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

ROMÁRIO CÉSAR DE ALMEIDA

Implantação de processos de gestão de dados e técnicas de recuperação de informações no Diário Oficial da União

Brasília
2019

ROMÁRIO CÉSAR DE ALMEIDA

Implantação de processos de gestão de dados e técnicas de recuperação de informações no Diário Oficial da União

Monografia apresentada ao Programa de Pós-graduação em Informática - Área de Concentração: Gestão de Tecnologia da Informação, da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do título de Especialista em Informática.

Orientador: Rodrygo Luis Teodoro Santos

Brasília
2019

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

Recuperação de Informações no Diário Oficial da União

ROMÁRIO CÉSAR DE ALMEIDA

Monografia apresentada aos Senhores:

Prof. Rodrygo Luis Teodoro Santos
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

© 2019, Romário César Almeida
Todos os direitos reservados

.

Almeida, Romário César

A447i Implantação de processos de gestão de dados e técnicas de recuperação de informações no Diário Oficial da União / Romário César Almeida. – Brasília, 2019.

xi, 44 f. : il.

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

Orientador: Rodrygo Luis Teodoro Santos

1. Computação – Monografias. 2. Gestão de Dados. 3. Recuperação de Informação. 4. DMBOK. 5. Administração Pública Federal. 6. Elastic; 7.Elasticsearch; 8. Logstash. 8. Kibana I. Orientador. II. Título.

CDU 519.6*

Resumo

Este documento trata de forma breve sobre possibilidades de implantação de funções de gestão de dados no órgão responsável pelo Diário Oficial da União e também nos demais órgãos mantenedores dos dados públicos, por meio do Guia DMBOK. Além disso, é apresentado um conjunto de ferramentas providas pela organização chamada Elastic, para otimização no armazenamento, indexação e recuperação das informações públicas.

Palavras-chave: Gestão de Dados; DMBOK; Diário Oficial da União; Administração Pública Federal; Elastic; Elasticsearch; Logstash; Kibana; Recuperação de Informação.

Abstract

This document deals briefly with possibilities for the implementation of data management functions in the body responsible for the Diário Oficial da União and also in the other bodies maintaining the public data, through the DMBOK Guide. In addition, a set of tools provided by the organization called Elastic is presented for optimization in the storage, indexing and retrieval of public information.

Keywords: Data Management; DMBOK; Diário Oficial da União; Federal Public Administration; Elastic; Elasticsearch; Logstash; Kibana; Information Retrieval.

Lista de ilustrações

Lista de tabelas

Lista de Siglas

APF	Administração Pública Federal
BACEN	Banco Central do Brasil
CF	Constituição Federal
CGU	Controladoria-Geral da União
CNPJ	Cadastro Nacional de Pessoa Jurídica
CPCC	Cartões de Pagamento do Governo Federal – Compras Centralizadas
CPDC	Cartões de Pagamento da Defesa Civil
CPF	Cadastro de Pessoa Física
CPGF	Cartões de Pagamento do Governo Federal
DOU	Diário Oficial da União
MPOG	Ministério do Planejamento, Orçamento e Gestão
RI	Recuperação da Informação
SPU	Secretaria do Patrimônio da União
STN	Secretaria do Tesouro Nacional
TCU	Tribunal de Contas da União
TI	Tecnologia da Informação

Sumário

1 Introdução

A década atual é marcada pelo auge da quarta revolução industrial, a revolução tecnológica. Um dos fatores de maior relevância para essa revolução é a análise de dados, que pode gerar informações de alto valor para determinados setores de negócio. Além disso, há também o avanço da transformação digital, fazendo com que mais dados sejam armazenados, transformados e analisados. É de conhecimento geral que empresas como Google e Facebook são as mais poderosas do planeta atualmente, e um dos fatores desse sucesso é a grande retenção de dados obtidos através de seus sistemas. A quantidade de dados dessas empresas é tão imensa, que é possível uma reinvenção do modelo de negócio de inúmeras formas diferentes.

Não obstante, além de ser algo bastante vantajoso para as empresas privadas, o setor público também se beneficia nessa onda da quarta revolução industrial. A digitalização de serviços públicos dá a oportunidade de obtenção de um grande número de dados relacionados às políticas e gestões públicas dos mais variados tipos. E com isso, é possível gerar inúmeras informações relevantes de forma com que o serviço público fique mais eficiente e transparente.

Este documento tem como objetivo a prospecção de melhorias acerca da gestão e recuperação de informações por meio do principal canal oficial de comunicação da Administração Pública Federal (APF): o Diário Oficial da União (DOU)

.

1.1 Problemas

Hoje, o canal oficial de divulgação de atos administrativos da esfera federal, o DOU, possui um sistema de recuperação de informações bastante pobre e ineficiente. Além disso, há também uma grande carência no cruzamento desses dados com dados de outras fontes de bases da APF.

Além do DOU, que dá a publicidade de atos oficiais, dentro da esfera do Poder Executivo Federal existem dois grandes portais de disponibilização de dados e informações. Um deles, o Portal da Transparência, dá publicidade de uma série de informações consolidadas com base em diversos dados advindos de sistemas dos órgãos. O segundo é o Portal de Dados Abertos, que tem como objetivo a disponibilização de dados brutos.

No entanto, não há dentro da APF, um sistema central abrangendo todos os Poderes: Executivo, Judiciário e Legislativo. E o problema gira em torno do DOU. Se já existe esse canal de comunicação oficial de todos os poderes da esfera federal, por que não integrá-lo com dados de sistemas dos órgãos e mostrar uma visão mais abrangente sobre Pessoas Físicas e Jurídicas?

1.2 Metodologia

Este documento será dividido em 2 partes:

- O Capítulo 2 apresentará a revisão dos normativos e da literatura que dispõe sobre o tema estudado;
- O Capítulo 3 apresenta a proposta, dividida em 2 partes: no tópico 3.2 uma revisão de literatura mais detalhada a respeito de práticas de gestão de dados e, no tópico 3.3 serão apresentados ferramentas e exemplos de utilização no contexto da proposta.

2 Revisão de normativos e de literatura

2.1 Princípio da Publicidade

A Constituição Federal de 1988 estabelece no Art. 37 que "a administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, **publicidade** e eficiência". A CF/88 garante expressamente o princípio da publicidade, e como complemento para a implementação desse princípio têm-se dois outros dispositivos legais: a Lei 12.527/2011 e o Decreto 9.215/2017.

A Lei 12.527/2011 dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso às informações, em consonância com o Art. 37 da CF/88. Segundo o 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."

Observa-se então, que desde 2011 existe uma previsão de utilização da tecnologia da informação como meio viável para o acesso à informação. Desde então, surgiram diversos meios dispersos de acesso à informação pela Internet.

Cabe ressaltar ainda, que a Lei 12.527/2011 conceitua informação como: "dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato".

