



**1ª CONFERÊNCIA
INTERNACIONAL
EM QUALIDADE
DA INFORMAÇÃO**

Métricas em Qualidade de Dados

Mario F. Cervi

Métricas fazem parte do nosso dia-a-dia

O índice de congestionamento em São Paulo hoje está em ...

... 40 km



Chego logo !

... 200 km



Nem saio de casa !!!

Métricas fazem parte do nosso dia-a-dia

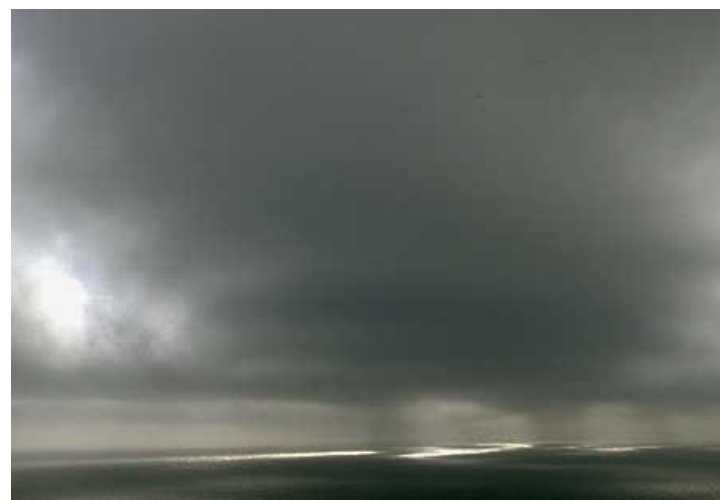
Hoje o tempo está ...

... bom



Pega o protetor !

... ruim



Cadê o baralho ?

Problemas Frequentes para as Empresas

- *Muitas faturas não chegam às mãos dos clientes*
- *Não conseguimos prever corretamente a reposição de estoque*
- *Não está claro qual é o perfil dos meus melhores clientes*



E onde estão as possíveis causas

- *Processos* ➡ *Auditar, Corrigir*
- *Pessoas* ➡ *Treinar, Motivar*
- *Informações* ➡ *Validar, Qualificar*

Como Qualificar Informações

Uma pergunta recorrente:



“De 0 a 10, qual a nota da qualidade dos meus dados ?”

Uma resposta simples e tentadora:

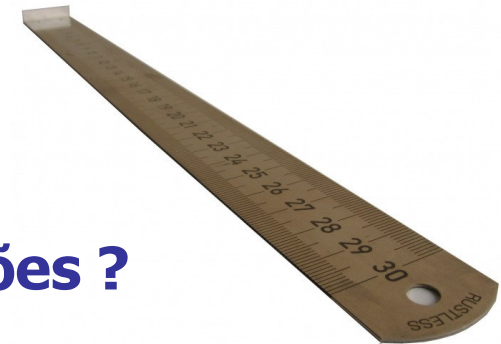
“De 0 a 10 ? Acho que entre 7 e 8.”

“Não, acho que é menos, entre 5 e 6.”

(cada um dá uma nota ...)

Como medir a qualidade das informações ?

Aliás, como *definir* qualidade das informações ?



***Para responder estas perguntas,
é preciso utilizar conceitos e métodos***

***Por isso, o MIT desenvolveu o
TDQM – Total Data Quality Management***

Conceitos de TDQM



Os Quatro Princípios

- 1. Entenda as necessidades de informação do usuário**
- 2. Gerencie a informação como produto de um processo de produção bem definido**
- 3. Gerencie a informação como produto com um ciclo de vida**
- 4. Defina um “Information Product Manager” para gerenciar o produto informação**

Conceitos de TDQM

Os Três C's



- **Produto informação: resultado da transformação de dados primários em informações que possuam valor para o consumidor**
- **No processo de produção, é preciso considerar a coleta, a custódia (ou armazenamento) e o consumo (ou utilização).**

