



GUIA DE SOBREVIVÊNCIA: GOVERNANÇA DE DADOS

SUMÁRIO EXECUTIVO	4
O QUE É GOVERNANÇA DE DADOS	5
PRINCIPAIS PILARES	6
<i>Qualidade de dados</i>	7
<i>Gerenciamento de dados mestre</i>	8
<i>Dicionário de dados</i>	8
<i>Segurança da informação</i>	8
CICLO DE GOVERNANÇA DE DADOS	9
BENEFÍCIOS DA GOVERNANÇA DE DADOS	10
<i>Decisões estratégicas com dados confiáveis</i>	10
<i>Eficiência operacional</i>	10
<i>Melhor conhecimento dos dados e sua linhagem</i>	11
<i>Maior qualidade de dados</i>	11
<i>Conformidade regulatória (compliance)</i>	11
<i>Receita Expandida</i>	11
GDPR E LGPD EM GOVERNANÇA DE DADOS	12
FRAMEWORK ATRA	13
FASE 1 – ASSESSMENT	15
<i>Levantamento</i>	15
<i>Entrevistas</i>	15
<i>Maturidade</i>	15
<i>Fontes de dados</i>	17
<i>Sistemas e gaps de tecnologia</i>	17
FASE 2 – OPERACIONAL	18
<i>Modelos Operacionais</i>	18
<i>Papéis e Responsabilidades</i>	23
<i>Políticas</i>	25
<i>Processos</i>	27
<i>Dados Mestre</i>	30
<i>Glossário de Negócios</i>	31
<i>Linhagem de Dados</i>	31
<i>Segurança da Informação</i>	32
FASE 3 – QUALIDADE DE DADOS	33
<i>Como classificar a importância de um dado?</i>	34
<i>Dimensões da qualidade de dados</i>	34
<i>Análise de dados</i>	36
<i>Definição de regras de padronização e higienização</i>	40
<i>Definição de regras de enriquecimento</i>	42
<i>Definição de regras de deduplicação</i>	42
FASE 4 – ROLL OUT	44
<i>Plano de comunicação</i>	44
<i>Treinamentos</i>	44

FIGURAS

<i>FIGURA 1: GOVERNANÇA DE DADOS</i>	5
<i>FIGURA 2: A ORQUESTRAÇÃO DE GOVERNANÇA DE DADOS</i>	5
<i>FIGURA 3: Os PILARES DA GOVERNANÇA DE DADOS</i>	6
<i>FIGURA 4: DIFERENÇA ENTRE GOVERNANÇA E QUALIDADE DE DADOS</i>	7
<i>FIGURA 5: EXEMPLOS DE DADOS E SUAS CLASSIFICAÇÕES</i>	9
<i>FIGURA 6: VISÃO CÍCLICA DA GOVERNANÇA DE DADOS</i>	10
<i>FIGURA 7: REGULAMENTAÇÕES E SUAS CONSEQUÊNCIAS</i>	13
<i>FIGURA 8: FRAMEWORK DE IMPLEMENTAÇÃO DE GOVERNANÇA DE DADOS</i>	14
<i>FIGURA 9: NÍVEIS DE MATURIDADE EM GOVERNANÇA DE DADOS</i>	16
<i>FIGURA 10: MODELO OPERACIONAL CENTRALIZADO</i>	19
<i>FIGURA 11: MODELO OPERACIONAL DESCENTRALIZADO</i>	20
<i>FIGURA 12: MODELO OPERACIONAL HÍBRIDO</i>	21
<i>FIGURA 13: MODELO OPERACIONAL FEDERADO</i>	22
<i>FIGURA 14: PIRÂMIDE DE PAPÉIS DENTRO DA GOVERNANÇA DE DADOS</i>	23
<i>FIGURA 15: TIPOS DE PROCESSOS</i>	27
<i>FIGURA 16: CICLO DE PROCESSOS</i>	28
<i>FIGURA 17: MATURIDADE E AS DIMENSÕES DE QUALIDADE DE DADOS</i>	35
<i>FIGURA 18: CICLO DE QUALIDADE DE DADOS</i>	36
<i>FIGURA 19: EXEMPLO DE PROFILE CRIADO COM A FERRAMENTA ANALYST DA INFORMATICA</i>	37
<i>FIGURA 20: DATA PREVIEW (DRILLDOWN) DA FERRAMENTA ANALYST DA INFORMATICA</i>	38
<i>FIGURA 21: EXEMPLO DE SCORECARD COM DIMENSÃO DE QUALIDADE</i>	39
<i>FIGURA 22: APLICAÇÃO DE REGRA DE PARSE E PADRONIZAÇÃO</i>	41
<i>FIGURA 23: APLICAÇÃO DA REGRA DE PARSE E PADRONIZAÇÃO PARA ENDEREÇO</i>	41

TABELAS

<i>TABELA 1: DEFINIÇÕES DE PAPÉIS E RESPONSABILIDADES</i>	25
<i>TABELA 2: EXEMPLO DE REGRA DE ENRIQUECIMENTO</i>	42
<i>TABELA 3: EXEMPLO DE DEFINIÇÃO DE MATCH</i>	43
<i>TABELA 4: EXEMPLO DE DEFINIÇÃO DE MERGE</i>	43

Este documento destina-se aqueles que têm interesse pela atividade de governança de dados, desde sua concepção, planejamento e execução. Visa esclarecer os aspectos funcionais de como realizar levantamentos, análises de maturidade, definição de modelo operacional de governança de dados, papéis e responsabilidades, políticas e processos.

- **OPORTUNIDADE:** Resolver problemas relacionados à dados e estar à frente de legislações que afetem os dados
- **MISSÃO:** Prover o melhor serviço de implementação e operacionalização de governança de dados
- **NOSSA SOLUÇÃO:** Framework flexível e adaptável às necessidades dos clientes, baseado nas melhores metodologias do mercado (DMBoK, DGI framework, DGPO, Informatica)



Figura 1: Governança de Dados

A governança de dados é responsável pelo funcionamento e monitoramento dos fluxos dos dados em uma empresa. É uma atividade que foca os esforços de várias áreas na atenção à qualidade, segurança e controles dos dados.

A governança de dados é a orquestração de pessoas, políticas, processos e tecnologia para que a empresa possa assegurar que os dados servem aos propósitos necessários ao cumprimento da estratégia, assim como prevenção a multas e processos legais que possam ocorrer devido ao manuseio errado dos dados.

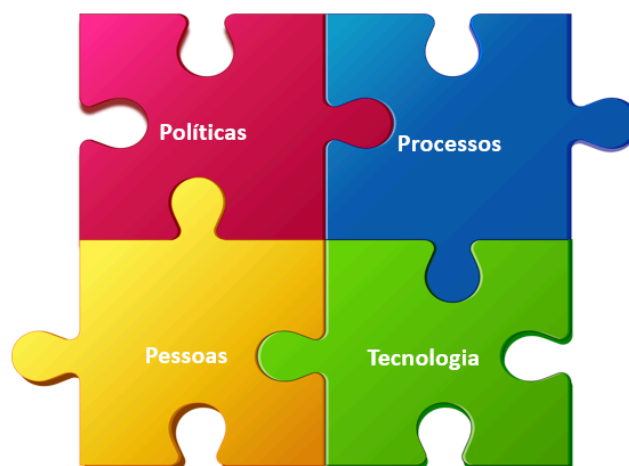


Figura 2: A orquestração de Governança de Dados

A governança de dados usa como base os fatores externos, como leis e regulamentações, mas também fatores internos, como políticas, normas e processos. Outra base importante na governança de dados são as pessoas e as tecnologias aplicadas.

PRINCIPAIS PILARES

Os principais pilares que sustentam a governança de dados, no nosso framework, são:

- **Qualidade de dados:** um dos principais motivadores das empresas para implementação de GD
- **Gerenciamento de dados mestre:** definição e gerenciamento dos dados corporativos
- **Dicionário de dados:** criação e gerenciamento de dicionário de dados, glossário de negócios e linhagem dos dados
- **Segurança da Informação:** definições e classificação de importância dos dados, monitoração de acesso e uso dos dados (pessoas e sistemas)

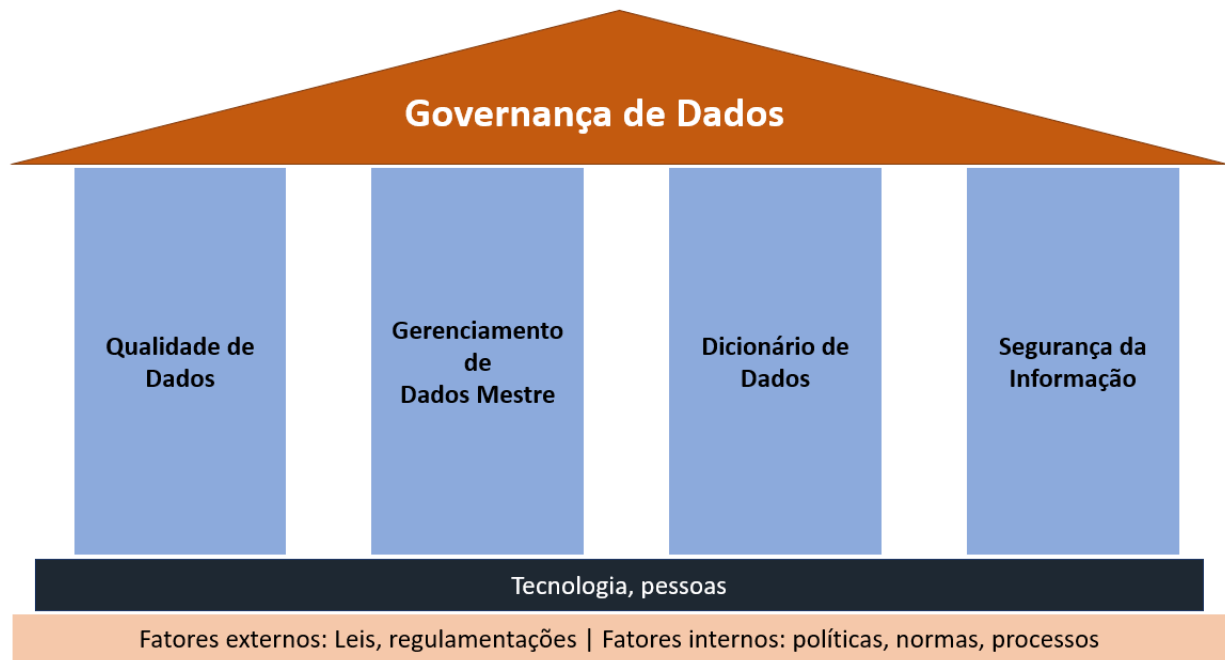


Figura 3: Os Pilares da Governança de Dados

A qualidade dos dados é um dos principais benefícios da implementação de governança de dados. A qualidade dos dados está intrinsecamente ligada a relatórios confiáveis, que por sua vez geram tomadas de decisões informadas e precisas, baseada em informações corretas.

Qual a diferença entre governança e qualidade?

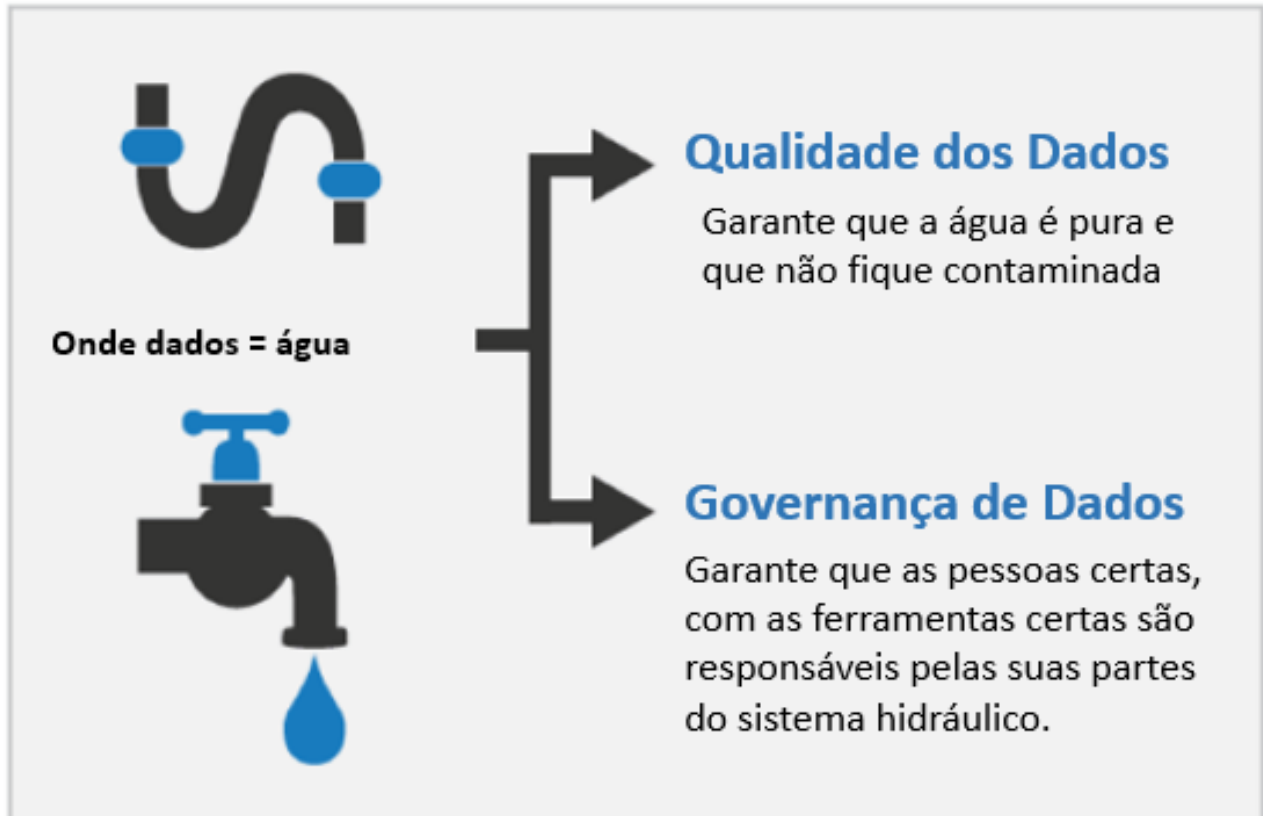


Figura 4: Diferença entre Governança e Qualidade de Dados

No DAMA Dictionary of Data Management, esses termos são definidos da seguinte maneira:

Qualidade de dados - o grau em que os dados são precisos, completos, oportunos e consistentes com todos os requisitos e regras de negócios.