Informações relativas aos atos com conteúdo normativo e aos atos oficiais são publicadas na íntegra no Diário Oficial da União - DOU, conforme Decreto 9.215/2017. Este Decreto dispõe sobre as normas gerais a serem seguidas na publicação do DOU. Uma delas estabelece, no art. 3º, que o DOU será exclusivamente eletrônico e será publicado no sítio eletrônico da Imprensa Nacional. No entanto, essa norma existe desde 2002, por meio do Decreto 4.520/2002. Há pelo menos 17 anos que a competência para publicar e dar acesso ao DOU é da Imprensa Nacional, órgão vinculado à Casa Civil, da Presidência da República. Decreto 9.215/2017 garante também que o meio de envio dos dados para o DOU deve ser exclusivamente eletrônico (art. 5º). Com essa medida, é possível

padronizar a forma de recebimento dos dados, facilitando o processo de extração, transformação e carregamento dos dados no DOU.

2.2 Dados Abertos

Além da legislação prever a disponibilização especificamente dos atos com conteúdo normativo e dos atos oficiais, há também previsão normativa sobre a disponibilização de dados públicos de forma geral. O Decreto 8.777/2016 institui a Política de Dados Abertos do Poder Executivo Federal. Conforme seu art. 1º, a Política tem como objetivo:

> "I - promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos;
>
> II - aprimorar a cultura de transparência pública;
>
> 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;
>
> IV - facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação;
>
> 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;
>
> VIII - promover o compartilhamento de recursos de tecnologia da informação, de maneira a evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações; e
>
> IX - promover a oferta de serviços públicos digitais de forma integrada."

Em relação aos objetivos listados pelo Decreto, é possível afirmar que os objetivos dos incisos I e III foram parcialmente implementados por meio da disponibilização do Portal de Dados Abertos, no endereço http://dados.gov.br. Parcialmente, pois nem todos os órgãos do Poder Executivo Federal disponibilizaram dados para serem publicados. Além disso, apesar de o portal ser encontrado na Internet, não há campanhas de divulgação para acesso do cidadão, tornando-o ineficiente para o objetivo do inciso III.

Sobre a transparência pública, citada no inciso II, será melhor detalhada na seção seguinte.

2.3 Transparência

A transparência pública está amparada pela Lei Complementar 131/2009, que altera outra Lei Complementar, a 101/2000. Naquela foi incluído o art. 48 parágrafo único, com destaque para o inciso II:

> "Parágrafo único. A transparência será assegurada também mediante:

> . . .
> II - liberação ao pleno conhecimento e acompanhamento da sociedade, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público; . . . "

Em 2004, a Controladoria-Geral da União - CGU lançou o Portal da Transparência (http://www.portaltransparencia.gov.br), em atendimento ao disposto na Lei Complementar. O Portal da Transparência foi reformulado em 2018 reforçando, assim, "com novos recursos e mais informações, sua razão de ser uma ferramenta que permita ao cidadão, de forma cada vez mais eficiente, fiscalizar e assegurar a boa e correta aplicação dos recursos públicos federais.".

No Portal da Transparência são disponibilizados dados sobre:

- Orçamento Anual
- Receitas Públicas
- Despesas Públicas
- Recursos Transferidos
- Gastos por Cartão de Pagamento
- Áreas de Atuação do Governo
- Programas de Governo
- Benefícios aos Cidadãos
- Programas e Ações Orçamentárias
- Emendas Parlamentares
- Órgãos do Governo
- Servidores Públicos
- Viagens a Serviço
- Imóveis funcionais
- Licitações
- Contratações
- Convênios e outros Acordos
- Sanções

No portal encontra-se ainda a origem dos dados que são disponibilizados. De acordo com o portal:

> "O Portal da Transparência integra e apresenta dados de diversos sistemas utilizados pelo Governo Federal para a sua gestão financeira e administrativa, objetivando prover transparência da gestão pública, além de instrumentalizar a sociedade para a realização do controle social.
> Os dados são recebidos com periodicidade diária, semanal e mensal, a depender do tema, e são de responsabilidade dos ministérios e outros órgãos do Poder Executivo Federal, por serem eles os executores dos programas de governo e os

responsáveis pela gestão das ações governamentais “.
http://www.portaltransparencia.gov.br

A Tabela 1 descreve a origem dos dados disponibilizados:

Tabela 1 – Origem dos dados do Portal da Transparência

ORIGEM/ARQUIVO	ÓRGÃO/ENTIDADE RESPONSÁVEL PELO ENVIO	PERIODICIDADE
DESPESAS PÚBLICAS		
Secretaria do Tesouro Nacional – STN (Origem SIAFI) - Despesas Diárias	Secretaria do Tesouro Nacional – STN	Diário
Transferências constitucionais e Royalties	Secretaria do Tesouro Nacional – STN	Mensal
RECEITAS PÚBLICAS		
Secretaria do Tesouro Nacional – STN (Origem SIAFI) - Receitas	Secretaria do Tesouro Nacional – STN	Diário
LICITAÇÕES E CONTRATOS		
SIASG - Sistema Integrado de Administração de Serviços Gerais	Ministério do Planejamento, Desenvolvimento e Gestão	Mensal
CONVÊNIOS E OUTROS ACORDOS		
Convênios - SIAFI	Secretaria do Tesouro Nacional	Semanal
Convênios - SICONV	Ministério do Planejamento, Desenvolvimento e Gestão	Semanal
BENEFÍCIOS AOS CIDADÃOS		
Caixa Econômica Federal - Seguro Defeso	Ministério do Trabalho e Emprego	Mensal

Caixa Econômica Federal - Bolsa Família	Ministério do Desenvolvimento Social	Mensal
Caixa Econômica Federal - Garantia Safra	Secretaria de Desenvolvimento Agrário	Mensal
Caixa Econômica Federal - PETI	Ministério do Desenvolvimento Social	Mensal
SERVIDORES		
Servidores - Banco Central do Brasil (BACEN)	Banco Central do Brasil (BACEN)	Mensal
Servidores - SIAPE	Ministério do Planejamento, Orçamento e Gestão - MPOG	Mensal
Servidores - Comandos Militares	Comandos Militares	Mensal
CARTÕES DE PAGAMENTO		
Banco do Brasil - Cartões de Pagamento do Governo Federal (CPGF)	Banco do Brasil	Mensal
Banco do Brasil - Cartões de Pagamento da Defesa Civil (CPDC)	Banco do Brasil	Mensal
Banco do Brasil - Cartões de Pagamento do Governo Federal - Compras Centralizadas (CPCC)	Banco do Brasil	Mensal
VIAGENS A SERVIÇO		
Diárias e passagens - SCDP	Ministério do Planejamento, Desenvolvimento e Gestão	Mensal
IMÓVEIS FUNCIONAIS		
Imóveis funcionais - Ministério da Defesa	Ministério da Defesa	Sob Demanda

