mais claro, também teremos maior clareza sobre as condições, padrões e normas de utilização capazes de tornar a informação significativa e valiosa”.

A informação de alto valor, necessária para as ações estratégicas dos órgãos e entidades federais, parte de uma necessidade informacional conhecida ou não. Conhecer o seu próprio dado também leva a uma melhor avaliação da qualidade e utilização da informação.

Para o TCU (2014, p. 32), a governança também se preocupa com a qualidade do processo decisório e sua efetividade. Com esse raciocínio, ele faz os seguintes questionamentos: Por quem e por que essas decisões foram tomadas? Os objetivos foram alcançados?

Segundo MOSLEY *et al.* (2012, p. 52):

Algumas decisões podem ser tomadas sem riscos para gestores individuais. Mas a necessidade de tomada de decisão compartilhada, e o controle de risco conduzem a maioria das organizações a uma forma representativa de governança de dados, de modo que todos os interessados e constituintes possam ser ouvidos.

Nesse sentido, é possível perceber que um dos papéis da GD dentro de uma instituição é proporcionar um ambiente em que seus servidores e gestores possam formular a melhor estratégia para conquistar os objetivos estratégicos da organização em cada tomada de decisão, onde cada responsável pelos dados de negócio possa expor as suas necessidades de informação.

MOSLEY *et al.* (2012, p. 56) afirma que:

Normalmente, uma estratégia de dados é um programa estratégico de gestão de dados para manter e melhorar a qualidade dos dados, integridade, segurança e acesso. No entanto, uma estratégia de dados pode incluir também planos de negócio para utilizar as informações para a vantagem competitiva e suportar os objetivos da organização. A estratégia de dados deve ser proveniente de uma compreensão das necessidades inerentes às estratégias de negócio. Esses dados precisam conduzir a estratégia de dados.

A estratégia de dados e informações não pode ser confundida com a arquitetura de dados adotada. Nesse caso, a estratégia pode influenciar a arquitetura, que por sua vez apoia a estratégia e, também orientam outras decisões. Em muitos casos, essa estratégia é definida por um comitê gestor de dados e informações.

Nesse contexto da GD, é preciso diferenciar o que são dados, informações, conhecimento e finalmente a inteligência adquirida. Conforme DUMONT *et al.* (2006, p. 32), “não se deve confundir dado com informação, assim como informação com conhecimento, tampouco conhecimento com inteligência”.

3.4.1. Dado

O dado pode ser conceituado como uma série de fatos que não foram contextualizados, integrados ou interpretados e que representam apenas uma parte isolada de acontecimentos e situações. Estes são a matéria bruta para a construção da informação.

Para MOSLEY *et al.* (2012, p. 11), “dado é a fundação da informação, do conhecimento e, finalmente, da sabedoria e da ação bem informada”. Ou seja, um dado é o pilar na construção da informação de alta qualidade que fundamenta o conhecimento sobre esta, e por fim, embasa a inteligência nas ações da alta gestão para tomada de decisão.

As bases de dados legadas e os dados semi e não estruturados podem conter muita inconsistência e erros de cadastro, porém é necessário um trabalho para a melhoria destes. Atualmente a necessidade por informação de alto valor é muito grande e esperar uma ótima qualidade nas ações da APF com dados ruins não é possível.

Conforme já dito anteriormente, os dados são a matéria prima no desenvolvimento da informação para tomada de decisão. Deve-se dar a devida importância às fontes primárias no que tange a qualidade e integração dos dados na disponibilização da informação de alto valor à alta gestão e também para a sociedade.

3.4.2. Informação

MOSLEY *et al.* (2012, p. 10) nos diz que a “informação são dados em contexto. Sem contexto, o dado não tem significado; nós criamos informações significativas ao interpretar o contexto em torno do dado”. Nesse sentido, a informação também contribui para a construção do conhecimento.

Sendo assim, a informação, após um processo de transformação dos dados, deve responder às necessidades de cada instituição, sejam elas para dar sentido a perguntas, solucionar problemas, auxiliar a tomada de decisão, apoiar em uma negociação ou dar sentido à algum acontecimento.

Segundo CHOO (2013, p. 20):

Os sistemas e serviços de informação não se limitam a responder perguntas, mas são concebidos como atividades que agregam valor à informação, de modo a ajudar os usuários a tomar melhores decisões e ter uma melhor percepção das situações.

Através de informações consolidadas e de qualidade, é possível tomar decisões mais alinhadas com a necessidade da organização, bem como melhorar o

direcionamento das políticas públicas dos órgãos integrantes da APF no atendimento às demandas da sociedade.

DUMONT *et al.* (2006, p. 34) explica que a “informação é a mensagem transmitida da interpretação de um ou mais dados, ou, de forma mais simples, de um dado ou de um conjunto de dados, com valor agregado implícito objetivando uma comunicação”.

3.4.3. Conhecimento

A construção do conhecimento vai além da informação, pois ele dá significado a uma ocorrência ou situação, bem como o direcionamento para sua devida aplicação. Conforme CHOO (2013, p. 20), “na organização do conhecimento, os sistemas e serviços de informação vão além do que as pessoas querem saber, para saber por que e como elas usam a informação”.

O uso estratégico da informação é percebido através do entendimento, da criação, da estruturação e do processamento da informação, a fim de produzir novos conhecimentos através do aprendizado.

Para MOSLEY *et al.* (2012, p. 10), o “conhecimento é a informação em perspectiva, integrado em um ponto de vista com base no reconhecimento e interpretação de padrões, tais como tendências formadas com outras informações e experiências”.

Assim, o conhecimento é capaz de apoiar a construção de novas ideias na ampliação do conhecimento humano. CHOO (2013, p. 53) afirma que “da construção do conhecimento resultam novos conhecimentos que levam a inovações, novos produtos e novas capacidades organizacionais”.

Nesse sentido, podemos interpretar o conhecimento como sendo a informação trabalhada para a geração do conhecimento através da experiência humana. Em outras palavras, é a capacidade de analisar e interpretar um conjunto de informações levando em consideração o contexto e o ambiente em que está inserido.

MOSLEY *et al.* (2012, p. 11) nos diz que:

O conhecimento pode ser explícito, o que uma organização ou comunidade aceita como verdadeiro, ou tácito, que está dentro da cabeça dos indivíduos. Nós ganhamos conhecimento quando compreendemos a informação.

CHOO (2013, p. 188) corrobora afirmando que:

O conhecimento tácito é o conhecimento implícito usado pelos membros da organização para realizar seu trabalho e dar sentido a seu mundo. O conhecimento tácito é difícil de verbalizar porque é expresso por habilidades baseadas em ação e não pode se reduzir a regras e receitas.