Governança de dados - O exercício de autoridade, controle e tomada de decisão compartilhada (por exemplo, planejamento, monitoramento e execução) sobre o gerenciamento de ativos de dados.

GERENCIAMENTO DE DADOS MESTRE

Compreende desde a definição dos dados corporativos (dados mestre, dados não transacionais), que são os dados usados por toda a empresa, até a sua manutenção e verificação periódica.

Os dados mestre podem estar em um sistema de gerenciamento, ou pode ser somente o conhecimento sobre eles e onde estão localizados. Geralmente, os dados mestre tendem a estar em um sistema apartado que faz a consistência dos dados de vários sistemas legados, elegendo o melhor dado para ser usado por toda a empresa.

A partir de sistemas de MDM (Master Data Management), os dados passam por duplicações, padronizações e correções, até chegar ao estado requerido pela empresa.

DICIONÁRIO DE DADOS

Também chamado de Glossário de Negócio, Business wiki ou Business Glossary, o dicionário de dados é uma ferramenta importantíssima na governança de dados.

Um dicionário de dados contém todas as definições mutuamente acordadas de termos de negócio, livrando os dados de definições incorretas e usos ambíguos dos dados em relatórios ou sistemas.

SEGURANÇA DA INFORMAÇÃO

Um item muito importante em tempos de legislações de proteção aos dados, já que ajuda as empresas a se prevenir contra multas pesadas quando há violações nos sistemas.

O item mais importante para a governança de dados dentro da segurança da informação, é a classificação dos dados. A classificação dos dados identifica os níveis de proteção que os dados devem ter. As empresas podem adotar vários níveis de classificação, mas geralmente, os dados são classificados em:

1. **Público:** Os dados devem ser classificados como públicos quando a divulgação, alteração ou destruição não autorizada desses dados resultarem em pouco ou nenhum risco para a empresa. Exemplos de dados públicos incluem comunicados de imprensa, informações sobre cursos e publicações de pesquisa. Embora pouco ou nenhum controle seja necessário para proteger a confidencialidade dos dados públicos, é necessário algum nível de controle para impedir a modificação não autorizada ou a destruição desses dados
2. **Confidencial:** Os dados devem ser classificados como confidenciais quando a divulgação, alteração ou destruição não autorizada desses dados pode resultar em um nível moderado de risco para a empresa. Por padrão, todos os dados corporativos que não são classificados como dados restritos ou públicos devem ser tratados como dados confidenciais. Um nível razoável de controle de segurança deve ser aplicado aos dados confidenciais.
3. **Restrito:** Os dados devem ser classificados como restritos quando a divulgação, alteração ou destruição não autorizada desses dados pode causar um nível significativo de risco para a empresa. Exemplos de dados restritos incluem dados protegidos por regulamentações de privacidade nacionais ou internacionais

e dados protegidos por acordos de confidencialidade. O nível mais alto de controle de segurança deve ser aplicado a dados restritos.



Figura 5: Exemplos de dados e suas classificações

CICLO DE GOVERNANÇA DE DADOS

Governança de dados não pode ser vista como um projeto. A governança de dados é um processo cíclico e contínuo. Sem pessoas capacitadas e o entendimento que o trabalho precisa ser sempre revisitado, a atividade não gera os frutos que as empresas querem e sempre culpam a governança de dados por ser ineficiente.

Vamos usar como exemplo uma gestação:

- Uma mulher engravida. Durante nove meses ela planeja o que virá pela frente e o bebê cresce (matura) dentro dela. Essa é a fase de levantamento de requerimentos para implementação de governança de dados.
- Após nove meses nasce o bebê. Ela tem que aprender a dar banho, dar comida, limpar o bumbum, colocar para dormir. A nova mamãe está aprendendo como lidar com o bebê. Essa seria a fase de implementação.
- Depois de alguns meses a mamãe já está bem treinada. Consegue lidar com tudo que aprendeu e continua aprendendo mais coisas. Essa é a fase de operacionalização.

Agora imagine que essa mamãe pense: “hoje não preciso dar comida para o bebê”, ou “vou deixar de dar banho porque ele se suja o tempo todo”, o que pode acontecer com o bebê? É o mesmo caso quando terminada a operacionalização, as pessoas envolvidas no processo de governança de dados não estejam cientes da importância de suas tarefas e o quanto isso afeta as verificações e monitoramento cíclicos que deveriam estar sempre presentes em uma atividade de governança de dados.

A governança de dados é um processo cíclico que precisa ser sempre examinado para definir prioridades, descobrir falhas e melhorar processos.

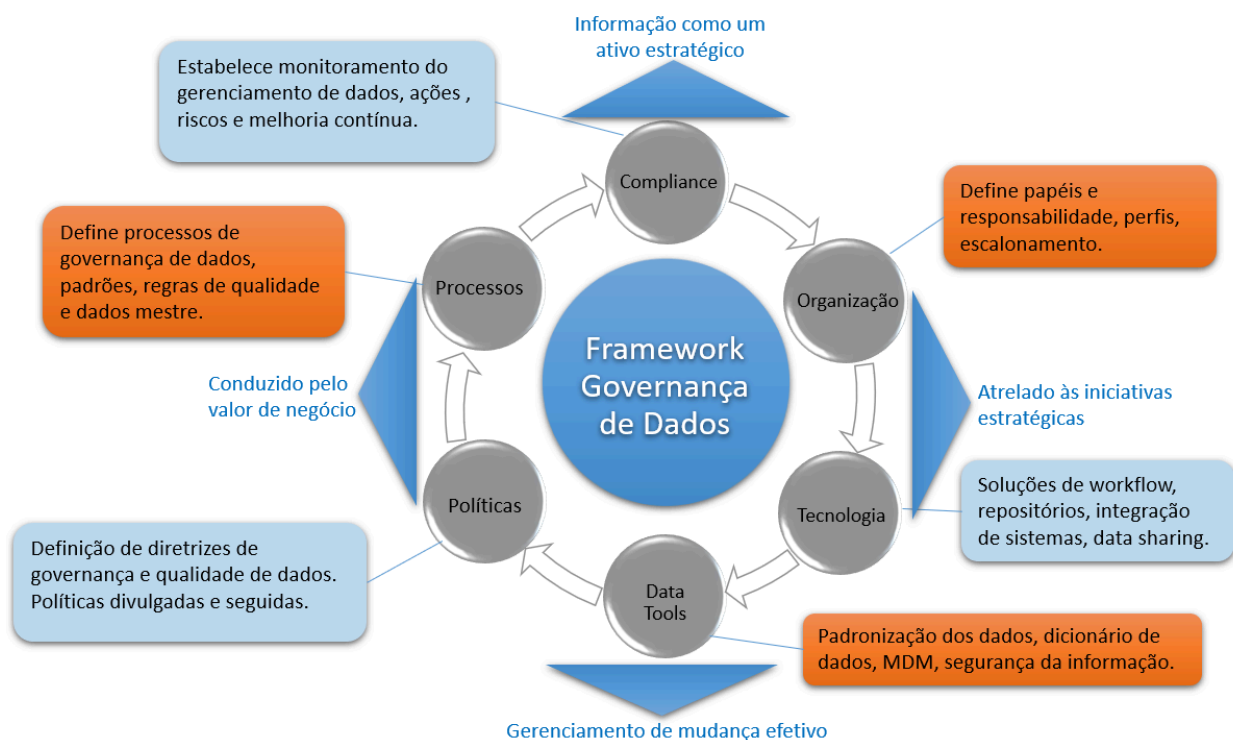


Figura 6: Visão cíclica da governança de dados

BENEFÍCIOS DA GOVERNANÇA DE DADOS

A governança de dados, aplicada corretamente, com os devidos controles, traz diversos benefícios para as empresas:

DECISÕES ESTRATÉGICAS COM DADOS CONFIÁVEIS

Um dos principais benefícios da governança de dados é a melhor tomada de decisão. Isso se aplica tanto ao processo de tomada de decisão quanto às próprias decisões.

Dados bem governados são mais detectáveis, tornando mais fácil para as partes relevantes encontrar informações úteis. Isso também significa que as decisões serão baseadas nos dados corretos, garantindo maior precisão e confiança.

EFICIÊNCIA OPERACIONAL

Os dados são muito valiosos na era dos negócios baseados em dados. Portanto, devem ser tratados como o ativo que é.

Considere os ativos físicos de uma empresa de manufatura, por exemplo. Empresas de manufatura bem gerenciadas garantem que suas máquinas de linha de produção sejam submetidas a inspeções, manutenção e upgrades regulares, para que a linha funcione sem problemas e com tempo de inatividade limitado. A mesma abordagem deve se aplicar aos dados.

MELHOR CONHECIMENTO DOS DADOS E SUA LINHAGEM

A governança de dados significa entender quais são seus dados e onde eles estão armazenados. Quando bem implementada, a governança de dados fornece uma visão abrangente de todos os ativos de dados.

Conhecer os dados e sua linhagem nos dá a oportunidade de conhecer os responsáveis por eles. Ao atribuir permissões, por exemplo, é muito mais fácil determinar quem é responsável por dados específicos.

MAIOR QUALIDADE DE DADOS

À medida que a governança de dados ajuda na descoberta, empresas com programas de governança de dados eficazes também se beneficiam da melhoria da qualidade dos dados, já que a qualidade de dados é um dos grandes pilares de governança.

Os dados podem ser padronizados, com consistência. A qualidade de dados quer saber como os dados são úteis e completos, enquanto a governança de dados quer saber onde os dados estão e quem é responsável por eles.

A governança de dados melhora a qualidade dos dados, porque responder a estes últimos torna mais fácil lidar com os primeiros.

CONFORMIDADE REGULATÓRIA (COMPLIANCE)

Se a sua empresa ainda não adotou um programa de governança de dados, a conformidade regulatória talvez seja o melhor motivo para isso. Multas pesadas com um limite superior de 20 milhões de euros, ou 4%, ou volume de negócios global anual - o que for maior - não são brincadeira.

Dito isto, as multas do GDPR – e a partir de 2020 LGPD – apenas incentivam algo que você já deveria estar interessado em fazer. As empresas orientadas por dados que não estão desfrutando dos benefícios mencionados estão fundamentalmente reprimindo o seu próprio desempenho.

Podemos até dizer que para uma empresa ser realmente orientada por dados, a governança de dados é uma obrigação.

RECEITA EXPANDIDA

O aumento da receita deveria, de fato, estar no topo da lista de benefícios da governança de dados. No entanto, está posicionado aqui porque os benefícios mencionados influenciam a receita cumulativamente.

Todos os benefícios da governança de dados ajudam as empresas a tomar decisões melhores e mais rápidas com mais certeza.

Isso significa que erros menos onerosos - dados ruins e até violações de dados - são feitos. Significa que você gasta menos dinheiro gerenciando riscos e fechando as brechas mais vulneráveis na segurança de sua empresa, em vez de mais dinheiro adiante, lidando com as consequências de perda de reputação e perda financeiras custosas.

As legislações de proteção aos dados estão ficando cada vez mais onerosas nas suas multas e muitas das condições de uso dos dados estão sendo controladas por essas legislações.