Imóveis funcionais Secretaria do Patrimônio da União - SPU	Secretaria do Patrimônio da União - SPU	Sob Demanda
Imóveis funcionais - Presidência da República	Presidência da República	Sob Demanda
Imóveis funcionais Ministério das Relações Exteriores	Ministério das Relações Exteriores	Sob Demanda
SANÇÕES		
Sistema Integrado de Registro do CEIS/CNEP - CNEP	Ministério do Transparência e Controladoria-Geral da União	A cada quatro horas
Sistema Integrado de Registro do CEIS/CNEP - CEIS	Ministério do Transparência e Controladoria-Geral da União	A cada quatro horas
SIAFI - CEPIM	Secretaria do Tesouro Nacional	Diário
Diário Oficial da União - CEAF	Ministério do Transparência e Controladoria-Geral da União	Mensal
ORÇAMENTO PÚBLICO		
Orçamento da despesa	Secretaria do Tesouro Nacional – STN	Diário
Orçamento da receita	Secretaria do Tesouro Nacional – STN	Diário

Fonte: Elaborado pelo autor

Como descrito anteriormente, o inciso II da Lei Complementar 131/2009 prevê a "liberação ao pleno conhecimento e acompanhamento da sociedade, **em tempo real**, de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público". Destaca-se o termo "em tempo real" que, em outras palavras, exige uma atualização a cada vez que um dado é atualizado. Porém, conforme observado na Tabela 1, de todos os dados que são disponibilizados no Portal da Transparência, nenhum possui a periodicidade "em tempo real".

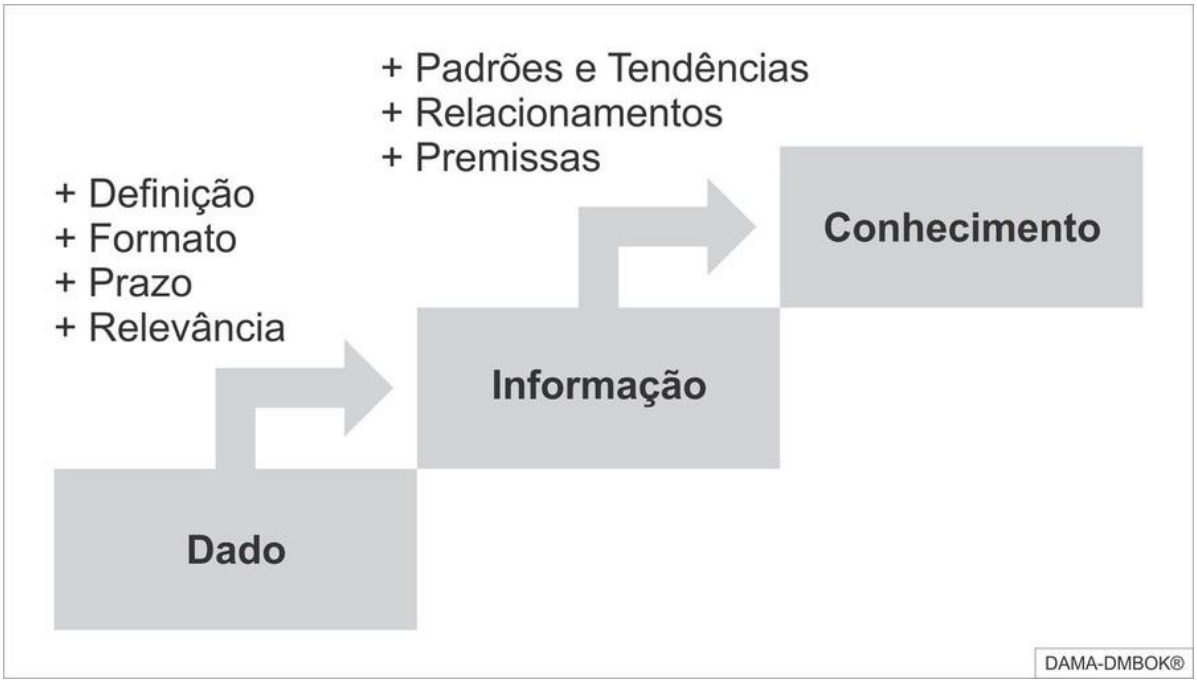
2.4 Gestão de dados - DMBOK

O DMBOK é um guia de boas práticas para gestão de dados, referência internacional no assunto. Em citação à Tom Peters (2001 apud DAMA, 2012; p. 9), o guia prega que ": as organizações que não entenderem a enorme importância da gestão de dados e informações como ativos tangíveis na nova economia não sobreviverão".

2.4.1 Dados x Informação

Dados são a representação de fatos em forma de textos, números, gráficos, imagens, som ou vídeo. Informação são dados em contexto. Sem contexto, o dado não tem significado; criamos informações significativas ao interpretar o contexto em torno do dado. Dado é a matéria-prima que nós consumidores de dados interpretamos para criar informações continuamente. (DAMA, 2012)

Figura 1 – Dado, Informação e Conhecimento



Fonte: DMBOK, 2012

2.4.2 Framework Funcional

O DMBOK institui um framework funcional com o objetivo de padronizar a aplicação da gestão de dados nas organizações. O framework é formado pela relação entre Funções de Gestão de Dados e Elementos Ambientais.

2.4.2.1 Funções de Gestão de Dados

O guia estabelece 10 funções para a gestão de dados. São apresentadas a seguir com uma respectiva descrição macro:

1) **Governança de Dados**: planejamento, supervisão e controle sobre uso e gestão de dados;
2) **Gestão da Arquitetura de Dados**: definição do diagrama para a gestão dos ativos de dados;
3) **Desenvolvimento de Dados**: análise, estruturação, implementação, testes, implantação e manutenção;
4) **Gestão Operacional de Dados**: presta suporte desde a aquisição de dados até a eliminação plena do dado;
5) **Gestão de Segurança de Dados**: garante a privacidade, confidencialidade e acesso apropriado;
6) **Gestão da Qualidade de Dados**: define, monitora e incrementa melhorias na qualidade dos dados;
7) **Gestão de Dados Mestre e Referência**: gerencia as versões "douradas" e réplicas;
8) **Gestão de Data Warehousing e Business Intelligence**: gera relatórios e análises;
9) **Gestão de Conteúdo e de Documento**: gerencia dados localizados fora de bases de dados;
10) **Gestão de meta-dados**: integra, controla e fornece meta-dados.