O conhecimento explícito é aquele que pode ser expresso formalmente com a utilização de um sistema de símbolos, portanto ser facilmente comunicado e difundido. O conhecimento explícito pode se basear em objetos ou regras.

Assim, o conhecimento está relacionado com a informação utilizada por uma pessoa, onde ela irá agregar a informação com seu conhecimento do contexto inserido, e por fim transformá-lo em inteligência.

3.4.4. Inteligência

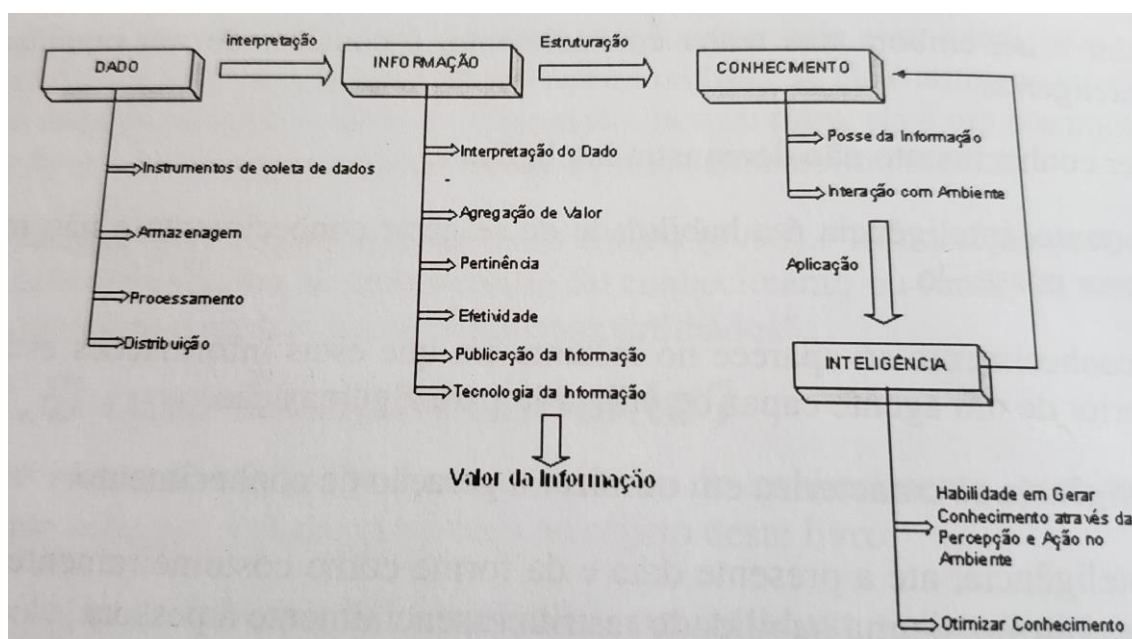
DUMONT *et al.* (2006, p. 34) exemplifica a inteligência como sendo a “habilidade de um agente em gerar conhecimento através da percepção e da ação no ambiente em que vive”, ou seja, através da inteligência o agente é capaz de criar conhecimento por meio de suas próprias atividades.

A inteligência é percebida por melhorar a capacidade de gerar conhecimento. Ela também é uma habilidade característica de seres humanos. Uma das estratégias voltadas para a aplicação da inteligência no ambiente corporativo é priorizar a geração do conhecimento e disseminação da informação de alto valor para toda a instituição.

As etapas para geração de conhecimento sempre se iniciam com os dados. Segundo DUMONT *et al.* (2006, p. 35) “o processo para a geração de conhecimento e, conseqüentemente, do desenvolvimento da inteligência, compreende um fluxo que começa com a geração do dado”.

Todas as etapas e métodos utilizados na construção do conhecimento, conforme DUMONT *et al.* (2006, p. 36), podem ser visualizadas conforme figura 3.

Figura 3 - Etapas da construção do conhecimento



Fonte: DUMONT *et al.* (2006, p. 36)

As ações de inteligência, conhecimento e sua gestão passam a ter um papel relevante na organização enquanto qualificam a informação e apresentam-se como uma forma de diminuir incertezas, suportar os gestores na tomada de decisão e otimizar as ações e estratégias organizacionais.

3.4.5. Integração de Dados

Os órgãos da Administração Direta e Indireta buscam enriquecer seus dados e transformá-los em informações de alto valor através da integração com diversas outras fontes de dados disponíveis. Integração esta que fornece um repositório estratégico único dos diferentes dados e/ou diferentes fontes que se fazem necessárias para cada setor em uma instituição pública.

Atualmente, as organizações têm focado suas estratégias na redução de custos, na melhoria da eficiência dos serviços prestados e na qualidade da tomada de decisão, sendo estes fatores determinantes para impulsionar a necessidade de simplificar e melhorar sua arquitetura de integração de dados.

Nesse sentido, a governança de dados ajuda as instituições a terem uma visão holística e a gerenciar dados no contexto de seus processos de negócio e a suportar as necessidades de integração de bases de dados dos diferentes órgãos para uma melhor geração de informação e conhecimento na elaboração de políticas públicas e no suporte a alta gestão.

Essa necessidade de integração na APF pode ser notada no Decreto Nº 8.789, que dispõe sobre o compartilhamento de bases de dados na Administração Pública Federal, nos diz em seu Art. 1º, a seguinte diretriz:

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, nos termos deste Decreto. (BRASIL, 2016b)

Corroborando com o objetivo a ser atingido, o Art. 2º do Decreto Nº 8.789, afirma-se que:

O acesso a dados de que trata o art. 1º tem como finalidades:

- I - a simplificação da oferta de serviços públicos;
- II - a formulação, a implementação, a avaliação e o monitoramento de políticas públicas;
- III - a análise da regularidade da concessão ou do pagamento de benefícios, ou da execução de políticas públicas;
- IV – a melhoria da qualidade e da fidedignidade dos dados constantes das bases dos órgãos e das entidades de que trata o art. 1º.

O tipo de base de dados mais comum para sistemas analíticos é o *Data Warehouse* - DW. O DW normalmente restringe o acesso ao banco de dados local para acessos somente leitura, fornecendo aos usuários e aplicativos institucionais um meio de integrar dados de diferentes fontes para os sistemas de suporte à decisão. Este será melhor explicado em um tópico específico.

A IBM, empresa atuante do setor de tecnologia, exemplifica que a “integração de dados é a combinação de processos técnicos e de negócios utilizados para combinar dados de fontes distintas em informações relevantes e de valor”².