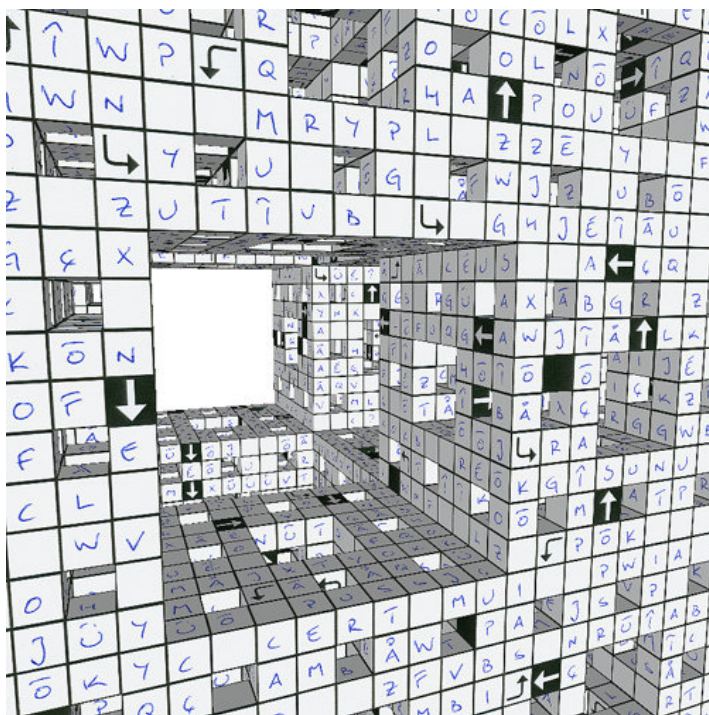


Conceitos de TDQM



As Dimensões da Qualidade da Informação

A Qualidade vai além da consistência



Qualidade de Dados

é

multi-dimensional

Conceitos de TDQM



As 16 Dimensões da Qualidade da Informação

Categoria	Dimensões
Intrínsecas	Integridade, Objetividade , Credibilidade, Reputação
Acessibilidade	Facilidade de Acesso, Segurança de Acesso
Contextuais	Relevância, Valor-Agregado, Atualidade, Completude, Volume de Dados, Facilidade de Uso
Representação	Interpretabilidade, Facilidade de Entendimento, Representação Concisa, Representação Consistente

Conceitos de TDQM



As 16 Dimensões da Qualidade da Informação

Categoria	Dimensões
Intrínsecas	Integridade, Objetividade , Credibilidade, Reputação
Acessibilidade	Facilidade de Acesso, Segurança de Acesso
Contextuais	Relevância, Valor-Agregado, Atualidade, Completude, Volume de Dados, Facilidade de Uso
Representação	Interpretabilidade, Facilidade de Entendimento, Representação Concisa, Representação Consistente

• Objetivas

• Subjetivas

Conceitos de TDQM

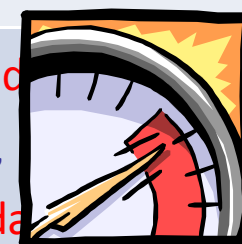


As 16 Dimensões da Qualidade da Informação

Categoria	Dimensões
Intrínsecas	Integridade, Objetividade, Credibilidade, Reputação
Acessibilidade	Facilidade de Acesso, Segurança de Acesso
Contextuais	Relevância, Valor-Agregado, Atualidade, Completude, Volume de Informação, Qualidade
Representação	Clareza, Precisão, Facilidade de Entendimento, Representação Concisa, Representação Consistente

• Objetivas

• Subjetivas



Aqui entram as métricas

Conceitos de TDQM



**Qualidade de Dados deve ser adequada ao uso
("fitness for use")**



flagrante



viagem

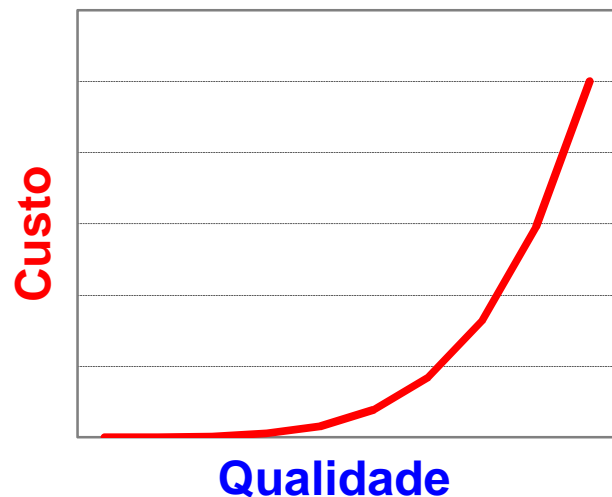


trabalho

Conceitos de TDQM



Qualidade de Dados deve ser adequada ao uso ("fitness for use")