A mais recente é a General Data Protection Regulation, GDPR, que entrou em vigor em maio de 2018 nos países da comunidade europeia. A GDPR afeta empresas no mundo todo, quando processa dados de cidadãos europeus. As diretrizes gerais são:

- **Jurisdição do regulamento:** Uma vez entrando em vigor, o GDPR passará a valer para todas as empresas que possuem dados pessoais de cidadãos residentes em países da União Europeia, independentemente da localização da empresa.
- **Multa:** A penalidade para empresas que não cumprirem com o regulamento pode chegar a 4% da receita global anual da empresa ou 20 milhões de euros – o que for maior. Esse valor máximo ultrapassa os 80 milhões de reais e é imposto a empresas que violarem os principais pontos do regulamento, como não pedir o consentimento ou violar os princípios de privacidade desde a concepção (ver abaixo).
- **Consentimento expresso:** O consentimento deve ser expresso. As empresas não poderão mais utilizar letras miúdas nem omitir o texto legal. Isso significa que a pessoa precisa estar ciente de que seus dados serão captados ao realizar a conversão – e que ele precisa aceitar ceder esses dados de forma explícita. Um checkbox visível com texto de fácil legibilidade ou um double opt-in, que é quando o cliente recebe um e-mail explicando que foi adicionado à base e precisa confirmar seu aceite, são o suficiente. É preciso ser fácil para aceitar ceder seus dados quanto negar o acesso a eles, ou retirar esse acesso, uma vez dado.
- **Direito de acesso:** Direito de saber se qualquer dado pessoal seu está sendo processado na base de uma empresa, onde esse dado está sendo processado e para que fim será utilizado. Ele também tem direito a acessar todo e qualquer dado que a empresa detenha sobre ele em seu sistema, que deve ser entregue a ele em formato eletrônico e sem custo algum.
- **Portabilidade de dados:** Além de ter direito a acessar e reivindicar seus dados pessoais, o indivíduo ganha o direito de fazer a portabilidade desses dados – isso é, mover esses dados para outro sistema que não seja o da empresa que os captou, sem perder a informação.
- **Direito de apagar dados:** Direito de ter seus dados apagados definitivamente da sua base, caso assim deseje. Ele também tem direito de impedir que a empresa continue disseminando esses dados e que os dados sejam processados por terceiros.
- **Notificação de violação do sistema:** Se o sistema da empresa, por qualquer motivo, for violado, e os dados forem roubados, expostos ou se tornarem vulneráveis, a empresa fica obrigada a notificar os indivíduos cujas informações foram afetadas em até 72h após descoberta da violação. Isso é válido tanto para processadores quanto para controladores de dados, sob risco de multa.
- **Privacidade desde a concepção:** O conceito de privacidade desde a concepção estabelece que o desenvolvimento de um sistema e as práticas de negócio para levá-lo ao mercado devem ser norteadas pelos conceitos de proteção de dados e privacidade de seus usuários. Isso significa que ao criar novos produtos ou desenvolver novas funcionalidades, é preciso ter sempre em mente os pontos colocados no GDPR.
- **Responsável por proteção dos dados (Data Protection Officer):** Se a empresa conduz atividades que requerem monitoramento regular de dados pessoais em larga escala, ela precisa ter um profissional responsável pela proteção desses dados, alguém que tenha familiaridade com normas e boas práticas.

A legislação brasileira, Lei Geral de Proteção aos Dados (LGPD), contém basicamente os mesmos princípios da GDPR, portanto, estando em compliance com a GDPR que é mais compreensiva, pode-se assumir que também estará em compliance com a LGPD.

A governança de dados é fundamental para que o compliance com essas legislações se realize de forma efetiva.



Figura 7: Regulamentações e suas consequências

“Conformidade com a GDPR e governança de dados estão inextricavelmente ligados; a conformidade com a GDPR não pode ser alcançada sem uma forte governança.”

- Bart Vandekerckhove, Collibra

FRAMEWORK ATRA

O framework desenvolvido leva em conta as melhores metodologias de mercado, como o framework da DAMA, DGI, Informatica, e outras fontes renomadas sobre o assunto, como o DataVersity e DGPO.

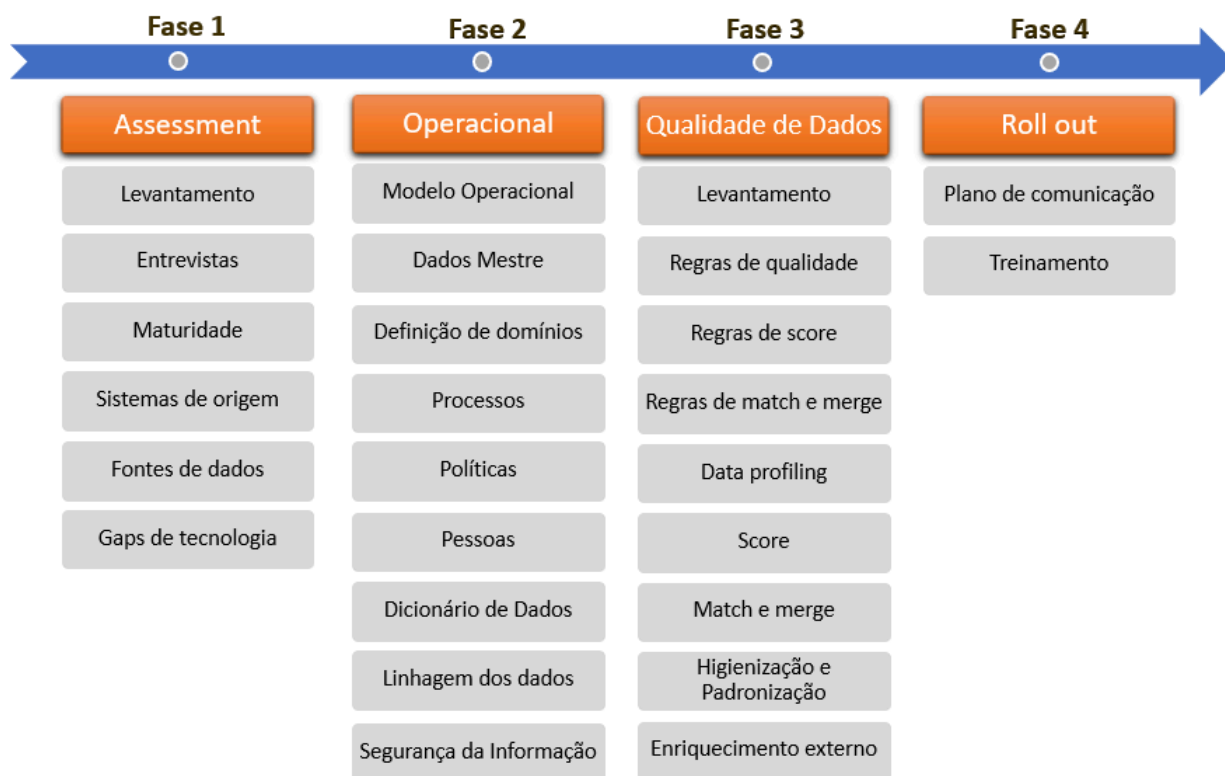


Figura 8: Framework de Implementação de Governança de Dados

O framework foi criado para poder ser implementado por fases, cada fase servindo o propósito final de implementação de um programa/escritório de governança de dados em empresas.

A fase de assessment é a fase inicial do framework, e por onde recomendamos que todo o processo se inicie. Aqui teremos levantamento de processos, políticas e ciclo de dados, entrevistas com as áreas relevantes, levantamento de maturidade por área e a média da empresa, verificação dos sistemas utilizados, quais as fontes de dados e quais os gaps de tecnologia.

A fase de assessment tem como grande entregável o documento de assessment, que contém o que foi encontrado, gaps e o que pode ser melhorado, assim como um roadmap e a análise final de maturidade. Esse documento é o guia e balizador para o trabalho a ser realizado na próxima fase, a Operacional.

LEVANTAMENTO

Nesta etapa é feita a análise das documentações disponibilizadas pelo cliente, processos, políticas, normas, regras, legislações relevantes ao negócio, organograma, papéis e responsabilidades e matrizes RACI onde disponível.

É feita uma verificação crítica dos gaps e das necessidades que serão geradas para que se tenha uma governança de dados efetiva.

ENTREVISTAS

A quantidade de áreas e entrevistas depende de um escopo definido, porém, geralmente, as entrevistas têm que acontecer com áreas de TI e as principais áreas de negócio que criam, alteram, removem, ou consomem os dados corporativos.

É importante salientar que entrevistas somente com TI não geram o valor necessário para criar um assessment correto da situação da empresa, pois TI não é responsável pelos dados, mas sim pelos sistemas que contém os dados.

MATURIDADE

Nenhuma organização começa a implementar um programa de governança de dados do zero; cada empresa provavelmente já tem algum nível de conhecimento em certas áreas. Determinar o nível de maturidade no assunto é um passo necessário no desenvolvimento de um plano de governança de dados.

Usamos o modelo de maturidade mencionado no framework da Informatica, que se baseia na abordagem de melhoria de processos do Capability Maturity Model Integration (CMMI) da Carnegie Mellon University para definir seis grandes estágios de maturidade.

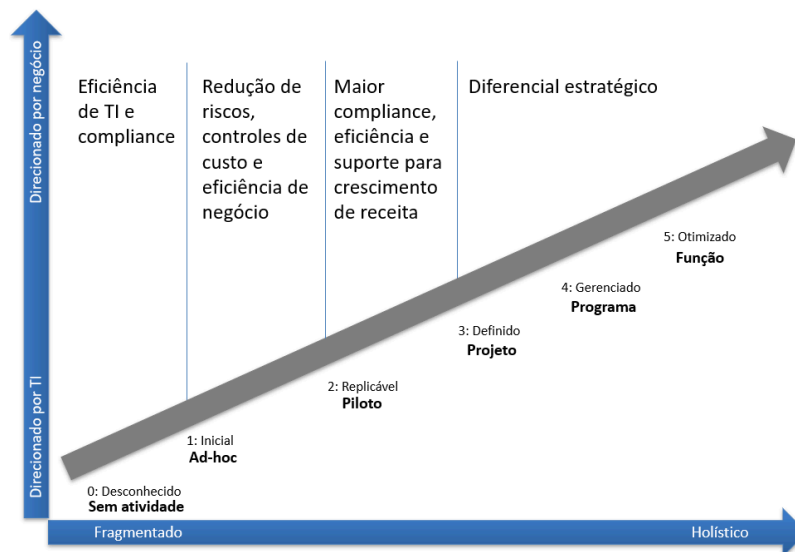


Figura 9: Níveis de Maturidade em Governança de Dados

0: DESCONHECIDO – SEM ATIVIDADE

Quando existe um foco mínimo em segurança ou qualidade de dados, os dados não são priorizados de nenhuma maneira significativa que possa gerar ações. Não há indicadores sobre governança de dados e a governança de dados não é gerenciada.

1: INICIAL – AD HOC

A governança de dados é basicamente dirigida por algumas pessoas. As regras são implementadas de forma ad hoc, com padrões como parte dos requisitos funcionais em um projeto de TI, que só é considerado bem sucedido se o projeto de TI for bem sucedido.

2: REPLICÁVEL – PILOTO

A governança de dados ainda é básica, mas ganha atenção no nível de gerência de TI. Existem padrões de governança e TI documentados que direcionam a reutilização de metadados e a colaboração em projetos de TI. O sucesso é medido principalmente em melhorias de TI. Geralmente, é gerenciado por meio de um projeto piloto.

3: DEFINIDO – PROJETO

Conduzido principalmente por TI através da adoção de centros de competência e centros de excelência. A liderança do projeto é principalmente através da TI, mas há envolvimento do negócio. O sucesso é medido em métricas operacionais em nível de projeto.

4: GERENCIADO – PROGRAMA

A Governança de Dados é patrocinada por líderes da empresa, iniciado como parte de um programa mais amplo de gerenciamento estratégico de dados empresariais. A Governança de Dados viverá através de esforços multi-fases, plurianuais, mas medidos com base no sucesso do programa.

5: OTIMIZADO – FUNÇÃO

Há patrocínio e apoio executivo. A governança de dados é adotada como uma função empresarial core e autossustentável que gerencia dados como um ativo corporativo. O sucesso é medido no impacto total para o negócio, não apenas confinado a programas ou estratégias específicas.

FONTES DE DADOS

Ao realizar as entrevistas esse item deve ser incluído como um tópico de levantamento. Ao realizar as perguntas sobre os ciclos dos dados, inevitavelmente deve-se perguntar qual a fonte desses dados. O conhecimento sobre as fontes de dados é essencial para desenvolver processos no futuro que garantam que essas fontes tem todas as travas necessárias, que os dados dessas fontes são confiáveis, atuais e padronizados.