Atualmente, para a realização de tarefas de integração de dados de várias fontes primárias, transformações nos dados com base em regras de negócio e seu

² <https://www.ibm.com/analytics/br/pt/technology/data-integration/> Acessado em: 05/02/2019.

carregamento em uma base de dados específica, temos as ferramentas de ETL, do inglês, *Extract, Transform and Load* – Extrair, Transformar e Carregar. Com essas ferramentas é possível mapear o caminho dado até sua inserção em uma base de dados, que em muitos casos é o DW. Neste trabalho, não iremos aprofundar o tema ETL, este será apenas para exemplificar como podem ser realizadas as etapas de integração das diversas fontes de dados.

Nesse sentido, iremos trabalhar com o conceito de integração de dados como sendo a disciplina que permite movimentar e visualizar dados de diferentes fontes, consolidá-las com outros tipos de dados necessários a instituição e aplicar transformações de negócios para que atendam às necessidades da organização, que no contexto deste trabalho é a APF.

Nesse sentido, DUMONT *et al.* (2006, p. 22) afirma que:

Dominar técnicas e ferramentas que permitem a geração e a aplicação do conhecimento é uma questão fundamental, pois este é o agente responsável que vai garantir às organizações maior agilidade, respondendo às demandas modernas de forma mais eficaz, eficiente e efetiva.

Pode-se perceber os benefícios dessa disciplina quanto à automatização das cargas de dados, integridade e precisão na transformação e a tempestividade, uma vez que podem ser definidos os períodos para as cargas dos dados nas bases analíticas, ou seja, possibilitar uma melhor tomada de decisão por meio de dados integrados e com maior qualidade em seu tempo correto.

A integração não precisa incluir a qualidade de dados, porém essas duas disciplinas podem ser implantadas em conjunto ou separadamente. No sentido de atingir melhores resultados, a qualidade deve ser priorizada, uma vez que integrar dados de baixa qualidade só fornece informação de má qualidade aos gestores e tomadores de decisão.

Para exemplificar o objetivo deste tópico, utilizaremos os dados do programa de governo PAA - Programa de Aquisição de Alimentos³ - e seus dados podem ser

³ <http://mds.gov.br/assuntos/seguranca-alimentar/programa-de-aquisicao-de-alimentos-paa> - Acesso em: 11/02/2019.

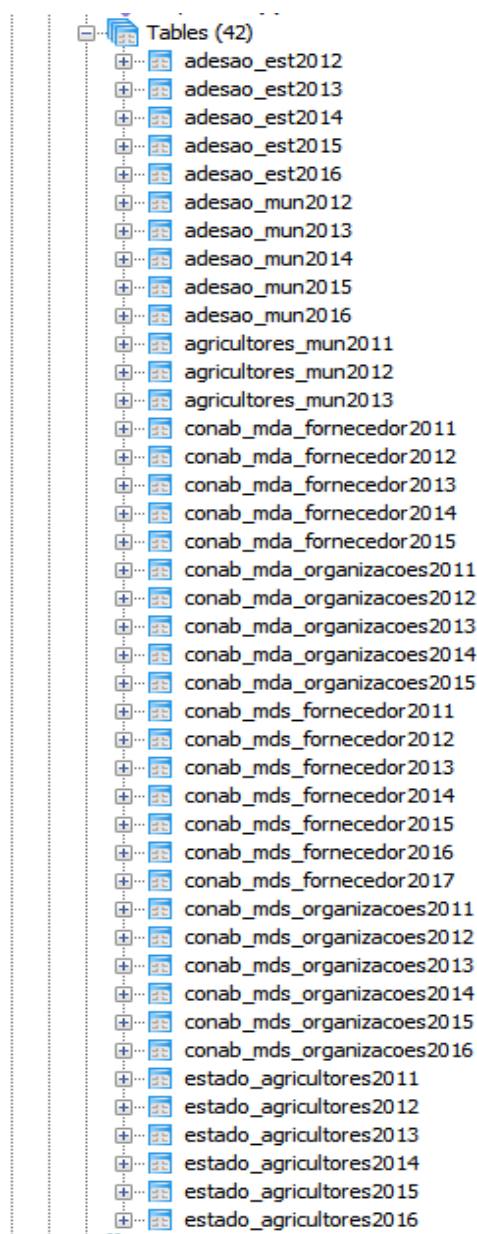
encontrados no portal do PAA Data⁴ em sua forma bruta. O objetivo é demonstrar de forma simples alguns métodos de integração de dados

Conforme a figura 4, é possível perceber que a base de dados utilizada desde a criação do PAA Data foi modelada separando os temas do programa, bem como os anos correspondentes. Neste exemplo utilizaremos apenas os dados de agricultores em todos os anos, que correspondem às tabelas de fornecedores, agricultores e em alguns anos as de adesão.

Em um primeiro momento foi necessário verificar os dados que são necessários para a área responsável pelo programa. Após essa validação, foi verificado se todos esses dados permanecem ao longo dos anos ou se sofreram algum tipo de mudança. Nesse caso específico, o conjunto dos dados praticamente se mantém inalterado.

⁴ https://aplicacoes.mds.gov.br/sagi/paa/visi_paa_geral/pg_principal.php?url=abertura - Acesso em: 11/02/2019.

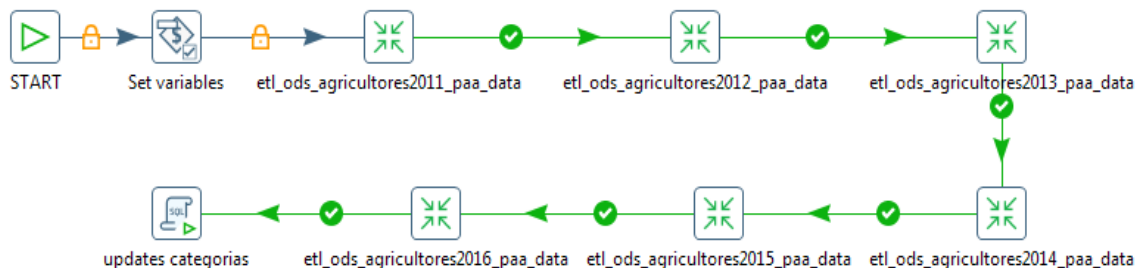
Figura 4 - Tabelas primárias PAA



Fonte: Elaborado pelo autor, 2019

Após a etapa de entendimento validação dos dados, foi utilizada uma ferramenta de ETL para extraí-los de suas tabelas de origem, transformar e enriquecer e finalmente inseri-los em uma nova tabela única que servirá como fonte para o DW. Nesse caso, os dados de agricultores serão mais um *Data Mart* dentro do DW. A estrutura geral dessa transformação pode ser visualizada na figura 5.