2.4.2.2 Elementos Ambientais

Além de identificar 10 funções de gestão de dados, o *framework* também identifica 7 Elementos Ambientais, nas quais cada uma se relaciona com todas as funções de gestão, conforme a Figura 2:

1) **Metas e Princípios**: as metas direcionais de negócios de cada função e os princípios fundamentais que dirigem o desempenho de cada função;
2) **Atividades**: cada função é composta de atividades de nível inferior. Algumas atividades são agrupadas em sub-atividades. As atividades são decompostas em tarefas e etapas.
3) **Entregas Primárias**: as informações e bancos de dados físicos e documentos criados como resultados intermediários e finais de cada função. Algumas entregas são essenciais, algumas são geralmente recomendadas, e outras são opcionais, dependendo das circunstâncias;
4) **Papéis e Responsabilidades**: os papéis da área de negócio e de TI envolvem o desempenho e supervisão das funções, e as responsabilidades específicas de cada função. Muitos papéis irão participar em múltiplas funções;
5) **Práticas e Técnicas**: métodos comuns e populares, e procedimentos utilizados para executar os processos e produzir entregas. Práticas e Técnicas podem também incluir convenções comuns, recomendações de melhores práticas, e abordagens alternativas sem elaboração;

6) **Tecnologia**: categorias de apoio à tecnologia, padrões e protocolos, critérios de seleção de produtos e curvas comuns de aprendizagem. Em conformidade com as políticas da DAMA International® vendedores ou produtos não são mencionados.
7) **Organização e Cultura**: essas questões podem incluir:
    - Gestão de Métricas de medidas de tamanho, esforço, tempo, custo, qualidade, eficácia, produtividade, sucesso e valor do negócio;
    - Fatores críticos de sucesso;
    - Relatório de estruturas;
    - Estratégias de contratação;
    - Questões relacionadas a orçamento e alocação de recursos;
    - Trabalho em equipe e dinâmica de grupo;
    - Autoridade e poder para decidir;
    - Valores e crenças compartilhadas;
    - Expectativas e atitudes;
    - Estilo pessoal e diferença de preferências;
    - Ritos culturais;
    - Herança organizacional;
    - Recomendações de gestão de mudanças.

Figura 2 – Matriz - Funções de Gestão de Dados x Elementos de Ambiente

Funções de Gestão de Dados	Metas e Princípios	Atividades	Entregas Primárias	Papéis e Responsabilidades	Tecnologia	Práticas e Técnicas	Organização e Cultura
Governança de Dados							
Gestão da Arquitetura de Dados							
Desenvolvimento de Dados							
Gestão de Operações de Dados							
Gestão da Segurança de Dados							
Gestão de Dados Mestres e Referência							
Gestão de DW e BI							
Gestão da Documentação e Conteúdo							
Gestão de Meta-Dados							
Gestão da Qualidade de Dados							

DAMA-DMBOK®

Fonte: DMBOK, 2012

2.5 Recuperação de Informação

Baeza-Yates e Ribeiro Neto (2013), conceituam Recuperação da Informação - RI da seguinte forma:

> "A Recuperação de Informação trata de representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse." (BAEZA -YATES; RIBEIRO NETO 2013, p. 1)

Segundo os autores, a pesquisa em RI inclui modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuário, visualização de dados, filtragem e linguagens. A área pode ser estudada sob dois pontos de vista distintos e complementares. O primeiro ponto de vista é centrado no computador, na qual a RI consiste principalmente na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento. O segundo é centrado no usuário, onde a RI consiste em estudar o comportamento do usuário, entender suas principais necessidades e determinar como esse entendimento afeta a organização e a operação do sistema de recuperação.

Segundo Baeza-Yates e Ribeiro Neto (2013) um sistema de RI deve de alguma forma "interpretar" o conteúdo dos itens de informação, isto é, dos documentos de uma coleção, e classificá-los de acordo com o grau de relevância à consulta do usuário. E isso envolve a extração de informações sintáticas e semânticas do texto do documento e sua utilização para satisfazer a necessidade do usuário. O autor destaca o problema da RI:

> "o objetivo principal de um sistema de RI é recuperar todos os documentos que são relevantes à necessidade de informação do usuário e, ao mesmo tempo, recuperar o menor número possível de documentos irrelevantes." ((BAEZA - YATES; RIBEIRO NETO 2013, p. 4)

Este trabalho focará na visão centrada no computador e na utilização de informações sintáticas.

3 Proposta

3.1 Visão Geral

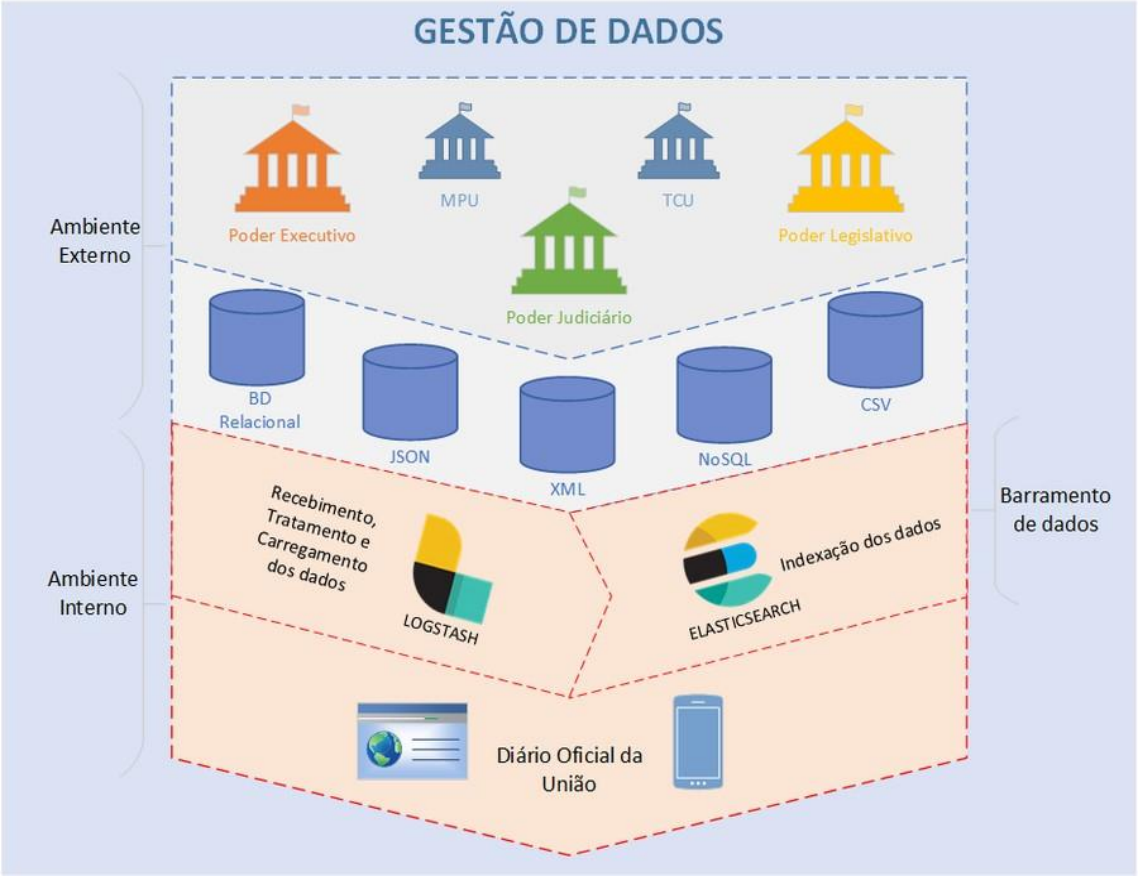
A revisão realizada anteriormente deixou clara a existência de diversos pontos de publicidade dos dados públicos. No entanto, é possível afirmar que não há integração entre eles. A questão é que, dados espalhados dessa forma deixam a busca ineficiente por informações, ainda mais quando o cidadão não sabe que existem tantas opções.