As fontes podem ser externas, tais como website, formulários, compra de dados ou internas, que podem incluir inclusão de clientes e fornecedores manualmente, assim como sistemas de RH.

SISTEMAS E GAPS DE TECNOLOGIA

A análise sistêmica e de gaps de tecnologia será realizada posteriormente às entrevistas, já que nessa fase já teremos os insumos das entrevistas para nortear as necessidades de mudanças ou inexistência de ferramentas levantadas pelas áreas de negócio e TI.

Geralmente realizada por um especialista em sistemas (para análise de capacity, etc.), esse é um passo importante para podermos ter uma fase de implementação de tecnologia, com a automatização de aplicação das regras de qualidade, da fase Qualidade de dados do framework.

A fase operacional visa implementar as necessidades levantadas na fase de assessment, assim como definir modelo operacional, políticas e novos processos.

Essa é uma etapa de extrema importância pois será o principal passo em direção à uma cultura de conhecimento dos dados.

MODELOS OPERACIONAIS

O modelo operacional de governança de dados é fundamental para dar início ao programa. Sem a definição do modelo, fica impossível definir papéis e responsabilidades, políticas e processos, pois todos requerem a direção que vem da definição de qual modelo será utilizado pela empresa.

O modelo operacional é a estratégia que será usada para conduzir a governança de dados. Na definição do modelo operacional também é feita a análise do organograma e, dependendo do modelo escolhido, onde a área de governança de dados poderia ser colocada.

Os modelos podem ser Centralizados, Descentralizados, Híbridos e Federados. Vamos entrar no detalhe dos tipos e quais os seus pros e contras.

MODELO CENTRALIZADO

Neste modelo, como em um projeto, existe um sponsor executivo e tudo é gerido pelo escritório de governança de dados e as pessoas envolvidas na governança e gerenciamento dos dados reportam diretamente ao líder de governança de dados.

A regra básica para determinar o melhor sponsor é identificar onde há força política além de uma posição (como CFO ou COO), porque o envolvimento e o suporte entre áreas são fundamentais para o sucesso nessa função.

PRÓS DO MODELO CENTRALIZADO

- Estabelece uma posição executiva formal da governança de dados
- O Steering reporta diretamente ao executivo
- A existência de um líder de governança de dados significa que há uma pessoa no topo; que torna a tomada de decisão mais fácil
- “One-stop shop” para todas as necessidades de dados
- Mais fácil de gerenciar por tipo de dados

CONTRAS DO MODELO CENTRALIZADO

- Ao implementar, geralmente há um grande impacto organizacional
- Novos papéis/funções/cargos provavelmente exigirão aprovação de Recursos Humanos
- Continua existindo a separação formal de funções comerciais e de arquitetura técnica

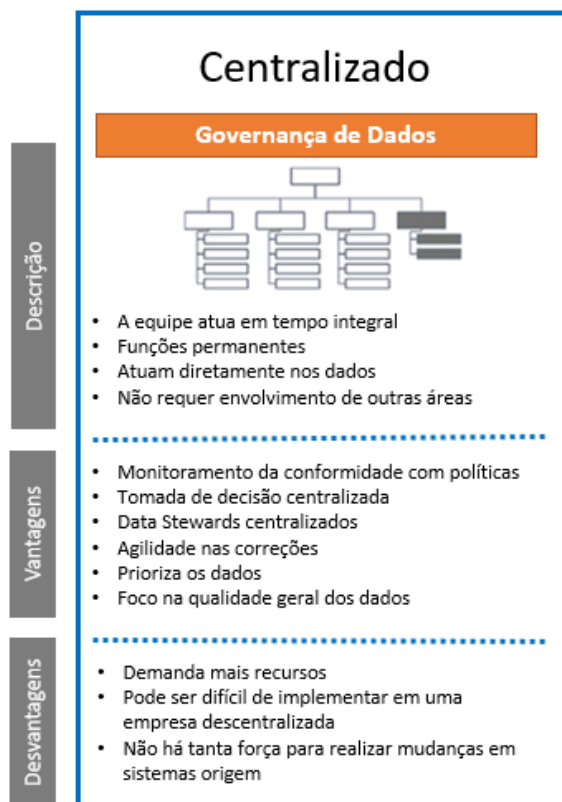


Figura 10: Modelo Operacional Centralizado

MODELO DESCENTRALIZADO

O modelo descentralizado é o oposto do centralizado: aqui tudo é resolvido em conselhos/comitês, então não há um líder de governança de dados. Geralmente, esse é o modelo encontrado nas empresas que começaram a falar sobre dados recentemente, portanto tem o básico para gerir dados.

PRÓS DO MODELO DESCENTRALIZADO

- Estrutura relativamente plana
- Conversas informais sobre Governança de Dados
- Relativamente rápido para estabelecer e implementar

CONTRAS DO MODELO DESCENTRALIZADO

- Discussões consensuais tendem a demorar mais do que os decretos centralizados
- Muitos participantes comprometem os órgãos de governança (tornando-os potencialmente indisciplinados)
- Pode ser difícil de manter com o tempo, devido à sua informalidade
- Tem um valor menor, pois tende a ser menos produtivo
- Coordenação muito difícil
- Nada muda nas áreas de negócio
- Falta de definição sobre coproprietários de dados e suas responsabilidades

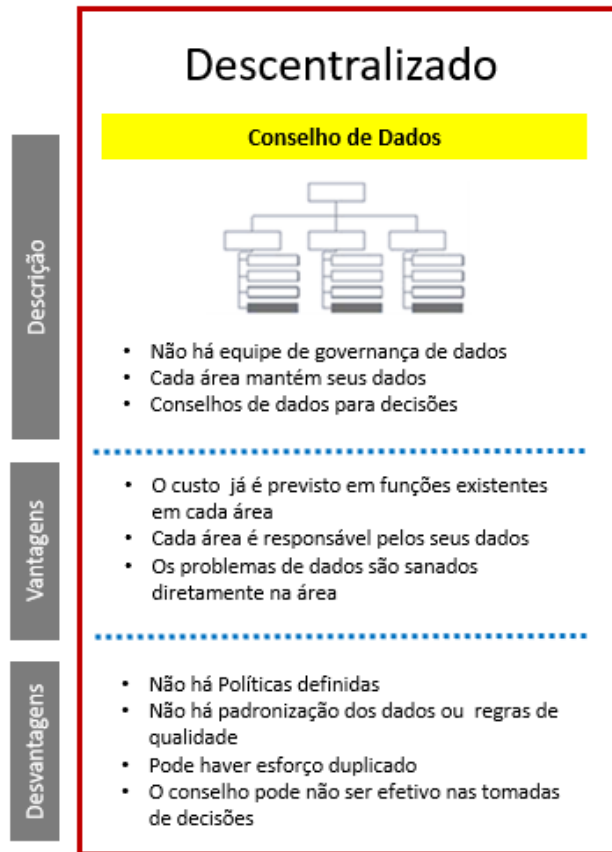


Figura 11: Modelo Operacional Descentralizado

MODELO HÍBRIDO

O modelo híbrido, também conhecido como mediano, como o próprio nome já indica é um híbrido entre o modelo centralizado e o descentralizado.

PRÓS DO MODELO HÍBRIDO

- Estrutura centralizada para estabelecer direção
- Função formal de liderança pelo escritório de governança de dados servindo como ponto único de contato e responsabilidade
- A posição de liderança da governança de dados é um papel dedicado em período integral, portanto a governança de dados recebe a atenção que merece
- Grupos de trabalho têm ampla participação para facilitar a colaboração e a decisão por consenso
- Potencialmente um modelo mais fácil de implementar inicialmente e manter ao longo do tempo
- Força a tomada de decisão pelas áreas
- Concentração em dados específicos
- Resolução de problemas sem necessidade de envolver toda a equipe

CONTRAS DO MODELO HÍBRIDO

- A posição de governança de dados é uma função dedicada em tempo integral, portanto pode exigir um novo número de funcionários

- A dinâmica do grupo de trabalho pode exigir a priorização de requerimentos de negócio conflitantes, pode não haver consenso/decisão
- Muitas camadas (áreas de negócio envolvidas)

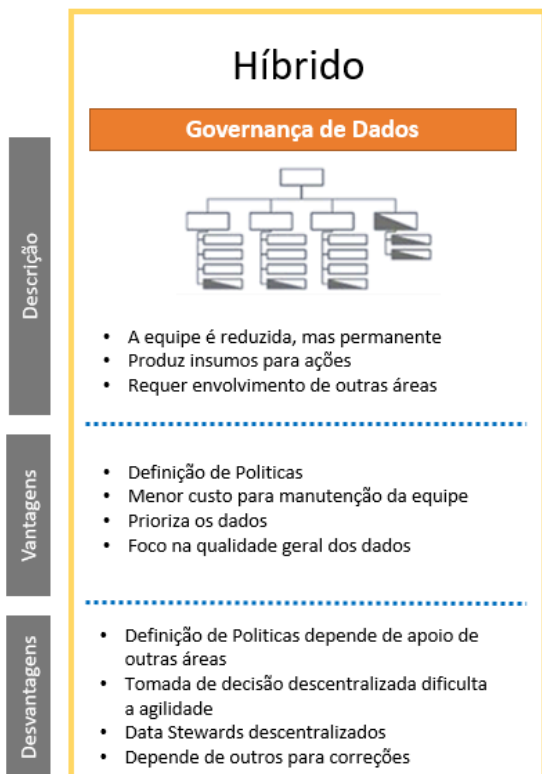


Figura 12: Modelo Operacional Híbrido

MODELO FEDERADO

Esse modelo é muito parecido com o híbrido, mas geralmente tem mais camadas (áreas/países) envolvidas. É utilizado para empresas com presença global (multinacionais ou multi-negócios) mas com uma tratativa única de governança de dados.

PRÓS DO MODELO FEDERADO

- Estratégia empresarial centralizada com execução e implementação descentralizadas
- Função formal de liderança pelo escritório de governança de dados servindo como ponto único de contato e responsabilidade
- Práticas de governança de dados realizadas por cada grupo/filial/negócio dentro de cada área de negócio para dar autonomia às divisões com requisitos diferentes
- Concentração em dados específicos, desafios de cada divisão ou prioridades regionais
- Força a tomada de decisão pelas áreas
- Concentração em dados específicos
- Resolução de problemas sem necessidade de envolver toda a equipe

CONTRAS DO MODELO FEDERADO

- Muitas camadas (áreas de negócio envolvidas)

- A autonomia no nível de empresa/área pode ser um desafio para coordenar
- Dífícil de encontrar o equilíbrio entre as prioridades da e as prioridades da empresa



Figura 13: Modelo Operacional Federado

PAPÉIS E RESPONSABILIDADES

A definição de papéis e responsabilidades dentro da estrutura de governança de dados é fundamental e necessário logo no início da implementação, pois assim que o modelo operacional está definido, é necessário saber quem será responsável por cada parte do processo.

Geralmente, as funções de governança de dados se espalham em diversos níveis organizacionais, como vemos abaixo:



Figura 14: Pirâmide de Papéis dentro da Governança de Dados

Existem vários nomes para as funções, aliás, podemos nomear do jeito que for mais conveniente para a empresa, porém, os nomes mais comuns para as funções estão descritos abaixo.

Papel	O que é?	Principais Responsabilidades
Data Steward	O data steward é o indivíduo que atua em nome do data owner para garantir que os dados sejam utilizados adequadamente e para garantir que as políticas e procedimentos sejam definidas, implementadas e cumpridas. É o ponto de contato principal dentro de cada área.	<ul style="list-style-type: none"> Garantir a conformidade com as políticas e as normas de governança dos dados. Aconselhar sobre o conceito e utilização dos dados. Auxiliar na definição e identificação de regras de qualidade de dados. Atuar em planos de ação que envolvam melhoria dos dados.
Data Owner	O data owner é responsável por definir e garantir que as medidas adequadas sejam promovidas para melhorar a qualidade e integridade dos dados, além de aprovar políticas e diretrizes para uso adequado dos dados e participar de reuniões/decisões sobre dados.	<ul style="list-style-type: none"> Especificar os controles apropriados, com base na classificação de dados, para proteger os recursos de informação a partir de modificações não autorizadas, exclusão ou divulgação. Garantir que os controles tenham sido implementados para fornecer qualidade e integridade dos dados. Participar em reuniões de definições e decisões que envolvam os dados de seu interesse.