Figura 5 - ETL de integração das tabelas do PAA

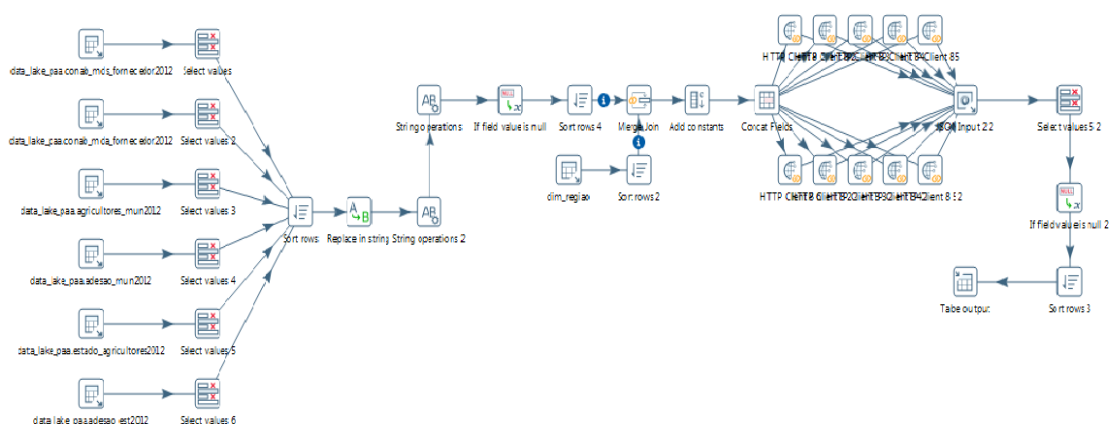


Fonte: Elaborado pelo autor, 2019

Toda essa cadeia de transformação utilizada como exemplo, foi apresentada como um todo. As especificidades de cada transformação não serão apresentadas, porém utilizaremos uma a título de conhecimento.

Nessa transformação, na figura 6, é possível visualizar a captura dos dados utilizados em sua fonte de origem, alguns processos de transformação e enriquecimento com outras fontes externas a da base de dados do programa PAA, e finalmente, sua inserção em uma base única que servirá como fonte de dados para o DW e outras possíveis formas de visualização e análise de dados.

Figura 6 - ETL de integração e qualidade



Fonte: Elaborado pelo autor, 2019

Apesar de tratarmos neste tópico sobre integração de dados, é possível realizar em conjunto procedimentos de integração e qualidade que será discutida no próximo tópico. Não existe uma regra que diga que cada procedimento deva ser

realizado separadamente, porém essas transformações dependem da estratégia adotada ou que melhor se adequa à necessidade.

Após todos os esses procedimentos, na figura 7, percebe-se que agora é possível realizar visualizações desses dados de forma horizontal e com série temporal. Nesse caso, foi utilizada uma ferramenta de *Data Discovery*. Se compararmos com o sistema atual, esse tipo de visualização de dados não é possível, sendo permitido somente a visualização por temas e anos específicos.

Dessa forma integrada, os gestores do programa podem tomar decisões melhores embasadas no que diz respeito a quantidade de agricultores anualmente, categoria e sexo. Essas duas dimensões foram utilizadas no exemplo para que se possa realizar a comparação em relação a qualidade do dado encontrado na base primária em um tópico específico.

Figura 7 - Visualização de dados com Data Discovery

Categoria	Sexo	2011	2012	2013	2014	2015	2016	Grand Total
AGRICULTOR FAMILIAR	FEMININO	23.945	34.457	23.218	30.070	26.890	23.021	161.601
	MASCULINO	62.314	68.975	32.831	36.837	33.941	26.836	264.734
	NÃO INFORMADO	242	11.859	182	358	16	-	12.657
AGRICULTOR FAMILIAR EM CONDIÇÕES ESPECIAIS (AUTORIZADOS PELO MDS)	FEMININO	100	-	58	82	94	98	432
	MASCULINO	283	-	36	62	63	84	508
	NÃO INFORMADO	2	-	1	3	1	-	7
AGROEXTRATIVISTA	FEMININO	157	404	402	1.288	1.543	2.777	6.571
	MASCULINO	608	710	233	1.227	1.561	1.477	5.816
	NÃO INFORMADO	15	-	-	-	-	-	15
AQUICULTOR FAMILIAR	FEMININO	189	395	270	218	44	19	1.135
	MASCULINO	387	817	423	441	95	59	2.222
	NÃO INFORMADO	-	-	-	-	-	-	-
ASSENTADOS DA REFORMA AGRÁRIA	FEMININO	4.447	9.849	6.545	6.913	4.333	2.921	34.808
	MASCULINO	11.489	17.543	7.229	6.756	4.280	2.683	49.960
	NÃO INFORMADO	179	1.492	4	9	-	-	1.684
COMUNIDADE INDÍGENA	FEMININO	206	404	276	208	222	289	1.605
	MASCULINO	520	659	238	238	266	235	2.156
	NÃO INFORMADO	-	-	1	1	-	-	2
FAMÍLIA ATINGIDA POR BARRAGEM	FEMININO	91	117	83	49	67	93	500
	MASCULINO	287	281	111	83	99	91	952
	NÃO INFORMADO	1	-	3	3	-	-	7
NÃO INFORMADA	FEMININO	6.302	247	8	-	-	-	6.557
	MASCULINO	12.549	247	26	-	-	1	12.823
	NÃO INFORMADO	-	-	-	-	-	-	-
PESCADOR ARTESANAL	FEMININO	466	831	842	468	377	457	3.241
	MASCULINO	2.087	1.989	1.129	748	517	566	7.036
	NÃO INFORMADO	3	-	2	4	-	-	9
QUILOMBOLA	FEMININO	487	962	308	741	634	658	3.790
	MASCULINO	551	746	257	528	390	436	2.908
	NÃO INFORMADO	-	-	1	2	-	-	3
TRABALHADOR RURAL SEM TERRA ACAMPADO	FEMININO	227	218	3	33	7	7	495
	MASCULINO	347	352	20	21	5	6	751
	NÃO INFORMADO	3	-	-	-	-	-	3

Fonte: Elaborado pelo autor, 2019

Apesar do exemplo utilizado ser simples, esses tipos de problemas de integração são muito comuns na APF, principalmente em bases de dados legadas ou de outras fontes solicitadas, como arquivos CSV e aplicações de Web Services.