Arrisco a dizer que a maior parte dos dados públicos são, de alguma forma, relacionados às Pessoas Físicas ou Pessoas Jurídicas. E é por meio do Diário Oficial da União (DOU) que há a publicidade de atos oficiais envolvendo-os. É fato que o DOU é o meio oficial para publicação desses atos, por força de Decreto 9.215/2017. E acredito que continuará assim por um longo tempo.

Mas eis a questão: se o DOU é o meio oficial para dar essa publicidade, por que não mostrar dados provenientes do Portal da Transparência e do Portal de Dados Abertos? E vou mais além. O DOU é responsável por dar publicidade não só dos atos oficiais do Poder Executivo Federal, mas também do Poder Judiciário, Poder Legislativo, MPU e TCU. Muitas pessoas físicas e jurídicas são citadas em atos de todos esses poderes. A junção de todos os dados dessas entidades em uma só ferramenta poderia gerar alta eficiência na busca de informações e transformação em conhecimento.

A Figura 3 demonstra, de forma macro, uma arquitetura para a solução que será proposta:

Figura 3 – Arquitetura macro da solução



Fonte:

3.2 Gestão de Dados

Antes de se falar em integração entre dados de inúmeras fontes diferentes, é necessário dar um passo pra trás e pensar: como gerir tamanha quantidade de dados e fazer com que agreguem o valor esperado? Para isso, existe o Guia DMBOK, conforme estudado anteriormente. Esse guia servirá para direcionar os trabalhos na implantação de uma gestão de dados eficiente.

A proposta, apresentada neste trabalho, possui uma abrangência gigantesca. Por isso, não basta desenvolver soluções, integrar e disponibilizar de forma *ad hoc*. É necessária uma camada anterior, na qual fará todo o planejamento e controle das ações que envolvem dados.

3.2.1 Fronteira de atuação

Podemos dividir a proposta entre duas linhas de atuação. A primeira é a atuação externa, dos órgãos proprietários dos dados públicos. A missão desses órgãos é manter uma gestão e infraestrutura de dados adequadas para a disponibilização eficiente dos dados para a Imprensa Nacional. Nesses órgãos, os dados públicos passam por todo o ciclo de vida, conforme

a Figura 4. Portanto, faz-se necessário que todo o Guia DMBOK seja observado, com todas as funções de gestão de dados devidamente customizadas para a realidade, e implementadas.

Figura 4 – O Ciclo de Vida do Dado

Fonte: DAMA, 2012 - adaptado

A segunda linha de atuação é mais pontual. Trata-se da atuação da Imprensa Nacional, que é receber os dados dos órgãos proprietários e integrá-los com os dados do DOU. Essa é a linha que será explorada neste trabalho.

3.2.2 Gestão de dados na Imprensa Nacional

O escopo de atuação da Imprensa Nacional é reduzido, se comparado ao escopo dos órgãos como um todo. É responsável apenas por receber, manipular e publicar os dados. Não faz parte das atribuições deste órgão a construção do dado. Portanto, não é a Imprensa Nacional que definirá regras de negócio para os dados em si, mas deve definir para a estrutura dos dados que serão recebidos.

O dado deve ser bem definido em sua própria origem. Os órgãos externos devem ser responsáveis por:

- Definir regras de negócio dos dados;
- Prover sistemas para o levantamento dos dados;
- Prover bases de dados bem estruturadas;
- Disponibilizar serviços para extração dos dados;
- Definir regras de acesso aos dados;
- Prover infraestrutura adequada para acesso aos dados;
- Entre outras.

Tais responsabilidades devem ser devidamente controladas pelo processo de gestão de dados de cada órgão. No entanto, após o dado entrar no domínio da Imprensa Nacional, deve haver também uma gestão desses dados.

3.2.2.1 Funções de gestão de dados

A seguir será mostrado de que forma é possível implementar cada uma das funções de gestão de dados, conforme Seção 2.4.2.1, no ambiente interno da Imprensa Nacional.

3.2.2.1.1 Governança de Dados

Segundo (DAMA, 2012), a Governança de Dados é a função central do framework de gestão de dados, pois, interage e influencia cada uma das nove outras funções, conforme Figura 5. A partir dela é realizado o planejamento, monitoramento e execução sobre a gestão de ativos de dados.

Figura 5 – Abrangência da Governança de Dados

Fonte: DMBOK

No contexto atual, são outros órgãos que possuem o domínio sobre os dados. Nesse caso, a Imprensa Nacional pode atuar como um centralizador, utilizando um Escritório de Governança de Dados, e cada órgão externo de forma descentralizada, indicando um Gestor de Dados para ser o ponto focal na interlocução com o Escritório de Governança de Dados.

3.2.2.1.2 Gestão da Arquitetura de Dados

De acordo com o DMBOK, o processo da função de gestão da arquitetura de dados define os dados necessários da organização e o desenho do diagrama mestre para atingir essas necessidades.

Essa função tem como objetivos:

1) Planejar com visão e previsão para prover dados de alta qualidade.
2) Identificar e definir requisitos comuns de dados.
3) Desenhar estruturas conceituais e planos para atingir requisitos correntes e de longo prazo da organização.

3.2.2.1.3 Desenvolvimento de Dados

Para o DMBOK, a função de desenvolvimento de dados trata da análise, projeto, implementação, implantação e manutenção de soluções de dados para maximizar o valor dos recursos de dados da organização.

Essa função tem como objetivos:

1) Identificar e definir requisitos de dados;
2) Projetar estruturas de dados e outras soluções para esses requisitos;
3) Implementar e manter componentes de soluções que atendam esses requisitos;
4) Garantir solução em conformidade para a arquitetura de dados e padrões apropriados;
5) Garantir a integridade, segurança, usabilidade e manutenibilidade da estrutura dos ativos de dados.

3.2.2.1.4 Gestão de Operação com Dados

O DMBOK prega que essa função trata sobre o desenvolvimento, a manutenção, e suporte de dados estruturados para maximizar o valor dos recursos de dados da organização. Promove suporte desde a aquisição de dados até a eliminação dos dados.

Essa função tem como objetivos:

1) Proteger e garantir a integridade dos ativos de dados estruturados;
2) Gerenciar a disponibilidade do dado em todo seu ciclo de vida;
3) Melhorar continuamente o desempenho das transações de bancos de dados.

As atividades principais para este contexto são:

3.2.2.1.5 Gestão de Segurança de Dados

A função de gestão de segurança de dados é responsável por planejar, desenvolver e executar procedimentos e políticas de segurança para prover autenticação, autorização, acesso e auditoria de dados e informação.

Essa função tem como objetivos:

1) Permitir apropriado (e prevenir inapropriado) acesso e alteração em ativos de dados;
2) Atender requisitos regulatórios para privacidade e confidencialidade;
3) Garantir que as necessidades de privacidade e confidencialidade de todas as partes interessadas sejam atendidas.

3.2.2.1.6 Gestão de Dados Mestres e de Referência

É um processo contínuo de reconciliação e manutenção dos dados mestre e dados de referência.