Escolher:

- as informações mais relevantes
- as dimensões mais importantes para o negócio
- o nível de qualidade suficiente

Escolha das Informações e Dimensões




- ☐ Identificação dos Stakeholders
 - Entendimento das necessidades
 - Ótica gerencial e operacional
- ☐ Seleção das Informações
 - Relevantes para o negócio
- ☐ Seleção das Dimensões
 - Relevantes para o negócio
 - Subjetivas: interpretáveis pelos Stakeholders
 - Objetivas: mensuráveis por processos automáticos

Métricas Subjetivas

- ❑ Seleção das dimensões
 - Não mais que 3 ou 4
- ❑ Definição de escala de avaliação
 - Simplicidade: ótimo, bom, regular, ruim, péssimo
- ❑ Pesquisa com os Stakeholders
 - Envolver os 3 C's: Collectors, Custodians, Consumers
 - Considerar a diferença de visões:
 - Necessidades e pontos de vista
 - Dados de repositórios diferentes ou com transformações diferentes
 - Momentos diferentes do ciclo de vida do dado




Métricas Subjetivas

Exemplo de Seleção e Escala de Avaliação

Classificação	Segurança	Credibilidade	Facilidade de uso
	Segurança de acesso inviolável ou eficaz	Informação provém sempre ou quase sempre de fonte com rigor na coleta	Informação utilizada com frequência, com ótima ou boa facilidade de manipulação
	Segurança de acesso com falhas	Informação provém predominantemente de boas fontes, mas parcela considerável vem de fontes ruins	Informação com uso regular, exige algum esforço para sua utilização
	Segurança de acesso sem controle ou inexistente	Informação provém predominantemente de fontes com baixa qualidade ou sem preenchimento	Informação com pouco uso, exige muito esforço para sua utilização