3.4.6. Qualidade de Dados

A necessidade de melhoria da qualidade de dados é um dos principais motivos existentes para a criação de um programa voltado a GD. Muitas vezes é iniciado para melhorar a qualidade das tomadas de decisões, fusões entre órgãos ou para alinhar as necessidades da organização com seu planejamento estratégico. Essas iniciativas podem começar com um escopo amplo, para toda a organização, ou para uma necessidade local de um departamento ou projeto.

Segundo MOSLEY *et al.* (2012, p. 309), a “qualidade de dados é sinônimo de qualidade da informação, tendo em mente que falta de qualidade nos dados resulta em informações imprecisas e um desempenho fraco de negócios”.

De forma geral, dados de qualidade podem atender a vários objetivos. Seja para uma iniciativa de análise de dados, mineração de dados ou para uso com inteligência artificial, a qualidade sempre irá afetar o resultado na geração de informação, seja ela para pior ou melhor. Neste trabalho iremos tratar os dados com foco na sua melhoria contínua e com toda a qualidade possível no uso destes.

MOSLEY *et al.* (2012, p. 310) define a gestão da qualidade de dados como sendo o “planejamento e implementação de atividade que aplicam técnicas de gestão de qualidade de dados para medir, avaliar, otimizar e garantir dados adequados para uso”.

A qualidade pode se concentrar em vários momentos: na extração, na organização, no armazenamento, no processamento ou da apresentação dos dados ou informações. Em geral, a responsabilidade por armazenar e gerenciar os dados é de responsabilidade da área de TI, já a melhoria da qualidade pode ser feita pela TI e pelas áreas finalísticas de cada organização.

Sendo assim, os sistemas mais comuns de análise de dados são os desenvolvidos sob a plataforma de um DW e frequentemente apresentam vários problemas em relação à confiabilidade e à qualidade dos dados integrados. A principal razão para isso é que muitas vezes as fontes de dados primárias que participam da integração dos dados para o DW contêm dados incorretos, imprecisos, desatualizados ou simplesmente de baixa qualidade.

Vale ressaltar que a simples implantação de um sistema analítico para visualização da informação ou a implantação de um *Data Warehouse* não garante uma informação de qualidade ou resolve problemas de má qualidade no dado.

Nesse sentido, as ferramentas de qualidade têm foco na avaliação da qualidade dos dados, na normalização de valores conforme as regras de negócio definidas e, também na capacidade de identificar duplicações e ruídos nos dados a fim de corrigi-los e apresentá-los de uma única forma correta.

Para exemplificarmos algumas etapas de transformação de qualidade nos dados, utilizaremos os dados do PAA conforme demonstrado no tópico de integração. Podemos perceber, conforme figura 8, as categorias de agricultores que se repetem ao longo de quase todos os anos. Porém, no ano de 2013 é possível verificar algumas delas duplicadas ou com simples alterações na descrição.

Se utilizados esses dados em sua forma primária, na visualização analítica essas categorias irão aparecer como se fossem distintas. Porém, é um erro de cadastro delas na base de dados. Dessa forma, é necessário aplicar métodos de qualidade para que essas categorias sejam distintas umas das outras ao longo dos anos sem que haja duplicação ou inconsistência nos registros.

Figura 8 - Categorias de agricultores sem qualidade

	ds_ano text	ds_categoria text
1	2011	AGRICULTOR ESPECIAL
2	2011	AGRICULTOR FAMILIAR
3	2012	AGRICULTOR FAMILIAR
4	2013	AGRICULTOR FAMILIAR
5	2013	AGRICULTOR FAMILIAR EM CONDICOES ESPECIAIS (AUTORIZADOS PELO MDS)
6	2011	AGROEXTRATIVISTA
7	2013	AGROEXTRATIVISTA
8	2012	AGROEXTRATIVISTA
9	2013	AQUICULTORES FAMILIARES
10	2011	AQUICULTOR FAMILIAR
11	2013	AQUICULTOR FAMILIAR
12	2012	AQUICULTOR FAMILIAR
13	2013	ASSENTADO DA REFORMA AGRÁRIA
14	2012	ASSENTADO DA REFORMA AGRÁRIA
15	2011	ASSENTADO DA REFORMA AGRÁRIA
16	2013	ASSENTADO DE REFORMA AGRARIA
17	2013	ASSENTADOS DA REFORMA AGRÁRIA
18	2012	ATINGIDO POR BARRAGEM
19	2013	ATINGIDO POR BARRAGEM
20	2011	ATINGIDO POR BARRAGEM
21	2012	COMUNIDADE INDÍGENA
22	2011	COMUNIDADE INDÍGENA
23	2013	COMUNIDADE INDÍGENA
24	2013	FAMILIA ATINGIDA POR BARRAGEM
25	2013	FAMÍLIA ATINGIDA POR BARRAGEM
26	2013	INDIGENA
27	2012	NÃO INFORMADA
28	2011	NÃO INFORMADA
29	2013	NÃO INFORMADA
30	2013	PESCADOR ARTESANAL
31	2012	PESCADOR ARTESANAL
32	2011	PESCADOR ARTESANAL
33	2013	QUILOMBOLA
34	2011	QUILOMBOLA
35	2012	QUILOMBOLA
36	2013	TRABALHADOR RURAL SEM TERRA ACAMPADO
37	2012	TRABALHADOR RURAL SEM TERRA ACAMPADO
38	2011	TRABALHADOR RURAL SEM TERRA ACAMPADO

Fonte: Elaborado pelo autor, 2019

Após verificação das categorias dos agricultores, foi possível construir uma estratégia para padronização destas. Esse padrão adotado para as categorias em todos os anos pode ser visualizado na figura 9.