Gestão de dados de referência é o controle sobre valores de domínio definidos (também conhecidos como vocabulários), incluindo o controle sobre termos padronizados, códigos e outros identificadores únicos, definições de negócio para cada valor dos códigos, relacionamentos de negócio dentro e entre listas de domínios, uso compartilhado de dados de referência, relevantes, consistentes, precisos, gerados em tempo e atualizados para classificar e categorizar os dados.

Gestão de dados mestres é o controle sobre os valores dos dados mestres para viabilizar o uso contextual consistente, compartilhado entre os sistemas, da mais acurada, gerada em tempo e relevante verdadeira versão sobre as entidades essenciais do negócio.

Essa função tem como objetivos:

1) Prover fonte autorizada para reconciliação, dados mestre e de referência com alta qualidade;
2) Baixar custos e complexidade por meio de reuso e alavancagem de padrões; 3) Suportar BI e esforços de integração de informações.

3.2.2.1.7 Gestão de Data Warehousing e Business Intelligence

O Data Warehousing (DW) é a combinação de 2 componentes primários. O primeiro é um banco de dados de suporte integrado à decisão. O segundo é um programa de *software* relacionado usado para coletar, limpar, transformar e armazenar dados de uma variedade de fontes operacionais e externas. Ambas as partes se combinam para atender às exigências

históricas, analíticas e de Business Intelligence (BI). Um Data Warehouse também pode incluir dependentes data marts, os quais são cópias de um subconjunto de um banco de dados de Data warehouse. Neste amplo conceito um Data Warehouse inclui qualquer depósito ou extrato de dados que sejam utilizados para dar suporte a entrega de dados para quaisquer fins de BI.

Essa função tem como objetivos:

1) Suportar e permitir efetiva análise de negócios e tomada de decisões por trabalhadores de conhecimento;
2) Construir e manter o ambiente/infraestrutura para suportar atividades de BI, especificamente alavancagem de outras funções de gestão de dados para entregar dados integrados consistentes para todas as atividades de BI.

3.2.2.1.8 Gestão de Documentos e Conteúdo

É o controle sobre a obtenção, o armazenamento, o acesso e a utilização dos dados e informações armazenados fora dos bancos de dados relacionais.

Essa função tem como objetivos:

1) Proteger e garantir a disponibilidade de ativos de dados armazenados em formatos menos estruturados;
2) Permitir recuperação e uso efetivo e eficiente de dados e informações em formatos não estruturados;
3) Cumprir com obrigações legais e expectativas de clientes;
4) Garantir a continuidade do negócio por meio da retenção, recuperação e conversão;
5) Controlar os cursos operacionais de armazenamento de documentos.

3.2.2.1.9 Gestão de Metadados

Segundo o DMBOK, metadado é a ficha do catálogo de cartões em um ambiente de dados gerenciados. São etiquetas descritivas ou o contexto que se aplica aos dados em um ambiente de dados gerenciados. Mostram aos usuários técnicos e de negócios onde encontrar informações em repositórios de dados. Além disso, fornecem detalhes sobre de onde vêm os dados e como chegaram lá, todas as transformações e seu nível de qualidade, fornecendo assistência com o que realmente significa os dados e maneiras de como interpretá-lo.

Essa função tem como objetivos:

1) Prover entendimento organizacional no uso de termos;
2) Integrar metadados de diversas fontes;
3) Prover fácil acesso integrado aos metadados;

4) Garantir metadados de qualidade e segurança.

3.2.2.1.10 Gestão da Qualidade de Dados

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.

Gestão da qualidade de dados é um processo contínuo para os parâmetros de definição que especificam níveis adequados de qualidade de dados de acordo com as necessidades do negócio, garantindo assim que a qualidade dos dados se mantenha nesse nível.

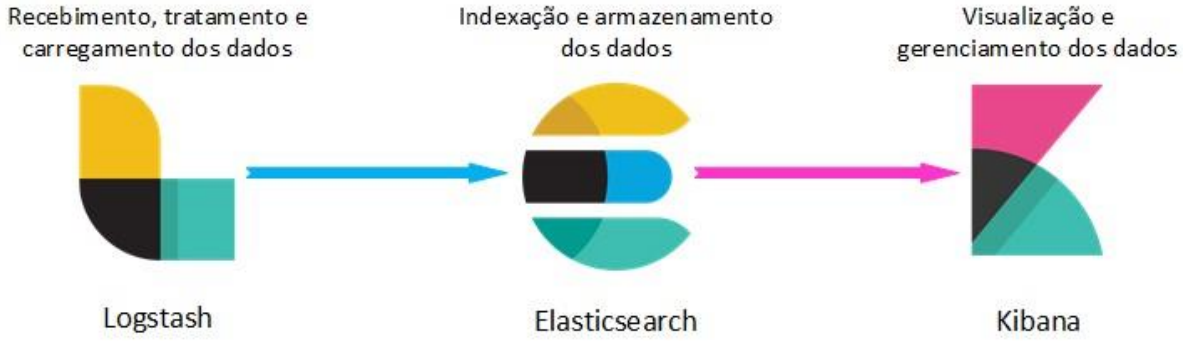
Essa função tem como objetivos:

1) Incrementar as medidas de qualidade de dados em relação às expectativas definidas por negócios;
2) Definir requisitos e especificações para integração de controle de qualidade de dados dentro do ciclo de desenvolvimento de sistemas;
3) prover processos definidor para medir, monitorar e reportar conformidade para níveis aceitáveis de qualidade de dados.

3.3 Recuperação de Informações com o ELK Stack

Segundo a página *web* do Elastic, ELK é o acrônimo para três projetos de código aberto: Elasticsearch, Logstash e Kibana. O Elasticsearch é um mecanismo de pesquisa e análise. O Logstash é um pipeline de processamento de dados do lado do servidor. O Kibana permite a visualização dos dados. A Figura 6 ilustra a utilização dessas ferramentas em conjunto.

Figura 6 – ELK Stack



FONTE:

3.3.1 Elasticsearch

Segundo o Guia de Referência, “o Elasticsearch é um *engine* de análise e busca de texto altamente escalável, e completamente em código aberto. Ele permite que você armazene, busque e analise grandes quantidades de dados e em tempo quase real. É utilizado, geralmente,

como um mecanismo fundamental em aplicações que possuem complexos recursos e requisitos de busca".

Esse *engine* é desenvolvido com base no Apache Lucene, que é uma biblioteca de *engines* de busca de textos, de alto desempenho e repleta de recursos, escrita em Java. É uma tecnologia adequada para praticamente qualquer aplicativo que exija pesquisa de texto completo.

O Elasticsearch realiza três funções básicas: armazenamento, indexação e busca de dados.

Para armazenar os dados que serão tratados, o Elasticsearch conta com um banco de dados similar ao banco de dados NoSQL baseado em documentos. Além disso, o Elasticsearch utiliza nativamente o padrão de arquivo JSON. Segundo PRAGMATEEK (2013), ao comparar o JSON com XML, conclui-se que o JSON é mais vantajoso, visto que arquivos nesse padrão são processados em menor tempo, e, além disso, possui forte integração ao ecossistema web, pois, o JavaScript utiliza JSON nativamente, para compor suas árvores de objetos.