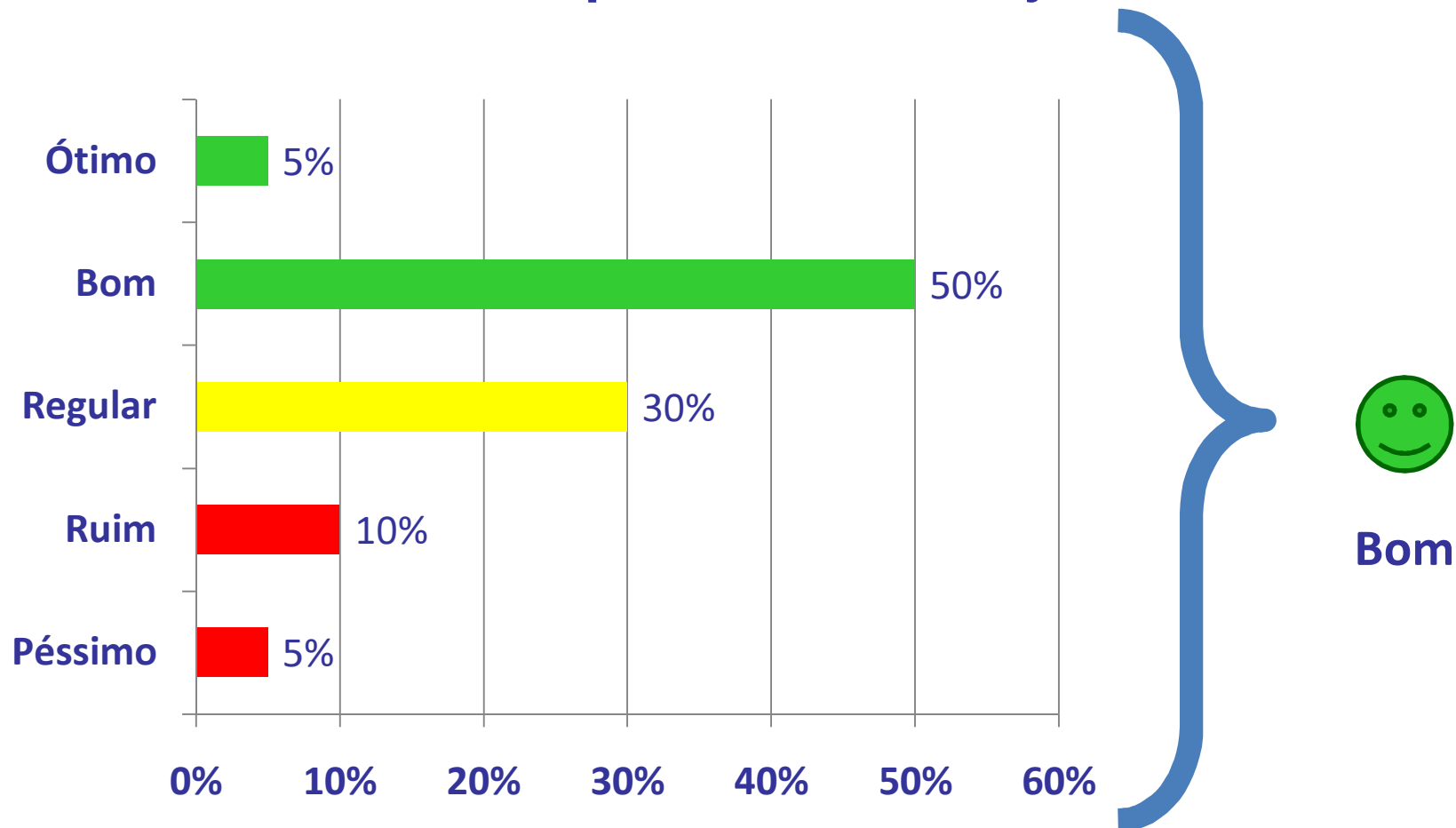
Métricas Subjetivas

Exemplo de Avaliação de um Stakeholder

Informação	Segurança	Credibilidade	Facilidade de Uso
CPF			
Nome			
Endereço			
Renda Comprovada			
Profissão			
Último Produto			
Qt. Compras no Ano			

Métricas Subjetivas

Exemplo de Avaliação Geral de uma Dimensão para uma Informação



Métricas Objetivas

- ❑ Identificação de Regras Mensuráveis
 - Preenchimento (Compleitude)
 - Data de captação (Atualidade)
(coleta ou último uso com sucesso)
 - Integridade de domínio, coluna, entidade, referencial
(Integridade, Representação Consistente)
 - Regras de negócio (Consistência)

- ❑ Definição de uma métrica para cada regra → métricas individuais

- ❑ Combinação das métricas individuais → métrica combinada = score de qualidade

Métricas Objetivas

Exemplo de Regras e Métricas Individuais

Informação: DATA DE NASCIMENTO

Regra	Métrica	Resultado
Validar preenchimento	M1 – Preenchido	Sim / Não
Calcular data da última captação	M2 – Idade da Última Captação	Idade da captação
Validar sintaxe e existência	M3 – Data Válida	Sim / Não
Validar intervalo permitido	M4 – Idade Válida	Sim / Não
Validar nível de repetição	M5 – Data Viciada	Sim / Não
Validar idade mínima versus negócio	M6 – Idade Mínima Válida	Válida Suspeita Inválida

Métricas Objetivas

Exemplo de Métrica Combinada

Informação: DATA DE NASCIMENTO

Métrica Combinada	Resultado
Se Idade Mínima Válida, no Intervalo Válido e Não Viciada	10
Se Idade Mínima Suspeita, no Intervalo Válido e Não Viciada	7
Se Data Válida e Não Viciada, mas Idade Fora do Intervalo ou Idade Mínima Inválida	5
Se Data Válida mas Viciada	2
Se Data Não Preenchida ou Inválida	0

Métricas Objetivas




Atualidade - Fator de Redução por Data de Captação

Informação: ENDEREÇO

Tabela de Redução Pela Idade da Última Captação	Fator Redutor
Menos de 1 ano	1
De 1 a 3 anos	0,8
De 4 a 6 anos	0,7
De 7 a 9 anos	0,6
De 10 anos em diante	0,5