Figura 9 - Categoria de agricultores após qualidade

Data Output		Explain	Messages	History
	ds_ano text	ds_categoria text		
1	2015	AGRICULTOR FAMILIAR		
2	2014	AGRICULTOR FAMILIAR		
3	2013	AGRICULTOR FAMILIAR		
4	2012	AGRICULTOR FAMILIAR		
5	2011	AGRICULTOR FAMILIAR		
6	2013	AGRICULTOR FAMILIAR EM CONDICOES ESPECIAIS (AUTORIZADOS PELO MDS)		
7	2011	AGRICULTOR FAMILIAR EM CONDICOES ESPECIAIS (AUTORIZADOS PELO MDS)		
8	2014	AGRICULTOR FAMILIAR EM CONDICOES ESPECIAIS (AUTORIZADOS PELO MDS)		
9	2015	AGRICULTOR FAMILIAR EM CONDIÇÕES ESPECIAIS (AUTORIZADOS PELO MDS)		
10	2013	AGROEXTRATIVISTA		
11	2011	AGROEXTRATIVISTA		
12	2012	AGROEXTRATIVISTA		
13	2014	AGROEXTRATIVISTA		
14	2015	AGROEXTRATIVISTA		
15	2011	AQUICULTOR FAMILIAR		
16	2014	AQUICULTOR FAMILIAR		
17	2012	AQUICULTOR FAMILIAR		
18	2013	AQUICULTOR FAMILIAR		
19	2015	AQUICULTOR FAMILIAR		
20	2012	ASSENTADOS DA REFORMA AGRÁRIA		
21	2013	ASSENTADOS DA REFORMA AGRÁRIA		
22	2014	ASSENTADOS DA REFORMA AGRÁRIA		
23	2015	ASSENTADOS DA REFORMA AGRÁRIA		
24	2011	ASSENTADOS DA REFORMA AGRÁRIA		
25	2013	COMUNIDADE INDÍGENA		
26	2012	COMUNIDADE INDÍGENA		
27	2015	COMUNIDADE INDÍGENA		
28	2011	COMUNIDADE INDÍGENA		
29	2014	COMUNIDADE INDÍGENA		
30	2014	FAMÍLIA ATINGIDA POR BARRAGEM		
31	2015	FAMÍLIA ATINGIDA POR BARRAGEM		
32	2013	FAMÍLIA ATINGIDA POR BARRAGEM		
33	2011	FAMÍLIA ATINGIDA POR BARRAGEM		
34	2012	FAMÍLIA ATINGIDA POR BARRAGEM		
35	2011	NÃO INFORMADA		
36	2013	NÃO INFORMADA		
37	2012	NÃO INFORMADA		
38	2014	PESCADOR ARTESANAL		
39	2015	PESCADOR ARTESANAL		
40	2011	PESCADOR ARTESANAL		
41	2013	PESCADOR ARTESANAL		
42	2012	PESCADOR ARTESANAL		
43	2011	QUILOMBOLA		
44	2014	QUILOMBOLA		
45	2012	QUILOMBOLA		

Fonte: Elaborado pelo autor, 2019

Assim, se removermos os anos na nossa extração de exemplo é possível perceber um único registro para cada categoria conforme figura 10. Nesse sentido, os gestores podem tomar decisões e elaborar políticas públicas suportadas em dados com maior qualidade e consistência.

Figura 10 - Categorias únicas de agricultores

	ds_categoria text
1	AGRICULTOR FAMILIAR
2	AGRICULTOR FAMILIAR EM CONDIÇÕES ESPECIAIS (AUTORIZADOS PELO MDS)
3	AGROEXTRATIVISTA
4	AQUICULTOR FAMILIAR
5	ASSENTADOS DA REFORMA AGRÁRIA
6	COMUNIDADE INDÍGENA
7	FAMÍLIA ATINGIDA POR BARRAGEM
8	NÃO INFORMADA
9	PESCADOR ARTESANAL
10	QUILOMBOLA
11	TRABALHADOR RURAL SEM TERRA ACAMPADO

Fonte: Elaborado pelo autor, 2019

Da mesma forma, conforme figura 11, o sexo relacionado aos agricultores em sua fonte de origem possui descrições variadas para cada ano.

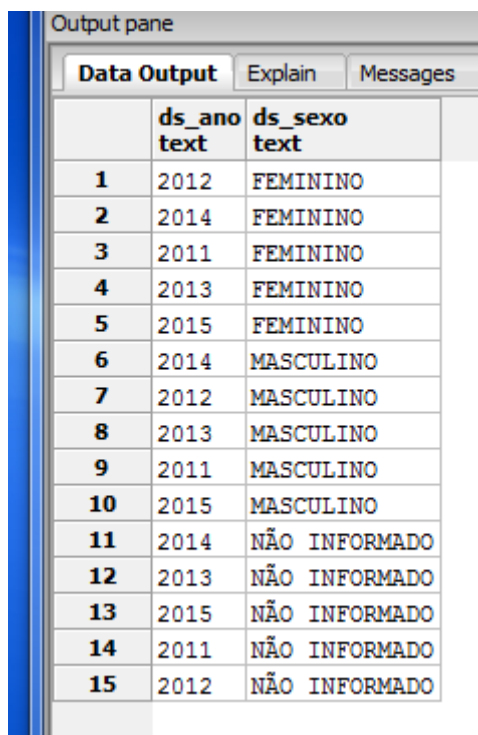
Figura 11 - Sexo agricultores sem qualidade

	ds_ano text	ds_sexo text
1	2015	F
2	2011	F
3	2014	F
4	2012	F
5	2013	F
6	2013	FEMININO
7	2015	FEMININO
8	2012	M
9	2013	M
10	2011	M
11	2015	M
12	2014	M
13	2013	MASCULINO
14	2015	MASCULINO
15	2014	NÃO INFORMADO
16	2013	NÃO INFORMADO
17	2011	NÃO INFORMADO
18	2012	NÃO INFORMADO
19	2015	

Fonte: Elaborado pelo autor, 2019

Após a conferência dessas classificações de sexo dos agricultores, foi realizado uma transformação com a finalidade de padronizar essas classificações. O resultado obtido pode ser visualizado na figura 12.

Figura 12 - Sexo agricultores após qualidade

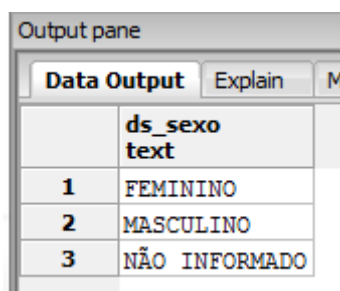


	ds_ano text	ds_sexo text
1	2012	FEMININO
2	2014	FEMININO
3	2011	FEMININO
4	2013	FEMININO
5	2015	FEMININO
6	2014	MASCULINO
7	2012	MASCULINO
8	2013	MASCULINO
9	2011	MASCULINO
10	2015	MASCULINO
11	2014	NÃO INFORMADO
12	2013	NÃO INFORMADO
13	2015	NÃO INFORMADO
14	2011	NÃO INFORMADO
15	2012	NÃO INFORMADO

Fonte: Elaborado pelo autor, 2019

Assim, se removermos a dimensão dos anos teremos apenas as classificações existentes na nova base de dados, que foi integrada e realizados tratamentos pontuais na melhoria da qualidade do dado disponibilizado, conforme figura 13.