A indexação é baseada na técnica de índice invertido. Segundo Baeza-Yates; Ribeiro Neto, 2013)

> "Um índice invertido é um mecanismo orientado a palavras para a indexação de uma coleção de texto a fim de acelerar a tarefa de busca. A estrutura do índice invertido é composta por dois elementos: o vocabulário e as ocorrências. O vocabulário é o conjunto de todas as palavras diferentes no texto. Para cada palavra do vocabulário, o índice armazena os documentos que contém esta palavra. Por essa razão, ele é chamado de índice invertido, pois podemos reconstruir o texto através do índice. Esse é o principal índice utilizado atualmente, e é também o mais antigo." (BAEZA-YATES; RIBEIRO NETO, 2013, p. 342):

Esse modelo de indexação é de alto desempenho, visto que com o mapeamento já realizado, basta que a busca verifique os vocabulários para encontrar os documentos em que há a ocorrência de determinada palavra.

Para maior eficiência da busca de informações, são definidas estratégias tanto no momento da indexação quanto no momento da busca. Fazendo analogia com o modelo cliente-servidor, as estratégias de indexação são representadas pelo lado servidor, no qual necessita de grande capacidade de processamento. E as estratégias de busca são representadas pelo lado cliente, que é a *interface* com o usuário, necessitando apenas do processamento local.

3.3.1.1 Estratégias de Indexação de Dados

Antes de executar a indexação no Elasticsearch, os dados devem ser corretamente caracterizados, de acordo com o contexto. No caso do contexto do Diário Oficial da União, a tabela 2 exemplifica quais dados e de que forma eles podem ser caracterizados.

Tabela 2 – Caracterização de dados do DOU

Nome do dado	**Descrição do dado**	**Tipo de dado**	**Observações**
artType	Tipo de publicação (ex.: portaria, extrato de contrato, etc.)	texto	Lista pré-definida
artCategory	Instituição origem da publicação	texto	Lista pré-definida
pubDate	Data da publicação	data	
texto	Conteúdo da publicação	texto	As informações contidas nesse dado são as mais relevantes para esse contexto.

Fonte Elaborado pelo autor

Como exposto, o dado "texto" é o mais importante, visto que é o conteúdo da publicação em si. Esse dado é constituído literalmente por um texto, variando conforme o tipo de publicação. No conteúdo desse texto, costuma aparecer dados como:

- CNPJ e/ou CPF;
- Nomes próprios, de pessoas físicas e/ou jurídicas;
- Conjuntos de frases;
- Número de matrículas;
- Número de processo eletrônico;
- Verbos característicos, tais como: nomear, designar, exonerar, determinar, requisitar, etc;
- Endereços;
- Entre outros.

A partir da caracterização dos dados, é possível definir uma estratégia de configuração da indexação dos dados no Elasticsearch. A estratégia é configurada com a utilização das seguintes funções do Elasticsearch:

- Mapping: função utilizada para mapear os tipos de dados, em casos em que haja alguma particularidade no dado, e também para definir um *analyzer* específico para aquele dado particular.

- Analyzer: essa função executa um processo de análise no texto e retorna um trecho do texto, chamado de *token*.
- Filter: aplicação de filtros específicos após realizado o processo do analyzer.
- Similarity: permite configurar um algoritmo para pontuar a similaridade dos dados por campo.

Para exemplificar, segue uma possibilidade de configuração para os dados de publicações do DOU:

Código 3.1 – Exemplo - Mapping

```
{
"publication": {
      "properties": {
          "artCategory": {
              "type": "keyword",
               "analyzer": "analyzer_keyword"
          }
      }
}
```

O Código 3.1 mostra um exemplo de como configurar um *mapping* em uma situação que exige um tratamento adequado. Nesse caso específico, utilizei o exemplo do dado “artCategory” que recebe um nome de uma instituição responsável pelo ato que será foi publicado. Na configuração, defini que o tipo de dado seria “keyword”, ou seja, uma palavra-chave, na qual poderia servir em filtros, agregações, etc. Além disso, defini também que esse campo seria processado com um “analyzer” próprio para palavras-chaves, de forma adequada para esse tipo de dado, deixando o dado inteiro, por exemplo, por se tratar de uma palavra-chave.

Código 3.2 – Exemplo - Analyzer e Similarity

```
{
"settings":{
    "filter": {
        "portuguese_stop": {
        "type": "stop",
        "stopwords": "_portuguese_"
        }
    },

    "analysis":{
        "analyzer":{
        "analyzer_keyword":{
```

```
13                                  "tokenizer":"keyword",
14                                  "filter":"lowercase"
15                              },
16                          "default": {
17                              "filter": [
18                                  "lowercase",
19                                  "portuguese_stop"
20                                  ],
21                              "type": "custom"
22                                      }
23                              }
24                          }
25                      }
26                  }
```

No exemplo do Código 3.2, foram configurados os *analyzers* e o algoritmo de similaridade. Na linha 12 é configurado o *analyzer* específico para palavra-chaves, conforme demonstrado no Código 3.1. Nessa configuração são utilizados os *tokenizers*, que são mecanismos para divisão de *tokens*, estes que podem ser uma única letra ou uma palavra, ou uma frase, a depender da estratégia. No caso específico das palavraschave, o objetivo é mantê-las inteiras, independente da quantidade de palavras, por isso foi utilizado o *tokenizer keyword*. Além disso, para evitar conflitos entre letras maiúsculas e minúsculas, todas as letras são transformadas em minúsculas, com o *filter lowercase*. Também foi configurado o *analyzer default*, que é executado no restante do texto. Esse analyzer utiliza um filtro chamado *stopwords*, o qual remove as pequenas e mais comuns palavras de determinado idioma. Para exemplificar a utilização do *stopwords*, a seguir um comparativo entre o texto original e o texto processado, em inglês:

- Texto original: *"The 2 QUICK Brown-Foxes jumped over the lazy dog's bone."*
- Termos processados: [*quick, brown, foxes, jumped, over, lazy, dog, s, bone*]

O mecanismo de *stopwords* remove as palavras que podem ser consideradas mais irrelevantes para um determinado contexto, dando maior eficiência e qualidade para os resultados de uma busca.

Após a caracterização dos dados e a configuração com base no contexto utilizado, os dados são submetidos para indexação no Elasticsearch. Com os dados já indexados, vem a busca, que também conta com estratégias baseadas no contexto.

3.3.1.2 Estratégias de Busca de Dados

O Elasticsearch possui uma série de possibilidades para otimizar uma busca, como, por exemplo as buscas baseadas em full-text e buscas que utilizam lógica *fuzzy*. Além disso, com o Elasticsearch, inúmeros filtros e agregações podem ser realizados para aumentar mais ainda a eficácia.