		<ul style="list-style-type: none"> • Influenciar colegas e superiores na importância da governança, qualidade e segurança dos dados.
Líder de Governança de Dados	É o papel que tem a responsabilidade pelo escritório de governança de dados; a principal responsabilidade - e habilidade - do líder de governança de dados é não se importar com quem “vence”.	<ul style="list-style-type: none"> • Coordenar tarefas para Analistas de GD e Data Stewards • Ajudar a comunicar decisões tomadas por administradores a partes interessadas relevantes • Impulsionar a auditoria de dados e métricas que avaliam o sucesso do programa • Principal ponto de escalada para o Conselho e Comitês de governança de dados. • Mediador de conflitos de interesse das áreas e impulsionador de decisões com escalada dos assuntos conflitantes
Analista de Governança de Dados	Parte do time do escritório de governança de dados, são as pessoas que fazem as engrenagens funcionarem e garantem que os processos de governança de dados estão sendo seguidos.	<ul style="list-style-type: none"> • Analisar perfis de dados • Verificar se regras de qualidade estão corretas • Analisar processos e seus resultados • Propor planos de ação para correção de dados • Participar de projetos • Fornecer visão de governança de dados em reuniões e treinamentos
Comitê de Governança de dados	Grupo de data owners que avaliam as necessidades de dados nas áreas	<ul style="list-style-type: none"> • Servir como primeira escala na resolução de problemas nos dados, tanto em processos, como funcionalidade e pessoas. • Avaliar, discutir e definir as necessidades de dados trazidas pelas áreas • Definir, monitorar e reportar as métricas táticas de governança de dados
Conselho de Governança de dados	São membros selecionados que lideram a estratégia de dados na empresa. Conta com representantes seniores das áreas mais críticas, levando em conta qual o assunto em pauta.	<ul style="list-style-type: none"> • Tomar decisões de alto nível envolvendo dados • Definir as políticas relativas aos dados • Definir as estratégias que devem ser seguidas • Influenciar tomadas de decisão em nível C e Steering meetings.
Steering Committee	O Steering Committee é o sponsor comercial e o defensor da estratégia e política de dados. Esses indivíduos supervisionam o sucesso do escritório de governança de dados.	<ul style="list-style-type: none"> • Tomar decisões de nível estratégico envolvendo dados • Patrocinar políticas e visão de Governança de Dados
Escritório de	Trabalha com líderes de negócios e de TI para desenvolver e implementar soluções de	<ul style="list-style-type: none"> • Participar em decisões envolvendo dados (todos os níveis), com evidências e papel mediador • Definir e manter glossário de dados

Governança de Dados	governança de dados. Também coordenam a integração entre várias disciplinas de controle de dados, além de administrar, monitorar e relatar atividades de controle de dados.	<ul style="list-style-type: none"> • Definir e manter processos de verificação dos dados • Definir e gerar reportes e KPIs sobre os dados • Ponto focal para assuntos relacionados a dados • Participar de projetos que envolvem dados
Parceiros de Governança de Dados	São os departamentos que a governança de dados trabalha em estreita colaboração. Geralmente incluimos nesse grupo: Compliance, Governança corporativa, PMO e TI. Algumas das funções possuem interdependência, dependendo do tipo de atividade a ser concluída.	<ul style="list-style-type: none"> • Participar de reuniões para assuntos relacionados a dados (quando afeta alguma das áreas) • Prover informações críticas relevante a sua área de conhecimento, quando requisitado

Tabela 1: Definições de papéis e responsabilidades

POLÍTICAS

Antes da implementação da função de governança de dados, é importante definir as políticas que vão guiar todos os esforços das pessoas/áreas, bem como informar para a organização os objetivos da governança de dados.

COMPONENTES DAS POLÍTICAS

O primeiro passo para definir as políticas é conhecer a empresa. Algumas são mais informais e, portanto, não terão políticas definidas, mas sim um manual ou diretrizes, portanto antes de começar a escrever algo muito formal, se familiarize com o material que existe na empresa, para o trabalho condizer com a cultura da empresa.

Outro fator importante é analisar as políticas que já existem e refletir sobre o que as políticas de dados vão acrescentar às que já existem. A análise crítica é importante para não criar uma política que não irá acrescentar nada e será somente mais um “pedaço de papel”.

POLÍTICA DE GOVERNANÇA DE DADOS

Na política de governança de dados devemos encontrar:

- O que é a Governança de Dados
- Porque a governança de dados é importante e necessária
- Definição do programa de governança de dados
- Como a governança de dados irá funcionar
- Quem são os responsáveis pela governança de dados (papéis e responsabilidades)
- Quem apoia a governança de dados
- Quando será aplicado

- Quando será compulsório e auditado

A política de governança de dados também deverá conter referência a qualquer outra política que afete ou complemente a sua função, por exemplo, políticas de compliance e de segurança da informação.

Geralmente, qualidade de dados é um assunto muito extensivo para estar contido dentro da política de governança de dados, por isso, é recomendado que para esse assunto haja uma política separada, e que seja referenciada na política de governança de dados, como uma política complementar.

Se a empresa ainda não tiver política de segurança da informação ou compliance, esses tópicos devem ser incluídos na política de governança de dados ou então devem ser criados separadamente, o que for de preferência da empresa.

POLÍTICA DE QUALIDADE DE DADOS

A política de qualidade de dados está intrinsecamente ligada à política de governança de dados e terá vários aspectos que serão os mesmos da governança de dados. Por isso, algumas empresas preferem criar uma política somente, porém, criar uma política para os dois assuntos a torna muito pesada e de difícil verificação. Quebrando os assuntos podemos ter um maior controle sobre o que está sendo realizado e garantir que a política está sendo seguida.

Na política de qualidade de dados devemos encontrar:

- O que é a qualidade de dados
- Porque a qualidade de dados é importante e necessária
- Definição do programa de monitoramento qualidade de dados
- Como a qualidade de dados irá funcionar
- Quem são os responsáveis pela qualidade de dados (papéis e responsabilidades)
- Quem apoia a qualidade de dados
- Quando será aplicado
- Quando será compulsório e auditado

PROCESSOS

Após as definições de modelo operacional e políticas, podemos começar a pensar em quais os processos que serão necessários para a governança de dados.

Frequentemente, são criados os processos para verificação e manutenção dos dados, dicionário de dados e linhagem de dados, que ficam sob a responsabilidade do escritório de governança de dados. Porém, necessitamos fazer a análise de todos os processos que lidam com dados (inclusão, exclusão, alteração, seja manual ou batch), bem como processos de compliance, controles internos, segurança da informação e projetos.

A governança de dados é responsável por todo o ciclo dos dados, por isso é tão importante conhecer todos os processos que afetam os dados e que são processos dependentes.

PROCESSOS INTERDEPENDENTES

O framework da Informatica¹descreve três tipos de processos que fazem parte do ciclo dos dados. Os processos que possuem dependência são chamados de Upstream, Stewardship e Downstream.

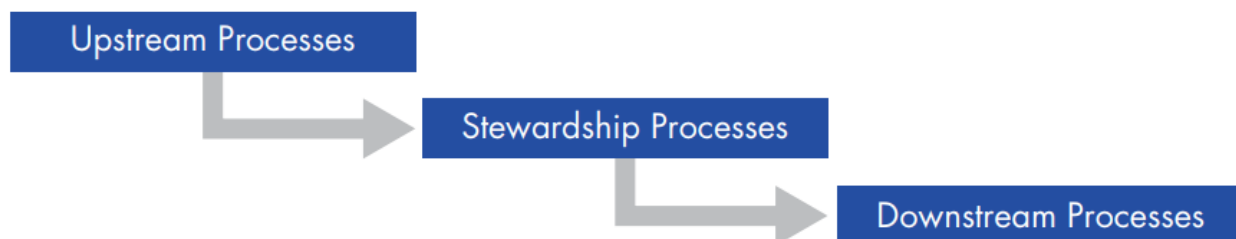


Figura 15: Tipos de Processos

PROCESSOS UPSTREAM

São aqueles processos de negócio que capturam, criam, importam, compram, transformam ou atualizam os dados nos sistemas de uma empresa. Um dos grandes desafios da governança dos dados é evangelizar as pessoas que fazem parte desse processo, de que existem outras pessoas usando esses dados e dos problemas causados no final da cadeia (processos downstream).

PROCESSOS DE STEWARDSHIP

Esses são os processos criados para fazer a verificação da qualidade dos dados, se as políticas, regras, padrões e definições de dados estão sendo aplicadas. Esses processos podem ser feitos manualmente ou através de ferramentas que verificam, higienizam, enriquecem, mascaram, identificam e resolvem duplicados, corrige e valida ou que melhora a segurança e qualidade dos dados. Em um fluxo manual, esses processos facilitam para o usuário a identificação, notificação, escalada, e mitigação de exceções em regras e políticas que estejam automatizadas.

PROCESSOS DOWNSTREAM

São os processos operacionais e analíticos que consomem, protegem, arquivam, saneiam, ou que produz conhecimento e valor sobre os dados. Os processos downstream são os que vão dar visibilidade dos problemas e possibilidades de melhorias nos processos upstream. As melhorias podem ser de processos, sistemas, culturais, treinamentos. Com esses relatórios, a governança de dados tem mais chances de obter suporte executivos para realizar as mudanças e melhorar toda a cadeia de dados, que ficará evidenciado nos processos de stewardship e downstream.

PROCESSOS ESPECÍFICOS

¹ P.17, 18 <https://www.informatica.com/resources.asset.06393c8448cb01a327701b27add88046.pdf>

Ainda seguindo o framework da informática, temos os processos específicos. Como a governança de dados é uma função corporativa como qualquer outra, obviamente tem processos que precisam ser definidos para a área, e que são específicos para essa função.

A figura abaixo ilustra os processos que compõem a função de governança de dados. Alguns exemplos incluem os processos que limpam, reparam, mascaram, protegem, reconciliam, escalonam e aprovam discrepâncias, políticas e padrões de dados. O framework segmenta os muitos processos de governança de dados em estágios fundamentais: analisar, definir, aplicar (regras e políticas) e medir/monitorar:



Figura 16: Ciclo de Processos

Além dos processos mencionados, vale lembrar de processos de atualização de glossário de negócio (dicionário de dados), linhagem de dados e processos de levantamento de risco.

ANALISAR

São os processos definidos para realizar o levantamento de como os dados estão hoje e quais os ciclos de dados que existem, quais as interdependências entre processos e/ou sistemas. O fluxo desses processos permite uma análise que dará base para as estratégias de governança de dados, propostas de priorização, políticas e padrões, assim como fornecer uma visão mais clara do futuro.

Inclui os processos de suporte de descoberta de dados, criação de perfil de dados, inventário de dados e processos de estado atual, análise CRUD (do inglês create, read, update, delete, ou criação, leitura, atualização e exclusão) no ciclo de vida de dados e avaliações de capacidades organizacionais, de pessoas e de tecnologia.

Os processos de análise geram insumos para os processos de definição, enquanto que os processos de definição criam um foco mais direcionado para os processos de análise.

DEFINIR

Os processos de definição documentam as descrições dos dados e o contexto de negócios associado à terminologia, outras taxonomias (identificação e classificação dos dados) e relacionamentos, bem como políticas,

regras, padrões, processos e estratégia de medição que devem ser definidos para operacionalizar os esforços de governança de dados.

Esse estágio inclui os processos de suporte de criação de glossário de negócios, classificação de dados, relacionamento de dados, definições de dados de referência e hierarquia e a definição de regras de negócios, políticas e principais indicadores de desempenho (KPIs).

APLICAR

Os processos da fase 'aplicar' são aqueles de operacionalização e garantia de conformidade com as políticas de governança de dados, assim como regras de negócio, processos administrativos, fluxos de trabalho e as funções e responsabilidades multifuncionais que foram descobertos nas etapas de análise e definição.

Inclui os processos que permitem a automação das regras e políticas dentro de sistemas, processos e serviços. Também inclui os processos que operacionalizam o suporte aos processos manuais e workflows de TI que mitiguem exceções de qualidade de dados e de segurança da informação manualmente.

MEDIR E MONITORAR

Aqui, os processos capturam e medem a eficácia e o valor gerado pela governança de dados e dos esforços de stewardship, rastreia a conformidade e as exceções às políticas e regras definidas e garante que os dados e seu ciclo de vida sejam transparentes e auditáveis.

Este estágio inclui os processos de suporte, tais como:

- Monitoramento proativo da qualidade dos dados e de compliance com as políticas
- auditoria reativa da qualidade dos dados operacionais
- Análise da linhagem dos dados para determinar a raiz do problema e levantamento de impactos.

Além disso, essa fase inclui os processos de medição quantitativa e qualitativa da efetividade do programa de governança de dados e o seu valor comercial.

DADOS MESTRE

O gerenciamento dos dados mestre é uma grande parte do trabalho de governança de dados. Por serem dados corporativos, usados e reusados por várias áreas de negócio e serem críticos para o negócio, esses dados são os que prioritariamente serão monitorados tanto proativamente quanto reativamente.

DEFINIÇÃO DOS DADOS MESTRE

O primeiro passo é a definição de quais são os dados mestre da empresa:

- Quais os dados que são comuns para todas as áreas?
- Quais os dados mais usados em relatórios gerenciais?
- Quais os dados obrigatórios nos sistemas?
- Quais os dados necessários para criar novos produtos?