Figura 13 - Sexo único agricultores



	ds_sexo text
1	FEMININO
2	MASCULINO
3	NÃO INFORMADO

Fonte: Elaborado pelo autor, 2019

Apesar dos exemplos estarem abordando apenas procedimentos de tratamento e padronização nos dados da base de dados utilizada, MOSLEY *et al.* (2012, p. 309) afirma que:

Na abordagem do programa, essas questões envolvem mais do que simplesmente correção de dados. Na realidade, eles envolvem gestão do ciclo de vida de criação, transformação e transmissão de dados para garantir que a informação encontre todos os consumidores de dados dentro da organização.

No caso deste trabalho, não abordaremos a questão do ciclo de vida do dado. Porém, no exemplo utilizado, os dados dos anos passados não sofrem alterações por se tratar de uma regra do próprio programa. Nesse sentido, ele não sofrerá mudanças conforme forem realizadas novas cargas de dados, sendo realizadas análises temporais comparando o momento atual com o passado.

3.4.7. Banco de Dados Corporativo

Após os dados estarem integrados, sido tratados, estarem confiáveis e com qualidade para suportar a tomada de decisão em forma de informação de alto valor, estes precisam de um ambiente para serem armazenados e futuramente analisados por ferramentas de análise e descoberta de conhecimento.

O armazenamento das informações deve ser feito em uma base de dados analítica, separada das outras bases existentes em uma organização, ou seja, nesta estará contido os ativos de informação mais precisos para gerar conhecimento e oferecer suporte de alto valor agregado na formulação de políticas públicas e tomada de decisão através da alta gestão.

Um modelo de banco de dados analítico existente é o *Enterprise Data Warehouse* (EDW) e segundo MOSLEY *et al.* (2012, p. 215) é “um *Data Warehouse* (DW) centralizado projetado para atender às necessidades de *business intelligence* (BI) de toda a organização. Um EDW adere ao modelo de dados da empresa para garantir a coerência das atividades de apoio à decisão em toda a empresa”.

Nesse sentido, existe uma diferença entre o *Data Warehouse* e o termo *Data Warehousing*. O primeiro está ligado ao repositório único e integrado que suporta as ações de análise de dados. O segundo é a utilização dos processos de ETL que mantêm os dados contidos em um DW.

Assim, MOSLEY *et al.* (2012, p. 241) corrobora quando afirma que o principal objetivo do DW é a integração de dados de várias fontes que serviram para finalidades analíticas.

4. INICIATIVAS NA ADMINISTRAÇÃO PÚBLICA FEDERAL

No âmbito do Governo Federal existe a “Plataforma de Análise de Dados do Governo Federal (GovData) que tem o objetivo de simplificar o acesso, compartilhamento e avaliação de gestores públicos a diferentes dados governamentais” (MPDG, 2018). Essa plataforma veio como resultado do Decreto Nº 8.789.

O objetivo dessa plataforma é que “será possível ter acesso a distintas bases de dados, infraestrutura de Tecnologia da Informação e, também às ferramentas necessárias para realizar o cruzamento de informações” (MPDG, 2018).

Para o devido alinhamento com a conformidade regulatória, MOSLEY *et al.* (2012 p. 60) nos diz que:

Parte da função da governança de dados é monitorar e garantir o cumprimento regulamentar. De fato, o cumprimento de regulamentação é frequentemente a razão inicial para a implementação de governança de dados. Governança de dados guia a aplicação de controles adequados para assegurar, documentar e monitorar a conformidade com as regulamentações relacionadas a dados.

A adesão a plataforma GovData é paga e destinadas aos órgãos da APF que têm interesse em realizar o cruzamento dos seus dados com o de outras instituições federais. A vantagem é que vários órgãos já realizaram uma parceria para disponibilização dos seus dados para a plataforma analítica do Governo Federal.

A construção, o desenvolvimento, a manutenção e as regras de negócio para adesão e utilização do GovData, ficam sob a responsabilidade administrativa do

Ministério do Planejamento e com utilização da estrutura do Serpro e Dataprev. Os conjuntos de dados e organizações que estão disponíveis na plataforma se encontram no catálogo de dados do GovData e que pode ser encontrado no endereço <http://ck.govdata.gov.br/>.

A plataforma também disponibiliza diversas ferramentas de visualização de dados, de uso estatístico e de *Data Discovery* (Descoberta de dados). O usuário pode optar por qualquer uma que melhor se alinhe com suas necessidades.

O GovData está sendo utilizado como a plataforma de análise e cruzamento de dados e informações no âmbito do Governo Federal. Seus benefícios já podem ser percebidos conforme notícias em mídias especializadas.

Ministro de Desenvolvimento Social, Alberto Beltrame:

As políticas sociais foram o primeiro alvo e, até por isso, se prestaram neste relançamento a servir de ‘garoto propaganda’ para o ‘Big Data’ do governo federal. Afinal, segundo o ministro de Desenvolvimento Social, Alberto Beltrame, até aqui 5,2 milhões de benefícios do Bolsa Família foram cancelados graças ao cruzamento de informações.⁵

Ministro do Planejamento, Dyogo Oliveira:

A GovData é uma preciosa ferramenta de tecnologia para o gestor público. O uso dos dados fornecidos na plataforma permitirá uma análise mais ágil das políticas públicas, como na concessão ou pagamento de benefícios sociais.⁶

Secretário Executivo do Ministério do Planejamento, Gleisson Rubin:

O GovData funciona como único repositório para onde vão as bases de dados de interesse. Mas mais do que isso, temos também uma série de recursos de análise estatísticas, geração de tendências e outras ferramentas de ciências de dados.⁷

Diante do exposto, nota-se que o uso de plataformas, sistemas ou mesmo bases de dados analíticas é uma necessidade atual e emergente. A APF precisa a cada dia melhorar as suas ações na elaboração de políticas públicas que realmente atendam às necessidades da sociedade.

⁵<http://www.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&inford=47765&sid=11> Acessado em: 18/02/2019.

⁶<http://www.brasil.gov.br/noticias/educacao-e-ciencia/2017/05/plataforma-govdata-auxilia-governo-no-combate-a-fraudes> Acessado em: 18/02/2019.

⁷<https://www.mixvale.com.br/2019/02/16/entenda-o-pente-fino-do-inss-em-2019-contra-irregularidades/> Acessado em: 18/02/2019.