Dando continuidade ao exemplo do dado “artCategory”, será exemplificado como uma busca pode ser configurada para um dado com essas características de palavraschave. Visto que um usuário muitas das vezes digita o nome da instituição no campo de busca, para que sejam retornadas as publicações e outros dados daquela instituição específica, precisamos preparar a busca para que se concentre nesse sentido. Portanto, o Código 3.3 pode representar um exemplo de uma busca desse tipo de dado:

Código 3.3 – Exemplo de busca

```
1   GET /_search
2       {
3           "query": {
4               "match_phrase": {
5                   "artCategory": "<TERMO_BUSCADO>"
6               }
7           }
8       }
```

No Código 3.3, na linha 4, é demonstrada a função “match_phrase”. Essa função é utilizada para os casos em que é desejado que o resultado contenha necessariamente o termo que foi informado. É o caso do termo digitado ser uma instituição que se enquadre nesse dado “artCategory”. E na linha 5, entra o termo que foi informado na busca.

O Elasticsearch dispõe de diversas outras funções para otimização da busca. É de suma importância que cada dado seja corretamente caracterizado para que a estratégia de busca seja corretamente definida, resultando em uma busca muito mais eficaz.

3.3.2 Logstash

Antes da indexação dos dados pelo Elasticsearch, é necessário que se realize um procedimento de normalização dos dados, que pode ser feito com diversas ferramentas para tratamento dos dados, ou com a ferramenta padrão do ELK Stack. Trata-se do Logstash, que é um mecanismo de coleta de dados. Ele pode unificar, de forma dinâmica, dados de diferentes fontes e normalizá-los.

O Logstash foi desenvolvido originalmente para coleta de logs, porém, seus recursos possuem potencial para ir além disso. Possui três principais vantagens:

1) É um pipeline de processamento de dados escalável com alta compatibilidade com o Elasticsearch e o Kibana;
2) Mistura, combina e orquestra diferentes entradas, filtros e saídas para garantir a harmonia do pipeline;
3) Possui mais de 200 plugins disponíveis, e é flexível para expansão.

No foco deste trabalho, que é centralizar dados de inúmeras fontes diferentes no Diário Oficial, o Logstash possui um papel muito importante. Será responsável por receber, rotineiramente, grandes quantidades de dados, e realizar o tratamento e envio para a indexação do Elasticsearch.

O Logstash ganha ainda mais força para ser a ferramenta mais adequada, devido à grande diversificação de formato de dados existentes na Administração Pública. Ele conta com diversos plugins que compatibilizam a entrada e saída dos dados. Não importa se os dados de entrada são JSON, XML, banco de dados relacional ou CSV, pois, com os plugins instalados e as entradas dos dados devidamente configuradas, o Logstash cumprirá o seu papel.

3.3.3 Kibana

O Kibana é uma plataforma de análise e visualização de dados, projetado para trabalhar com o Elasticsearch. Ele é utilizado para pesquisar, visualizar e interagir com dados armazenados em índices do Elasticsearch. Com o Kibana, é possível executar facilmente análises avançadas e visualizar os dados em uma variedade de gráficos, tabelas e mapas. O Kibana facilita a compreensão de grandes volumes de dados.

A proposta de utilização do Kibana no DOU é simples. Por meio da ferramenta, todo o ambiente do Elasticsearch pode ser monitorado em tempo real. A partir disso, é possível utilizar a plataforma para análises com vistas à melhoria contínua da configuração do Elasticsearch, das estratégias de indexação e das estratégias de busca.

4 Conclusão

Como exposto, a proposta gira em torno da ideia da centralização de dados de inúmeras fontes no Diário Oficial da União. E para isso, foi sugerido a utilização de ferramentas como o conjunto ELK Stack. Não obstante, para que tudo funcione com certa fluidez e organização, foi exposta a necessidade de implantação de funções de gestão de dados, conforme o Guia DMBOK.

Apesar do escopo deste trabalho ser bastante amplo, foi possível resumir a proposta em nível teórico e exemplificativo. Executar uma proposta tão ampla, que envolve diversos órgãos, exige uma maturidade alta em Governança de TI, o que, segundo o (TCU, 2018), não é a realidade atual da Administração Pública. No entanto, com a transformação digital em alta evidência, a tendência é que a maturidade seja alcançada e projetos grandiosos sejam colocados em prática.

REFERÊNCIAS

1) BRASIL. **Constituição Federal nº 1988, de 5 de outubro de 1988**. CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE1988. Brasília, 5 out. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm Acesso em: 7 mar. 2019.
2) BRASIL. **Lei nº 12.527 de 2011, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal Brasília, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 7 mar. 2019.
3) BRASIL. **Decreto nº 4520, de 16 de dezembro de 2002**. Dispõe sobre a publicação do Diário Oficial da União e do Diário da Justiça pela Imprensa Nacional da Casa Civil da Presidência da República. Brasília, 16 dez. 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/2002/d4520.htmAcesso em: 8 mar. 2019.
4) BRASIL.. **Decreto nº 9215, de 29 de novembro de 2017**. Dispõe sobre a publicação do Diário Oficial da União. Brasília, 29 nov. 2017. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Decreto/D9215.htm. Acesso em: 8 mar. 2019.
5) BRASIL. **Decreto nº 8777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo Federal. Brasília, 11 maio 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm Acesso em: 8 mar. 2019.
6) BRASIL. **Lei Complementar nº 131, de 27 de maio de 2009**. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal. Brasília, 27 maio 2009. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp131.htm Acesso em: 8 mar. 2019.
7) BAEZA-YATES, R.; RIBEIRO-NETO, Berthier. **Recuperação de informação: conceitos e tecnologia das máquinas de busca** . 2. ed. Porto Alegre, RS: Bookman, 2013. xxiii, 590 p.
8) DAMA. **The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK)**. Brasil: Technics Publications, 01/06/2012.
9) ELASTIC. **Elasticsearch Reference**. [S. l.: s. n.], 2019. Disponível em: https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html. Acesso em: 6 abr. 2019.
10) ELASTIC. **Logstash Reference.** [S. l.: s. n.], 2019. Disponível em: https://www. elastic.co/guide/en/logstash/current/index.html. Acesso em: 6 abr. 2019.
11) ELASTIC. **Kibana Reference.** [S. l.: s. n.], 2019. Disponível em: https://www.ela stic.co/guide/en/kibana/current/index.html. Acesso em: 6 abr. 2019.
12) PRAGMATEEK. **JSON vs. XML: Some Hard Numbers About Verbosity.** [S. l.], 2013. Disponível em: https://www.codeproject.com/Articles/604720/JSON-vsXML-Some-hard-numbers-about-verbosity. Acesso em: 13 abr. 2019.
13) TCU. **Acórdão nº 588, de 21 de março de 2018**. Levantamento realizado em 581 órgãos e entidades integrantes da Administração Pública Federal, em 2017, com o objetivo de obter e sistematizar informações sobre a situação de governança pública e gestão de tecnologia da informação (TI), contratações, pessoas e resultados. Brasília, 21 mar. 2018. Disponível em: https://pesquisa.ap ps.tcu.gov.br. Acesso em: 13 abr. 2019.