Muitas vezes os dados mestre podem ser confundidos com dados transacionais, porém, apesar de às vezes compor os dados mestre, dados transacionais geralmente não fazem parte dos dados mestre.

Por exemplo, em uma análise de crédito, incluímos obrigatoriamente o salário da pessoa para que os cálculos possam ser feitos. Porém, esse valor não é necessariamente um dado mestre, já que muda muito frequentemente e é usado somente para um objetivo, que é gerar o score do cliente para uma transação específica.

Listas de referência ou lookups, podem fazer parte dos dados mestre e quando são gerenciados, a exclusão, alteração ou inclusão de valores nessas listas precisam passar pelo processo de verificação e autorização da área de governança de dados.

criação de domínio de dados

Após a descoberta dos dados mestre, é útil classificar os dados e coloca-los em domínios. Os domínios servem para demarcar as responsabilidades dos data owners e também para facilitar as análises dos dados, que vão ser sempre feitas seguindo a definição de domínios.

Eles podem ser divididos em dados pessoais PF, dados PJ, contato, financeiro, correios, etc.

Gerenciamento

O gerenciamento dos dados mestre se dá através da realização de profile dos dados, análise de scorecards, análise dos dados, insumos gerados por exceções às regras de qualidade, etc.

Deve haver um processo específico para o monitoramento e gerenciamento dos dados mestre.

Ao iniciar um programa de governança de dados é interessante classificar os dados em níveis de criticidade e focar nas resoluções dos problemas dos dados mais críticos, mas sem perder de vista o plano de resolução dos outros dados.

Glossário de Negócios

O glossário de negócios, também chamado de dicionário de dados, não deve ser confundido com definição de dados técnicos, embora, se conhecido, os nomes técnicos e suas localizações podem fazer parte do glossário, e as definições de negócio podem estar atreladas à linhagem dos dados feita sistemicamente.

Porém, podemos começar a coletar informações sobre os dados muito antes de criar qualquer coisa técnica para fazer linhagem e/ou descoberta dos dados.

O glossário de negócios é muito importante e simples de ser iniciado. Muitas vezes, áreas diferentes dentro de uma empresa podem ter seus glossários locais. Se esse for o caso, a governança de dados irá obter esses glossários locais e transformá-los em um glossário de negócios corporativo. Quando não existir nenhum glossário na empresa, a governança de dados deve reunir as áreas de negócios para iniciar a coleta e definições dos termos.

Nos dois casos, quando há divergência nas definições, as áreas devem entrar em um consenso para que toda a empresa possa usar a mesma definição do termo. Quando é impossível conciliar as definições, a governança de dados deve pensar em qual meio é o mais eficiente para capturar essa informação, pensando também nos dados

técnicos, linhagem e uso dos dados, para que os dados não sejam usados erroneamente, pois se áreas diferentes usam o mesmo campo técnico para armazenar dados com objetivos diferentes, relatórios e KPIs podem estar sendo mascarados por dados incorretos.

A manutenção do glossário geralmente fica sob a responsabilidade do escritório de governança de dados e deve haver um processo definido para isso, demonstrando os periodicidade de verificação e o fluxo de manutenção.

CONTEÚDO DO GLOSSÁRIO

O glossário de dados será como uma enciclopédia para aquele dado. Pode conter:

- Definição unificada
- Histórico do dado
- A quais processos pertence
- Em quais sistemas está localizado
- Quem é o data owner / área responsável
- Relatórios e KPIs de que faz parte
- Datas e propósitos de atualizações/inclusões/deleções batch

LINHAGEM DE DADOS

A linhagem de dados é importante para a governança de dados pois permite rastrear um dado desde sua origem e todo o percurso que está fazendo, passando por outros sistemas, downloads, relatórios, bases de BI, Excel, data lakes, etc.

É um item importante para compliance com legislações de dados que querem saber porque os dados são capturados, para que são usados e quando são expurgados.

A linhagem de dados está muito conectada ao glossário de dados, e a junção desses dois aspectos gera valor tanto para áreas de negócio quanto para TI.

Quando a linhagem de dados é realizada, podemos relacionar esses dados puramente técnicos (nomes de campos em tabelas de sistemas diferentes) com os termos de negócio que foram definidos no glossário de dados.

O valor desse relacionamento está na facilidade com que áreas de negócio podem acessar essas informações e entendê-las sem que alguém de TI tenha que ajudar, isto é, não é preciso abrir um chamado para descobrir se, por exemplo, existe “faturamento” em algum sistema. Da mesma maneira, isso facilita o entendimento do campo por TI, de maneira confiável e rápida. Beneficia também as definições de projetos de TI, demandando menos tempo em descoberta de dados e esclarecimentos nas definições dos termos de negócio.

SEGURANÇA DA INFORMAÇÃO

Geralmente essa é uma área onde as empresas já tem um maior controle, com políticas, normas e processos já definidos. Porém, caso haja a necessidade de definição dessa área, vamos focar nos assuntos principais.

O que é segurança de dados? Está ligada ao compartilhamento de dados (proteção, mascaramento, arquivamento), aos acessos concedidos (quem pode acessar o que/porque/quando) e a classificação de importância dos dados.

PAPÉIS E RESPONSABILIDADES

A política deve definir quais os papéis e responsabilidades, assim como um organograma e sistema de escalonamento de solução de problemas.

COMPARTILHAMENTO DE DADOS

A política de segurança de dados deve descrever quando os dados podem ser compartilhados e sob quais condições.

Normalmente, a política ditará quais os meios aceitáveis para compartilhamento de dados, quais os métodos de criptografia a serem usados, a quantidade de dados que podem ser compartilhados, etc.

A política também deve conter informações sobre mascaramento de dados, como dever ser feita e sob quais circunstâncias. Geralmente o mascaramento de dados é necessário para desenvolvedores e testes de novas aplicações, portanto esses usuários não têm privilégios ou a necessidade de acessar dados de produção.

A política também deve descrever quais os parâmetros usados para o arquivamento e eliminação de dados inativos dos sistemas. Geralmente o arquivamento/eliminação de dados tem um requerimento legal, e, portanto, devem sempre fazer referência às leis que regem essa definição.

ACESSOS

A política deve definir como os dados podem ser acessados e quem concede e controla os privilégios.

Os acessos podem ser garantidos por sistema e se possível por dados a serem visíveis dentro de cada sistema, pois podem existir dados que não sejam apropriados para certos níveis e propósitos de acesso.

CLASSIFICAÇÃO DOS DADOS

Após a definição dos níveis de classificação, a política deve definir como novos dados devem ser classificados, qual o processo para tanto e quais as responsabilidades das áreas de negócio e de TI.

FASE 3 – QUALIDADE DE DADOS

A implementação de um programa de qualidade de dados vai depender dos objetivos de cada empresa. Algumas querem somente melhorar o que existe, mas sem as práticas da governança de dados.

Isso é possível? Sim. Se o objetivo for, por exemplo uma migração, a fase de qualidade de dados pode ser aplicada, porém é necessário uma definição de responsabilidades bem cedo no projeto, pois invariavelmente haverá várias exceções nas regras de qualidade aplicadas que precisaram ser verificadas por áreas de negócio e se não houver essa conscientização, muito dos frutos de uma qualificação dos dados para uma migração podem ser perdidos, ou até descartados completamente.

Porém, o mais recomendado é que a qualidade de dados seja vista como um processo cíclico que precisa estar em constante monitoramento e evolução. Essa mentalidade, junto com a governança de dados garantem que os dados estejam sempre monitorados e melhorados, para que possam gerar decisões com dados confiáveis e de onde todos saibam qual a sua fonte.

Um ponto importante em qualidade de dados: é impossível que todos os dados estejam 100% qualificados 100% do tempo. Por isso, ao iniciar um programa de qualidade de dados, é importante ser realista e saber definir o nível de importância dos dados, e começar a trabalhar com os mais críticos, resolvendo todos os problemas possíveis, antes de iniciar uma nova fase, com dados menos críticos. Portanto, a dica é: Defina seus dados mestre o quanto antes, e classifique todos quanto à sua importância. Esse será o quadro que deverá seguido para resolução de problemas de qualidade.

COMO CLASSIFICAR A IMPORTÂNCIA DE UM DADO?

Analise os dados e sistemas, faça entrevistas com as áreas. Algumas perguntas a serem respondidas:

- É dado mestre?
- O dado é de preenchimento obrigatório nos sistemas?
- O dado é usado em vários relatórios gerenciais?
- O dado é crítico para o funcionamento de uma área?
- O dado é usado na criação de novos produtos / campanhas de marketing?

Quanto mais o dado é usado, mais ele pode ser considerado crítico, porém não esqueça que alguns dados “business core” podem não ser tão usados, mas são os mais importantes para o funcionamento da empresa de forma corporativa, portanto isso deve ser levado em conta quanto for medir a criticidade.

Em alguns projetos, dependendo da quantidade de dados mestre, devemos limitar as análises dos dados aos dados de maior importância, pois muitas não há tempo hábil para analisar, definir e implementar todas as regras de todos os dados – e esse não deveria ser o objetivo de um projeto inicial de qualidade de dados – portanto uma visão crítica e a classificação dos dados focam os esforços e evidenciam as vantagens de um programa de qualidade de dados.

DIMENSÕES DA QUALIDADE DE DADOS

Existem seis dimensões essenciais na qualidade²:

- **Completeness:** Completude – Proporção de dados armazenados contra o potencial para 100%
- **Uniqueness:** Singularidade – Não deverá haver mais de uma instância para o mesmo registro
- **Timeliness:** Fidelidade – O grau em que o dado representa a realidade quando requerido
- **Validity:** Validade – Dado é válido se está em conformidade com as definições (syntaxe, tipo, intervalo)
- **Accuracy:** Acuracidade ou Exatidão – O grau em que o dado corretamente descreve o “mundo real” do objeto ou evento sendo descrito
- **Consistency:** Consistência – A ausência de diferença, quando comparando duas ou mais representações do mesmo objeto contra as definições.

Ao longo da jornada de criação, definição e ciclos de governança de dados, as organizações devem escolher algumas dimensões, já que fica inviável monitorar todas desde o início, onde não há maturidade para tanto.

Assim como a maturidade e as funções de governança de dados, as dimensões de qualidade de dados devem atingir um certo nível de maturidade para que novas dimensões possam ser medidas.

Podemos atrelar as dimensões de qualidade de dados aos níveis de maturidade apresentados anteriormente:

² DAMA DMBOK2, p.457

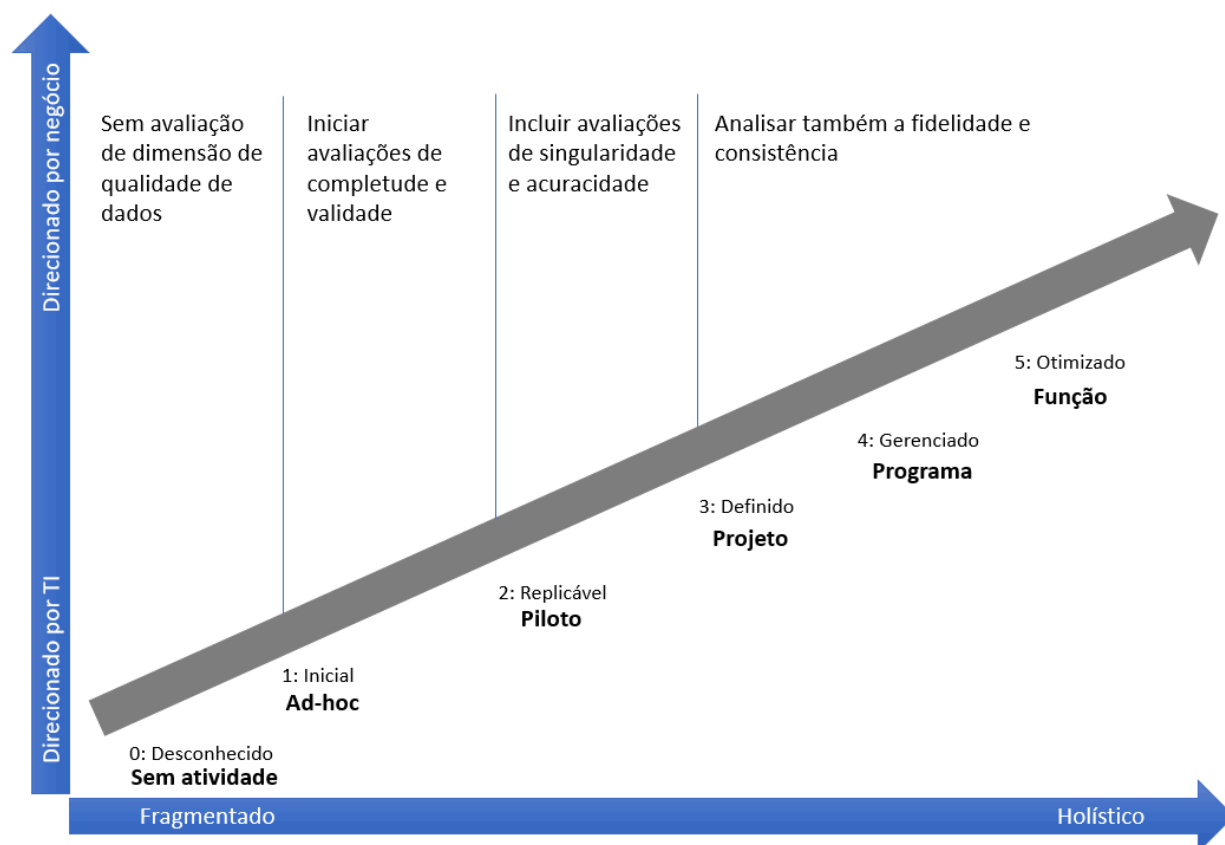


Figura 17: Maturidade e as dimensões de qualidade de dados

Ao iniciar um projeto de qualidade de dados, seja ele inicial ou já parte do ciclo de monitoramento de governança de dados, vale sempre lembrar que todo o fluxo se inicia com a análise de dados, quer seja ação reativa ou proativa.



