



UNIVERSIDADE FEDERAL
DE SANTA CATARINA

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Big Data

Acesso e organização de dados abertos (Inteligência Analítica)

Gabriela Colonetti
Alexandre Marquez

Roteiro

- Problema
- Objetivo
- Big Data
- Dados Abertos
- Plano de Dados
- Exemplos de Plano de Dados e Dados Abertos
- Aplicações

Problema

Os dados são abertos mas o acesso e organização depende de planejamento e desenvolvimento de aplicações.

Objetivo

Relacionar as tecnologias de Big Data para acesso e organização de dados abertos

O que é o Big Data

- Big Data é usado para descrever um enorme volume de dados estruturados e não estruturados que, por serem tão grandes, é muito difícil processá-los usando uma base de dados e técnicas de software tradicionais.
- Refere-se a um grande armazenamento de dados e maior velocidade. Diz-se que o Big Data se baseia em 5 "V" : velocidade, volume, variedade, veracidade e valor.

Características

- **Volume.** Dados coletados de uma grande variedade de fontes.
- **Velocidade.** Os dados fluem em uma velocidade sem precedentes e devem ser tratados em tempo hábil.
- **Variedade.** Todos os tipos de formatos - de dados estruturados, dados numéricos em bancos de dados tradicionais, até documentos de texto não estruturados.
- **Veracidade.** É preciso separar o que é rico em conteúdo em meio a tanta informação, ao garantir isso, sobram apenas os dados que possuem valor para um negócio.
- **Valor.** Geração de valor para os resultados que retornam do Big Data. Enriquecimento da base de dados já existente.

Quem utiliza

- **Ensino:** Podem implementar um sistema melhor de avaliação.
- **Saúde:** Registros de pacientes. Planos de tratamento. Informações de prescrição.
- **Manufatura:** Os fabricantes podem aumentar a qualidade e a produção, minimizando o desperdício.
- **Varejo:** Os varejistas precisam saber a melhor maneira de vender aos clientes, a maneira mais eficaz de lidar com transações, e a maneira mais estratégica de aumentar o número de negócios repetidos

Big Data Analytics

- Processos de Big Data;
 - Gerenciamento do Big Data
 - Big Data Analytics
 - Big Data Analytics compreende modelagem, análise e interpretação dos dados.
 - Existem técnicas em BI como mesmo propósito mas que não estão preparadas para a variedade de dados em cenários de Big Data.
 - Análises de textos, redes sociais, áudios, vídeos e preditivas.
 - Dados não estruturados, estruturados e semi-estruturados.

Dados Abertos

Em 2009, o tema começou a aparecer na grande mídia, com governos de vários países (como Estados Unidos, Reino Unido, Canadá e Nova Zelândia) anunciando iniciativas voltadas a abrir sua informação pública.

Brasil

Política de Dados Abertos do poder Executivo Federal, instituída pelo Decreto nº 8.777 de 11 de maio de 2016, tem por objetivo, entre outros, 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 e informações acessíveis ao público, visando o aprimoramento da cultura de transparência pública.

Definição

Segundo a definição da **Open Knowledge Internacional**, em suma, dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura.

Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta.

Características

O especialista em políticas públicas e ativista dos dados abertos David Eaves propôs as seguintes “leis”:

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

As leis foram propostas para os Dados Abertos Governamentais, mas pode-se dizer que elas se aplicam aos Dados Abertos de forma geral.

Princípios

- Completos
- Primários
- Atuais
- Acessíveis
- Processáveis por máquinas
- Acesso não discriminatório
- Formatos não proprietários
- Livres de licença

As principais fontes de dados abertos

- Dados abertos na ciência
- Dados abertos no governo
- Dados abertos de empresas privadas
- Dados abertos nas universidades

Potencial

- O leitor sabe exatamente quanto dos seus impostos é gasto com iluminação das ruas ou pesquisas contra o câncer?
- Qual é o caminho mais curto, mais seguro e com a melhor vista da sua casa até o seu trabalho?
- E qual é a qualidade do ar que respira ao longo desse trajeto?
- O leitor sabe onde, na sua região, pode encontrar as melhores oportunidades de emprego e os locais mais arborizados?
- Sabe quando e como influenciar leis ou decisões públicas sobre temas com os quais se preocupa?