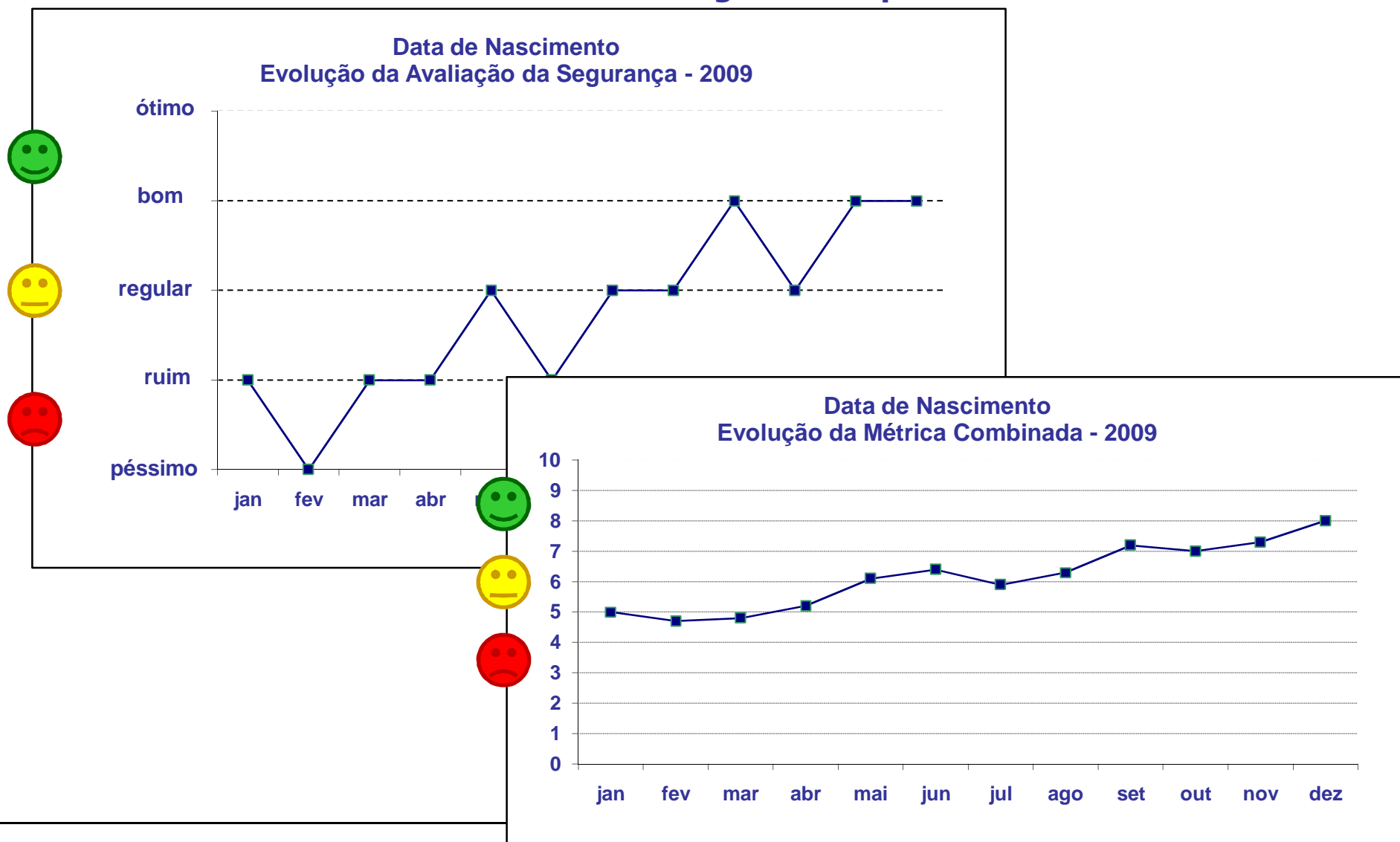
Métricas Objetivas

Metas Para as Métricas Combinadas

INFORMAÇÃO			
CPF	De 9 a 10	De 7 a 8,9	Menor que 7
Nome	De 9 a 10	De 7 a 8,9	Menor que 7
Endereço	De 7 a 10	De 5 a 6,9	Menor que 5
Renda Comprovada	De 7 a 10	De 5 a 6,9	Menor que 5
Profissão	De 7 a 10	De 5 a 6,9	Menor que 5
Último Produto	De 9 a 10	De 7 a 8,9	Menor que 7
Qt. Compras no Ano	De 9 a 10	De 7 a 8,9	Menor que 7