Figura 18: Ciclo de Qualidade de Dados

A análise dos dados possibilita o conhecimento dos dados e seus problemas. Pode ser feita de várias maneiras e sem uso de ferramentas específicas, mas frequentemente as empresas que possuem ferramentas mais sofisticadas para essas análises querem sugestões para que possam melhorar o nível de suas análises.

Quando o analista conhece um pouco de SQL statements, portanto consegue criar queries simples para obter a informação necessária, esse pode ser o caminho mais simples e menos custoso.

Outra alternativa é criar uma amostra para o analista, porém os dados não podem ser manipulados ou limpos, por exemplo, se o CPF está separado, a extração tem que trazer do mesmo jeito que está na tabela para não causar falsas impressões que podem gerar regras incorretas ou desnecessárias.

Os perfis dos dados geralmente são providos por ferramentas especializadas para tanto. Com os dados disponibilizados, o analista pode criar os perfis dos dados e obter os padrões mais facilmente.

Nas imagens abaixo está um profile produzido pela ferramenta Analyst da Informatica. O profile mostra as informações do dado ZIPCODE.

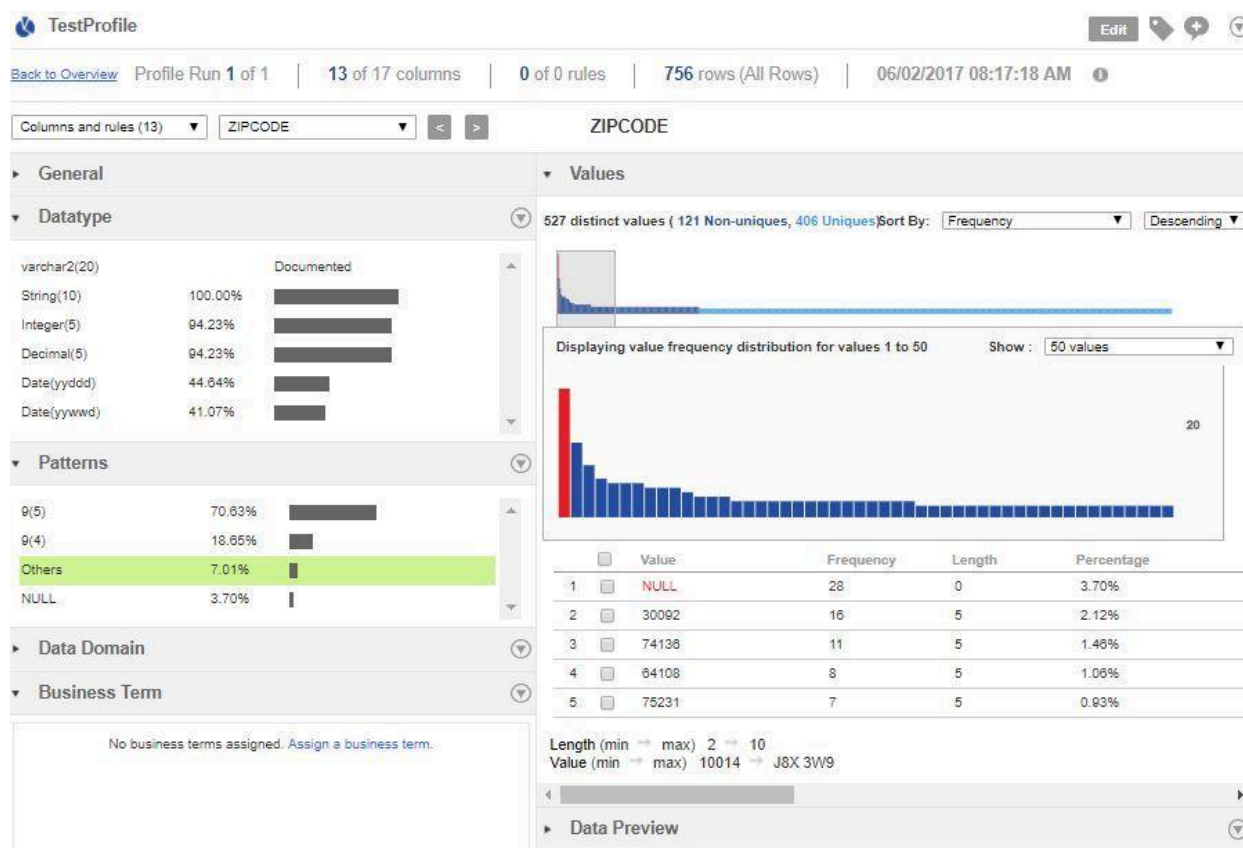


Figura 19: Exemplo de Profile criado com a ferramenta Analyst da Informatica

Na sequência temos o “drilldown”, isto é, mostra os dos dados onde o padrão (pattern) é “outros” (others).

▼ Data Preview

[ZIPCODE : "Others"]

53 rows

	PARTY_ID	NAME	ADDRESS	ADDRESS2	CITY	STATE	ZIPCODE	COUNTRY
1	2742600	CORRIE BERTS...	489 Pittsfield Road	NULL	Lenox	MA	124	US
2	2769900	VENA FRIEDLY	235 Old Connect...	NULL	Framingham	MA	170	US
3	2727900	CIDA DITTMER	2985 Berlin Turn...	NULL	Newington	CT	611	US
4	2759400	MRS PAULETTA...	248 Flanders Road	NULL	East Lyme	CT	633	NULL
5	2870700	456	46 Springer Drive	NULL	BANGOR	ME	440	US
6	8318100	DENIS ALPERM...	102 Mid St Sout...	NULL	Redhill	Surrey	234-2341	US
7	2637600	JILLI B. MCQUAID	9200 Highway 1...	NULL	Alabaster	AL	35007-5344	NULL
8	3666600	ZARIA SORTO	2900 Veterans H...	NULL	Metairie	LA	70	US
9	3628800	FORTUNATO KI...	1217 E Chestnut	NULL	Chilton	WI	53	US

Figura 20: Data Preview (drilldown) da ferramenta Analyst da Informatica

Esse tipo de ferramenta facilita a verificação da padronização dos dados, e quais tipos de correções seria preciso aplicar para corrigir as disparidades.

Para mais informações sobre profiles e scorecards, feitos nas ferramentas Informatica, visite o canal do YouTube informatica support: <https://www.youtube.com/user/INFASupport/videos>

Realizar uma pontuação dos dados é um bom meio de seguir a evolução dos dados e possibilita a criação de KPIs e uma visão mais clara do que precisa ser monitorado com mais cuidado.

Geralmente, o score é uma média das notas que o dado recebe nas dimensões de qualidade, por exemplo completude. Portanto, para o cálculo, precisamos definir quais os níveis aceitáveis dentro de cada critério.

No caso de completude, o ideal é 100%, isto é, todos os campos escolhidos estão preenchidos e, portanto, completos. Então: 100% = bom, 99% -90% aceitável, 89% - 0% inaceitável. Os pesos também devem ser definidos após a definição das métricas dos campos. Os pesos são a importância de cada campo no final do score, por exemplo nome e CPF são mais importantes que data de nascimento, portanto o peso desses dois campos será maior que o campo da data de nascimento.

Se houver a ferramenta analyst da Informatica, isso pode ser incluído diretamente na criação de profile/scorecard.

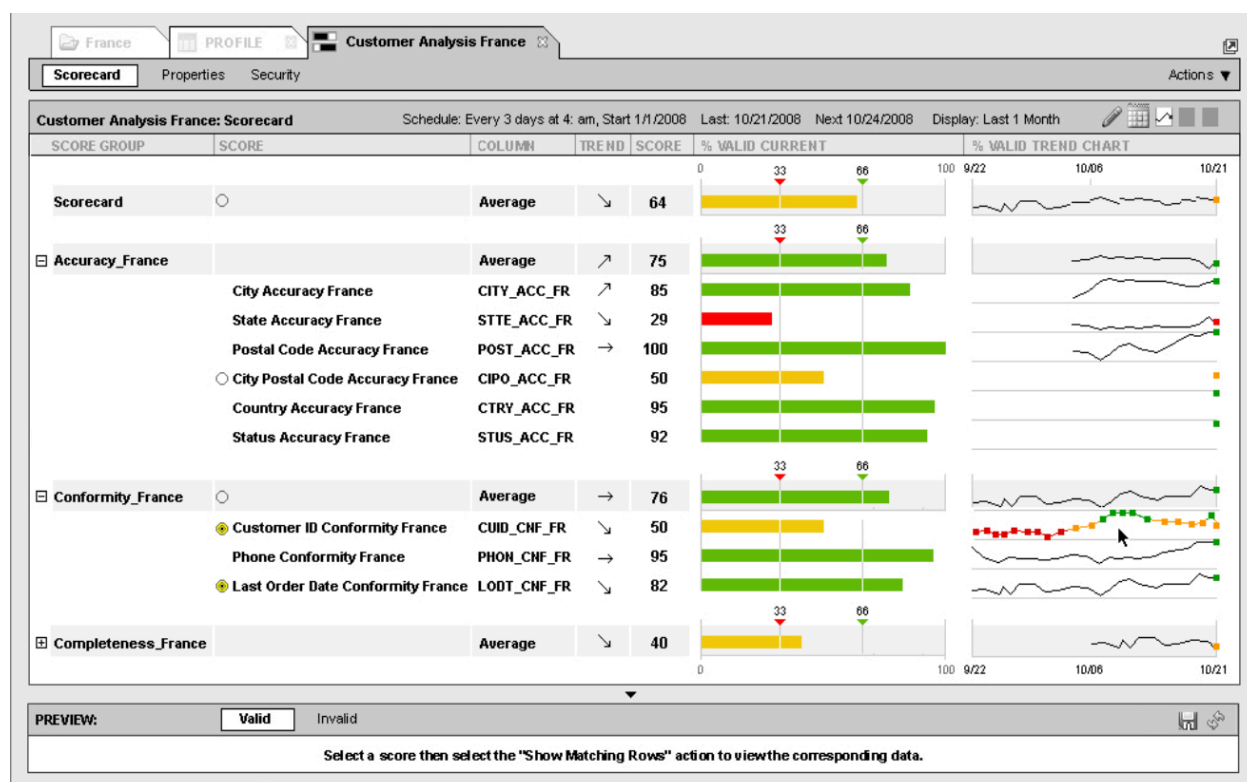


Figura 21: Exemplo de Scorecard com Dimensão de Qualidade

Para mais informações sobre profiles e scorecards, feitos nas ferramentas Informatica, visite o canal do YouTube informatica support: <https://www.youtube.com/user/INFASupport/videos>

Após a realização das análises dos dados, podemos começar a definir as regras que precisam ser aplicadas nos dados. As regras podem ser de padronização ou de higienização, separação de dados e validações.

REGRAS DE PADRONIZAÇÃO

São aquelas que colocam todos os dados dentro de um mesmo padrão a ser definido. Por exemplo:

- Campos que contêm datas devem sempre seguir o padrão dd/mm/yyyy
- Campos de números (telefone, CPF, etc.) não devem conter caracteres especiais ou letras.
- Todas as letras devem ser capitalizadas
- E-mail tem que ter estar no padrão definido internacionalmente
- Padronizar valores de acordo com lista de de-para, por exemplo no campo número do logradouro, se houver S/N, S.N., Sem Num, padronizar para SN.