Os gestores e agentes públicos que tomam decisões no seu dia-a-dia precisam estar seguros da decisão tomada com base em dados confiáveis e de qualidade. Deve-se levar em conta que as ações do Governo Federal por meio de seus agentes afetam diretamente as estratégias adotadas para o crescimento e desenvolvimento do Brasil.

5. CONCLUSÃO

Após as diversas conceituações de Governança, exemplos de integração e qualidade na manipulação de dados e informações e como a APF vem trabalhando na disponibilização da informação e geração de conhecimento através do GovData, fica claro a necessidade da modernização em todas essas etapas.

Essa necessidade é citada por DUMONT *et al.* (2006, p. 28) quando ele afirma que modernizar é um processo contínuo de transformação através do uso das novas tecnologias existentes, novos conceitos de gestão da informação e no atendimento às demandas da sociedade. Demandas estas, que também são requisitos na tomada de decisão e elaboração de políticas públicas por parte da alta gestão nas instituições federais.

Para CHOO (2013, p. 225), “o conhecimento estratégico de qualquer organização reside em sua capacidade de gerar conhecimentos duradouros, que são construídos ao longo do tempo”.

É importante ressaltar que a instituição deve estar envolvida como um todo no programa de governança de dados. Um problema que é comum de ser encontrado na APF é a retenção dos dados por parte de áreas específicas ou mesmo pelos próprios gestores que detêm o conhecimento em relação a certo dado ou informação.

CHOO (2013, p. 243) nos diz que:

A partilha de informações é uma condição necessária da construção do conhecimento. Ironicamente, quanto mais intensiva é a busca de informações na organização, menos probabilidade têm seus membros de partilhar livremente a informação.

Assim, é necessária uma mudança cultural na instituição no que tange a disponibilização dos dados, a construção da informação de alto valor e na geração do conhecimento para a tomada de decisão inteligente por parte dos gestores públicos no desempenho de suas funções.

Outro ponto importante é o entendimento de que a governança de dados se faz necessária em tempo de grandes volumes de dados a serem analisados. A construção e disponibilização da informação de alto valor, bem como a geração do conhecimento não são objetivos simples e fáceis de serem atingidos sem um

planejamento bem definido e o envolvimento da alta administração e instituição como um todo.

Apesar de terem sido demonstradas apenas as áreas de integração e qualidade de dados, a governança de dados trata outros temas conforme demonstrados nas figuras 1 e 2. Assim, os conjuntos de todas essas áreas da GD oferecem o suporte necessário para uma melhor gestão dos dados de um órgão da APF.

Desse modo, as áreas da APF que fazem a gestão da informação, gestão estratégica e a área de TI, precisam estar alinhadas na disponibilização da informação. Todas elas devem ter o seu papel definido na construção da informação a fim de gerarem melhores resultados no suporte a alta gestão para a tomada de decisão e elaboração de políticas públicas que realmente atendam os interesses da sociedade.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão (MPDG). PORTARIA Nº 58, DE 23 DE DEZEMBRO DE 2016 - Dispõe sobre procedimentos complementares para o compartilhamento de bases de dados oficiais entre órgãos e entidades da administração pública federal direta e indireta e as demais entidades controladas direta ou indiretamente pela União. Disponível em:

<<<https://www.governoeletronico.gov.br/documentos-e-arquivos/Portaria-MP%20no58%20de%202016.pdf/view>>>. Acesso em: 13/03/2018.

BRASIL. Decreto Nº 8.789, DE 29 DE JUNHO DE 2016 - Dispõe sobre o compartilhamento de bases de dados na administração pública federal. Disponível em: <<http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8789.htm>>. Acesso em: 13/03/2018.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão (MPDG). **GovData. Plataforma de Análise de Dados.** Disponível em:

<<<http://www.planejamento.gov.br/govdata-privado/plataforma-de-analise-de-dados-govdata>>>. Acesso em: 15/03/2018.

BRASIL. Tribunal de Contas da União - TCU. **Referencial básico de governança aplicável a órgãos e entidades da administração pública / Tribunal de Contas da União.** Versão 2 - Brasília: TCU, Secretaria de Planejamento, Governança e Gestão, 2014. 80 p.

CHOO, CHUN WEI. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões** | Chun Wei Choo; tradução Eliana Rocha. São Paulo: Editora Senac São Paulo, 2003.

DAMA-DMBOK. **Guide to the Data Management Body of Knowledge (DAMA-DMBOK v2)** 2017. Disponível em: <<<https://dama.org/content/body-knowledge>>>. Acesso em: 13/03/2018.

DUMONT, Danilo M. **Inteligência Pública na Era do Conhecimento** / Danilo M. Dumont, J. Araujo Ribeiro, Luiz A. Rodrigues, - Rio de Janeiro: Revan, 2006. 332p.

GRAHN, J.; AMOS, B.; PLUMPTRE, T. Institute of governance - IOG. **Governance principles for protected areas in the 21th century**, CA, 2003. Disponível em: <<https://iog.ca/docs/2003_June_pa_governance2.pdf>>. Acesso em: 07/10/2018.

IBGC. **Código das Melhores Práticas de Governança Corporativa**. 5. ed. São Paulo: IBGC, 2015. Disponível em: <<<http://conhecimento.ibgc.org.br/Paginas/Publicacao.aspx?PubId=21138>>>. Acesso em: 02/10/2018.

IFAC. International Federation of Accountants. **Governance in the public sector: a governing body perspective**. In International public sector study no 13., 2001. Disponível em: <<<http://www.ifac.org/sites/default/files/publications/files/study-13-governance-in-th.pdf>>>. Acesso em: 07/10/2018.

IFAC. International Federation of Accountants. **Good governance in the public sector: consultation draft for an international framework, 2013**. Disponível em: <<<https://www.iaasb.org/system/files/publications/files/International-Framework-Good-Governance-in-the-Public-Sector-supplement-IFAC-CIPFA-June-2014.pdf>>>. Acesso em: 07/10/2018.

LADLEY J. **Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program**. The Morgan Kaufmann Series on Business Intelligence. Morgan Kaufmann. 2012.

MATIAS-PEREIRA, J. **Governança no setor público**. São Paulo: Atlas, 2010.

MOSLEY, M.; BRACKETT, M.; EARLEY, S. HENDERSON, D. **DAMA DMBOK Guia para o corpo de conhecimento em gerenciamento de dados**. Technics Publications, versão brasileira 2012.

WORLD BANK. The International Bank for Reconstruction and Development. **Governance and Development**. 1992. Disponível em: <<<http://documents.worldbank.org/curated/pt/604951468739447676/Governance-and-development>>>. Acesso em: 03/10/2018.