Áreas em destaque

- Participação popular;
- Empoderamento dos cidadãos;
- Melhores ou novos produtos e serviços privados;
- Inovação;
- Melhora na eficiência de serviços governamentais;
- Melhora na efetividade de serviços governamentais;
- Medição do impacto das políticas;
- Conhecimento novo a partir da combinação de fontes de dados e padrões.

Acesso e organização de dados abertos (Inteligência Analítica)

- Diversas Instituições, sistemas e integrações entre esses
- Desafio de acesso e coleta.
 - Diversas fontes, tipos de arquivos e APIs
- Desafio de Organização
 - Arquivos estáticos de anos anteriores
 - Sistemas de Análises de Dados, tecnologias Big Data, NoSQL, Business Intelligence e desenvolvimento de sites e aplicativos
- Inteligência Analítica, coletar dados e processá-los para suporte a decisões, resultados e estatísticas.

Planejamento

Os primeiros desafios que muitas organizações enfrentam quando se tornam divulgadoras de dados abertos são técnicos, em grande parte centrados em como publicar dados em formatos fáceis de usar, com claro licenciamento.

Compreender esta mudança organizacional é um aspecto importante para ser um divulgador de dados abertos eficiente.

Publicação e reutilização

[Guia do Modelo de Maturidade de Dados Abertos: Avaliando a publicação e utilização de dados abertos](#)

Plano de Dados

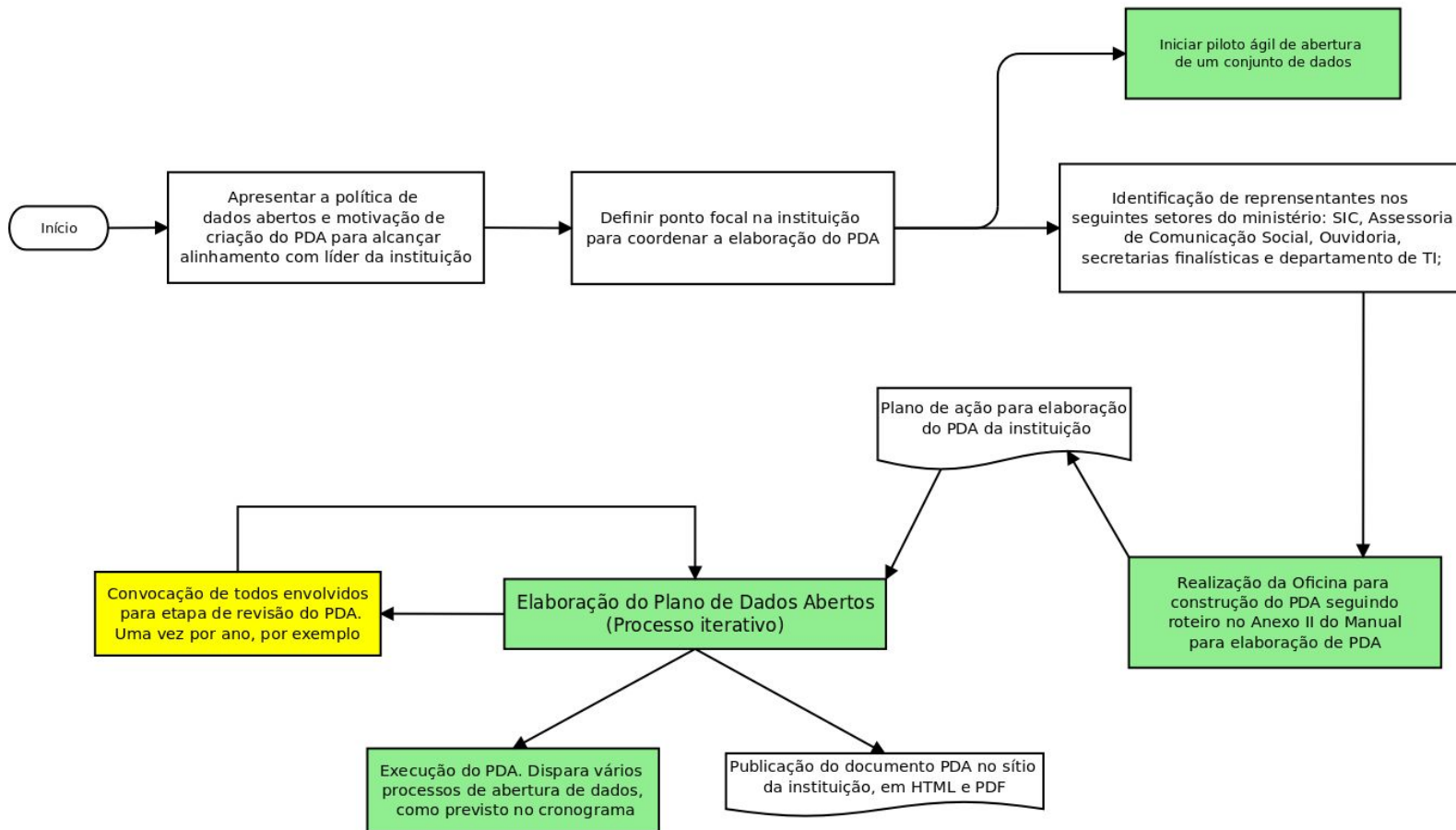
O Plano de Dados Abertos (PDA) orienta as ações de implementação e promoção de abertura de dados, inclusive geoespacializados, de uma organização.