REGRAS DE HIGIENIZAÇÃO

São aquelas regras que removem “ruídos” dos dados, por exemplo:

- Remover espaços duplos
- Remover espaços antes da primeira palavra e depois da última (leading and trailing spaces)
- Remover acentuação
- Remover caracteres especiais

REGRAS DE SEPARAÇÃO

São regras que realizam a separação dos dados contidos em um campo que deveriam estar em vários campos diferentes.

Exemplos de regras de separação, também conhecidas como parse:

- Move os dados para os campos corretos para permitir uma melhor limpeza, padronização e verificação dos dados.

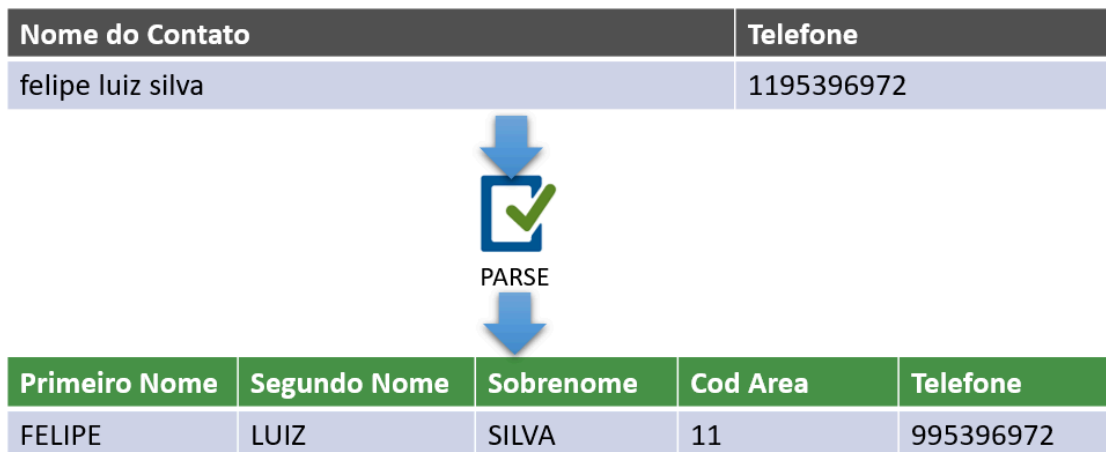


Figura 22: Aplicação de regra de parse e padronização

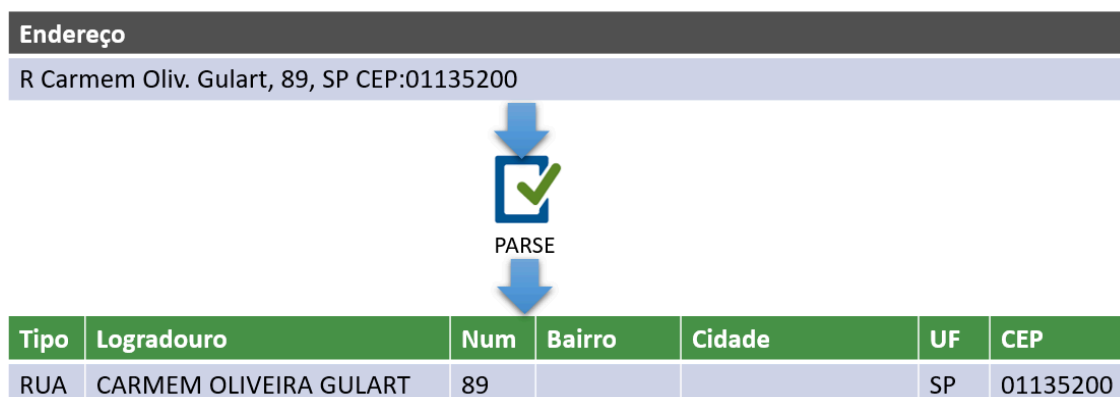


Figura 23: Aplicação da regra de parse e padronização para endereço

REGRAS DE VALIDAÇÃO

As regras de validação verificam se o dado contido é válido baseado em um parâmetro dado. Os parâmetros podem ser cálculos, tabelas de apoio, ou verificações de padronização, etc.

As regras de validação não modificam o dado, apenas indicam se o dado inserido está de acordo ou não com a validação.

Ao definir esse tipo de regra, deve-se definir também o que acontece com as suas exceções, pois se a validação for crítica, por exemplo a validação do CPF, e falhar, talvez o ideal seja que todo o cadastro seja enviado para verificação manual.

As regras de validação podem também ter somente um indicador se a regra foi atendida (0) ou não (1), para que quando realizamos o profile e score do dado possamos utilizar esses valores para definir se existe necessidade de verificação posterior do dado.

Alguns exemplos de regras de validação:

- Validar dígito verificador de CPF e CNPJ
- Validar se telefone contém somente números
- Validar UF contra tabela de apoio
- Validar nome contra lista de nomes impróprios

DEFINIÇÃO DE REGRAS DE ENRIQUECIMENTO

As regras de enriquecimento devem ser definidas quando houver a necessidade de incrementar os dados existentes com fontes externas ou tabelas auxiliares.

As fontes de enriquecimento mais comuns são os bureaux de informação (TransUnion, Serasa, Boa vista, etc.) e informação de endereços (geralmente fornecida pelos correios, porém existem empresas que fornecem dados de endereços com atualizações mais frequentes que os correios).

As definições devem conter quais os parâmetros para encontrar o mesmo registro e os campos que podem ser atualizados e quais as regras para tanto.

Exemplo:

RE4 - PESSOA FÍSICA - ENRIQUECIMENTO BIRÔ				
Esta regra deverá ser aplicada quando o registro pertencer à pessoa física, onde o cadastro com está marcado como "Pessoa Física".				
ID	Dado	Descrição da regra	Exceção	Observação
1	CPF	Verificar se o CPF combinado com a data de nascimento existe na receita federal. Se sim, atualizar todos os campos relevantes e possíveis.	Enviar para verificação manual caso esses dados não sejam encontrados na receita federal.	
	Data de Nascimento			
2	Nome	Atualizar		
3	CPF Status	Atualizar		
4	Data de atualização	Incluir data da verificação/atualização para fins de verificação periódica dos dados.		

Tabela 2: Exemplo de regra de enriquecimento

DEFINIÇÃO DE REGRAS DE DEDUPLICAÇÃO

As regras de deduplicação, ou match & merge, devem ser definidas quando há a necessidade de unificar cadastros duplicados na base de dados.

As definições devem conter como realizar o match, que é a descoberta de possíveis duplicados e quais os critérios para a unificação automática e manual, além de informar como o merge deve ser realizado. O merge é a unificação em si, portanto é necessário definir parâmetros para que os dados sejam unificados corretamente, especialmente no modo automático.

O match & merge também pode unificar informações dentro do mesmo cadastro, como é o caso de múltiplos endereços, contatos, telefones, etc. Essas regras também precisam ser definidas e detalhadas.

Exemplo:

MM1 - PESSOA JURÍDICA			
Os cadastros de PJ deverão ser unificados onde houver similaridades em níveis suficientemente confiáveis para que a unificação possa ser feita automaticamente, mantendo a integridade e histórico e amarrações dos cadastros.			
ID	Dado	Descrição da regra	Exceção
1	Razão Social/ CNPJ	PARA CADASTROS ONDE PAÍS = BRASIL A busca por cadastros duplicados deve ser feita: 1. Pelo CNPJ e então Razão Social ou 2. Pela Razão Social e então CNPJ Quando razão social e CNPJ forem 100% iguais em dois ou mais cadastros, unificar automaticamente, de acordo com especificações em 'Regras de Merge'.	PARA CADASTROS ONDE PAÍS = BRASIL 1. QUANDO MATCH DE CNPJ = 100% RAZÃO SOCIAL < 100% Enviar todos os casos para verificação e decisão manual. 2. QUANDO MATCH DE RAZÃO SOCIAL = 100% E CNPJ < 100% Se os dois campos estiverem em níveis entre 99% e 75% de igualdade, incluir os "potential match" para verificação e decisão manual. Quando o CNPJ não for igual nos registros e os níveis de igualdade menores que 74% os cadastros não devem ser considerados duplicados.
		PARA CADASTROS ONDE PAÍS <> BRASIL A busca por cadastros duplicados deve ser feita: 1. Pelo campo CNPJ e então Razão Social ou 2. Pela Razão Social e então CNPJ Quando razão social e CNPJ forem 100% iguais em dois ou mais cadastros, unificar automaticamente, de acordo com especificações em 'Regras de Merge'.	PARA CADASTROS ONDE PAÍS <> BRASIL Se os dois campos estiverem em níveis entre 99% e 75% de igualdade, incluir os "potential match" para verificação e decisão manual. Níveis de igualdade menores que 74% não devem ser considerados duplicados. Porém a busca deve continuar no próximo item.
2	Razão Social	SOMENTE PARA CADASTROS ONDE PAÍS <> BRASIL A busca por cadastros duplicados deve seguir com a verificação da Razão Social somente, pois para outros países o campo CNPJ pode não ser obrigatório ou pode conter informações conflitantes. Quando as duas forem 100% iguais em dois ou mais cadastros, enviar para verificação manual.	SOMENTE PARA CADASTROS ONDE PAÍS <> BRASIL Se os dois campos estiverem em níveis entre 99% e 75% de igualdade, incluir os "potential match" para verificação e decisão manual. Níveis de igualdade menores que 74% não devem ser considerados duplicados.

Tabela 3: Exemplo de definição de match

MM1 - PESSOA JURÍDICA			
ID	Tipo	Tabela	Descrição
1	Cliente	XXUC_HZ_CUST_RELATIONSHIPS	Realizar 'Group By' pelos seguintes campos: ACCOUNT_NUMBER RELATED_ACCOUNT_NUMBER ORG_ID
2	Cliente	XXUC_HZ_PARTY_RELATIONSHIP	Realizar 'Group By' pelos seguintes campos: PARTIES_PARA_PARTY_NUMBER PARTIES_PARA_PARTY_NUMBER_RELATED
3	Cliente	XXUC_HZ_CUST_ACCOUNTS	Realizar 'Group By' pelos seguintes campos: ID_ORIGEM_PARA_EMPRESA ID04_CADASTRO_PARA_EMPRESA

FASE 4 – ROLL OUT

Esta última fase do framework tem como objetivo comunicar a organização e treinar os participantes dos processos desenhados.

Essa etapa não precisa, necessariamente, ser feita somente ao final da implementação, por exemplo o plano de comunicação deve ser feito o quanto antes pois quanto mais cedo, mais familiaridade as pessoas terão com o assunto.

PLANO DE COMUNICAÇÃO

Toda e qualquer mudança organizacional deve ter um forte plano de comunicação. Pode começar cedo, com a comunicação da necessidade, como será realizado a análise, quem irá decidir sobre os assuntos e quem será afetado.

A comunicação contínua é chave para a aceitação e entendimento de um assunto que na maioria das vezes é desconhecido pela maioria das empresas.

Pensar nos canais de comunicação é fundamental na esteira de planejamento e implementação de governança de dados e deve contar com apoio executivo, quando necessário.

Quando possível devemos pensar também nos “change champions” pessoas selecionadas dentro das áreas que serão os evangelizadores das mudanças que irão ocorrer.

MEIOS DE COMUNICAÇÃO

Os anúncios podem ser feitos de diversas maneiras e usando diversos canais, o planejamento tem que levar em conta o que é mais efetivo dentro da empresa.

Os meios mais comuns são:

- Apresentações formais (powerpoint)
- E-mails
- Newsletters
- Intranet
- Quadro de avisos

Dependendo dos estágios de implementação podem ser realizados “roadshows” com a apresentação mais formalizada do que está acontecendo, irá acontecer e o objetivo final.

TREINAMENTOS

Os treinamentos formais são essenciais para o sucesso da implementação, especialmente no caso de utilização de novas ferramentas.

No caso de treinamento de ferramentas, os usuários devem receber manuais e instruções práticas de como utilizar as funções necessárias para a suas atividades.

No caso de participantes nos processos de governança de dados, deverá ser feito um overview do processo e o que é esperado de cada pessoa envolvida.

FONTES

<https://erwin.com/blog/top-6-benefits-of-data-governance/>

<http://blog.rdstation.com.br/gdpr/>

<https://www.edq.com/blog/data-quality-vs-data-governance/>

<https://www.informatica.com/resources.asset.06393c8448cb01a327701b27add88046.pdf>

http://www.b-eye-network.com/blogs/oneal/archives/2015/07/the_right_operating_model.php

<https://blogs.perficient.com/2013/01/02/data-governance-roles-and-responsibilities/>

<https://www.youtube.com/watch?v=qiPeeQd0ZHs>

<https://www.slideshare.net/Dataversity/rwdg-webinar-how-to-construct-a-data-governance-policy>

<http://www.datagovernance.com/>

<https://searchdatamanagement.techtarget.com/ehandbook/GDPR-data-protection-edicts-make-good-data-governance-a-must>

[DAMA DMBOK, 2nd Edition](#)