Métricas Subjetivas e Objetivas

Medir ao Longo do Tempo



Exemplos de Ações para Melhorar a Qualidade de Dados

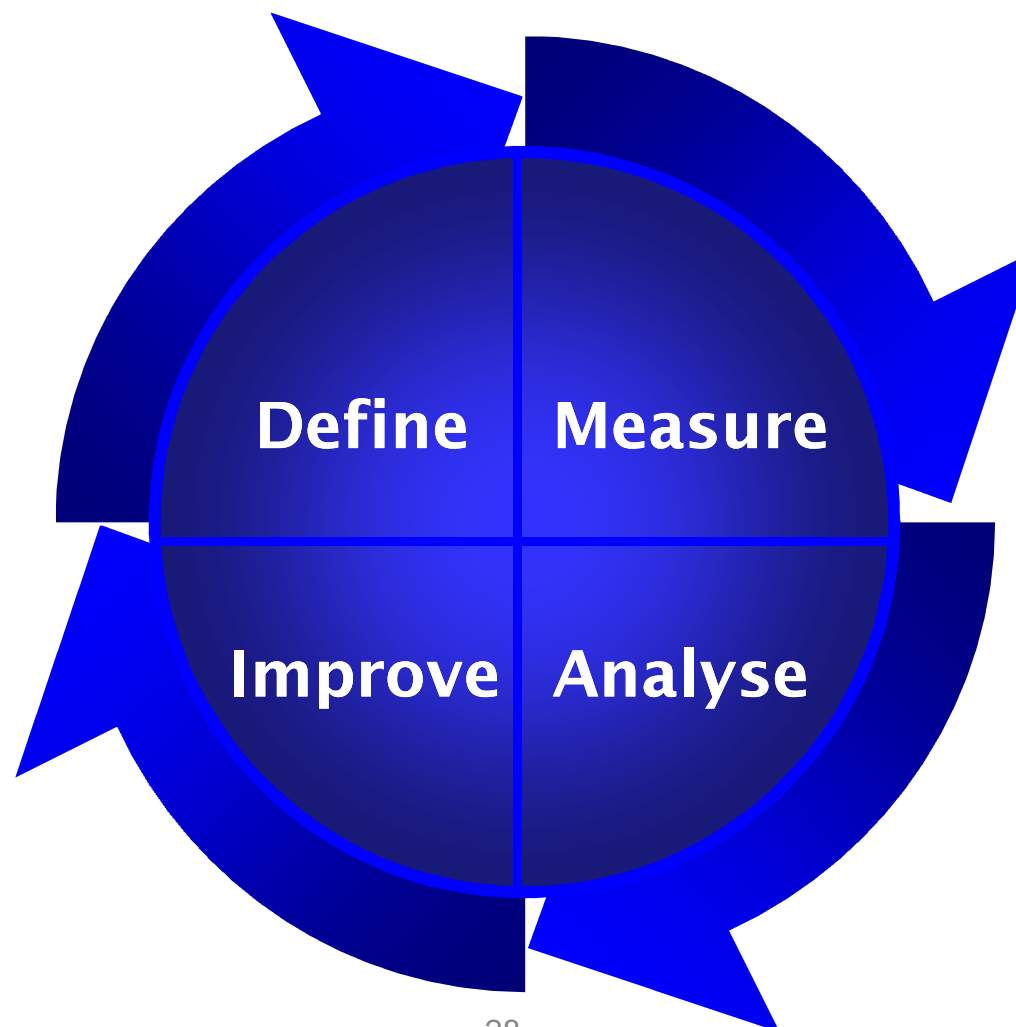
- ☐ Treinamento da equipe
- ☐ Validação de regras na coleta de dados
- ☐ Criação de processos de correção automática de dados
- ☐ Tratamento de dados por back-office
- ☐ Enriquecimento de dados com outras bases internas e externas

Avaliar os Resultados

- ☐ Monitorar a incidência dos problemas que motivaram o programa
- ☐ Aferir o retorno tangível e intangível



Qualidade de Dados é um Processo



Obrigado !

Mario F. Cervi
Assesso Engenharia de Sistemas
mario.cervi@assesso.com.br
mario.cervi@qibras.org
(11) 4195-5535