O documento faz parte da **política de dados abertos**, e seu objetivo é organizar e padronizar os processos de publicação de dados abertos do estado, resultando na melhoria da qualidade dos serviços prestados ao cidadão e à própria administração pública.

Processo Sistêmico

- Apresentar a política de dados abertos
- Definir ponto focal para coordenar o trabalho
- Processo Piloto Ágil
- Oficina para construção do PDA da instituição
- Plano de ação para elaboração do PDA
- Revisão do PDA
- Publicação do PDA
- Execução do PDA

Processo sistêmico de um Plano de Dados Abertos - PDA



Exemplos de Planos de Dados e Dados Abertos

[Plano de Dados Abertos da UFG - vigência 2017/2018](#)

[Plano de Dados IBCT](#)

http://cnpq.br/dados_abertos

UFSC?

Aplicações

- Orçamento Federal
 - Aplicação para navegar no banco de dados de despesas do governo federal brasileiro no formato RDF.
- Basômetro
 - Ferramenta interativa que permite medir o apoio dos parlamentares ao governo e acompanhar como eles se posicionaram nas votações legislativas. Cada um é representado por uma...
- Saúde Acessível
 - Ferramenta ajuda o cidadão a encontrar um posto de saúde mais próximo.



Em **139** votações, **419** deputados votaram com o governo em **50%** das vezes ou mais; e **94** votaram com o governo em menos de **50%** das vezes

Todos **75%**

PT	97%
PTC	97%
PRTB	91%
PCdoB	94%
PSB	90%
PP	88%
PSL	89%
PMDB	90%
PTB	84%
PRB	87%
PSD	79%
PSC	81%
PTdoB	79%
PRP	79%
PMN	76%
PDT	82%
PHS	82%
PR	78%
PV	75%
PPS	37%
PSOL	40%
DEM	30%
PSDB	27%
PEN	55%

Por bancada

RURALISTAS
EVANGÉLICOS
AMBIENTALISTAS

GOVERNO

100%

%



PASTOR
MARCO
FELICIANO
PSC SP

taxa de
governismo:

82%

votos/governo

contra: a favor:
21 95

abstenção: 0
não votou: 23

Taxa de apoio ao governo: 0%

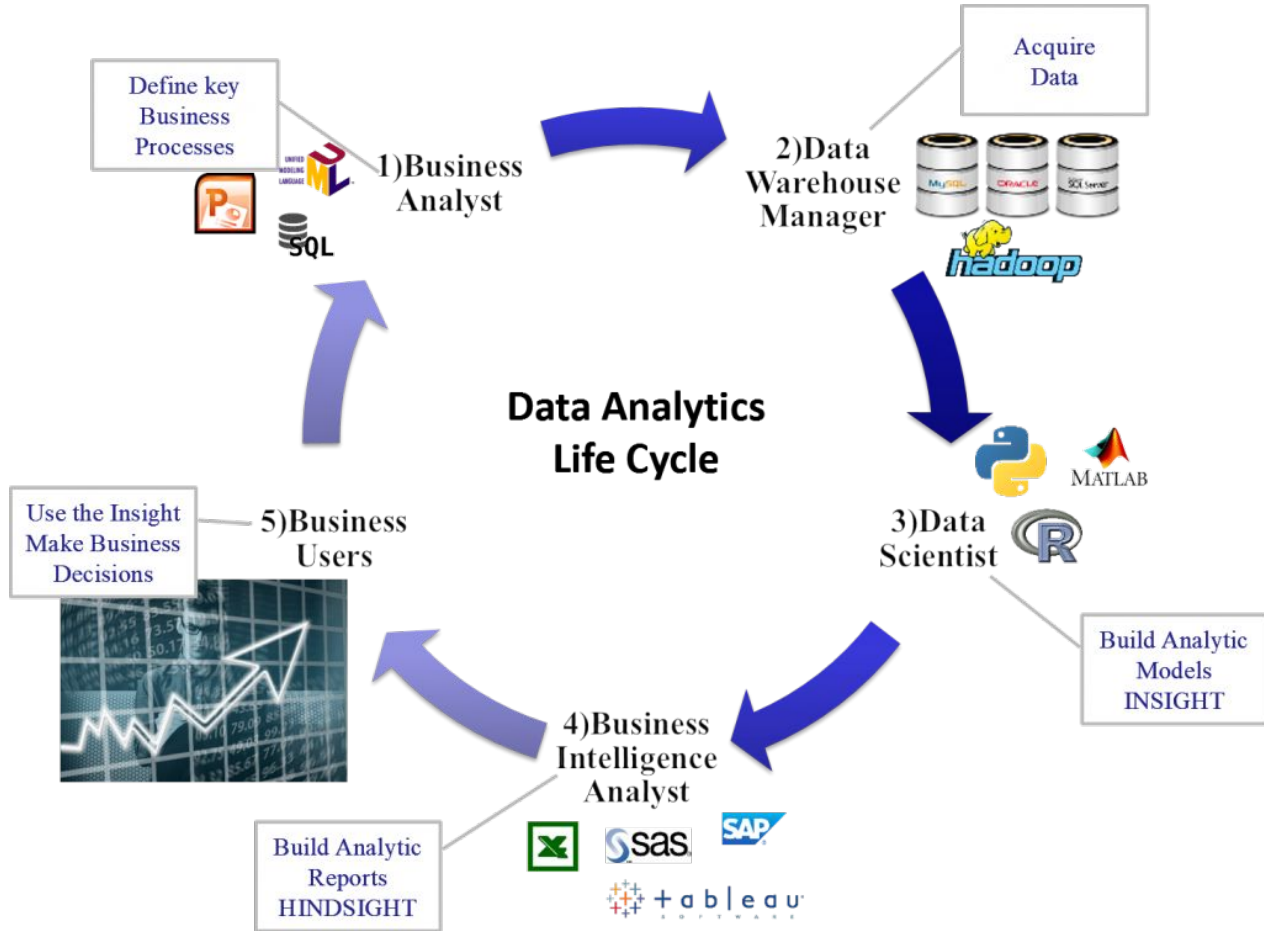


139 - 3/4/2013 17h36 - null (5263)

AC
AM
AP
PA
RO
RR
TO
AL
BA
CE
MA
PB
PE
PI
RN
SE
DF
GO
MS
MT
ES
MG
RJ
SP
PR
RS
SC

Proposta

- Descrever arquiteturas de Big Data relacionadas com acesso e organização de dados abertos
 - Coleta e armazenamento de dados abertos
 - Análise de dados abertos
 - Publicação de dados abertos
- Ferramentas para estatísticas e extração de dados por meio de ferramentas de Business Intelligence .
- Reutilização de dados abertos processados
- Big Data permite criar uma base universal para dados abertos



Ferramentas e Tecnologias Big Data

- Hadoop
 - MapReduce
 - Cluster
 - NoSQL
- Análise de Dados
 - Real Time
 - Machine Learning
 - Deep Learning
 - Preditiva

Referências

Portal Brasileiro de Dados Abertos <<http://dados.gov.br>>. Acessado em 15/08/2017

Plano de Dados Abertos (PDA) <<http://wiki.dados.gov.br/Plano-de-Dados-Abertos.ashx>>. Acessado em 15/08/2017

Processo sistêmico de um Plano de Dados Abertos < <http://kit.dados.gov.br/Processo-sist%C3%AAmico/>> .Acessado em 15/08/2017