

Analisando grandes volumes de dados abertos

Daniel Panizzo

Trilha Inovação com dados em nuvem

Analisando grandes volumes de dados abertos

Daniel Panizzo

Trilha Inovação com dados em nuvem

Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhável 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.





github.com/dspanizzo



Daniel Panizzo
Engenheiro de Soluções

Oracle Public Sector



linkedin.com/in/dspanizzo

TDC
ONLINE

Sala
ORACLE

Inovação com dados em nuvem

TRILHA

#TheDevConf
Oracle





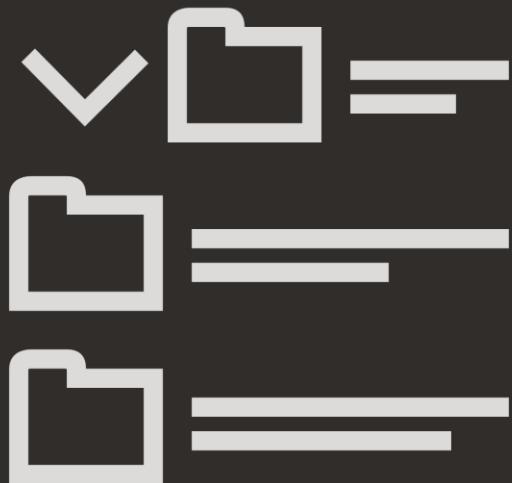
Open Knowledge
Foundation

“Dados abertos são dados que podem ser livremente usados, modificados e compartilhados por qualquer um para qualquer propósito.”

Open Knowledge Foundation

Fonte: [The Open Definition](#)

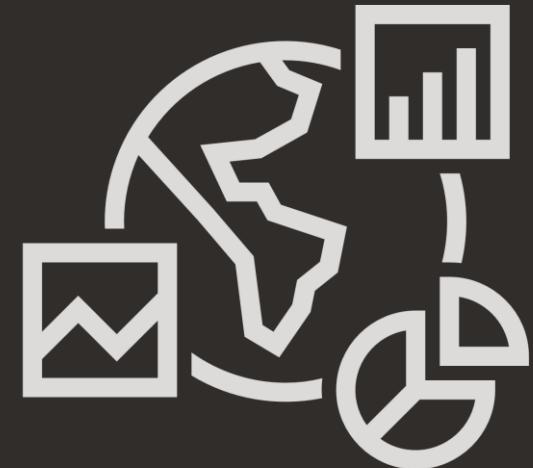
O que significa ser “aberto”?



Disponível e acessível



Reutilizável e distribuível



Participação universal

Tipos de dados abertos



Cultural



Estatístico



Clima



Científico



Financeiro



Ambiental

Open Government Partnership



“A Parceria de Governo Aberto é baseada na ideia de que um governo aberto é mais acessível, mais responsável e mais responsável pelos cidadãos, e que melhorar o relacionamento entre as pessoas e seu governo tem benefícios exponenciais de longo prazo para todos.”

Fonte: [Open Government Partnership - Approach](#)

CGU

CONTROLADORIA-GERAL
DA UNIÃO



Responsável pela gestão e monitoramento
das Políticas de Dados Abertos.

Fonte: [CGU – Acesso à informação – Dados Abertos](#)

Por quê dados abertos?



Transparência



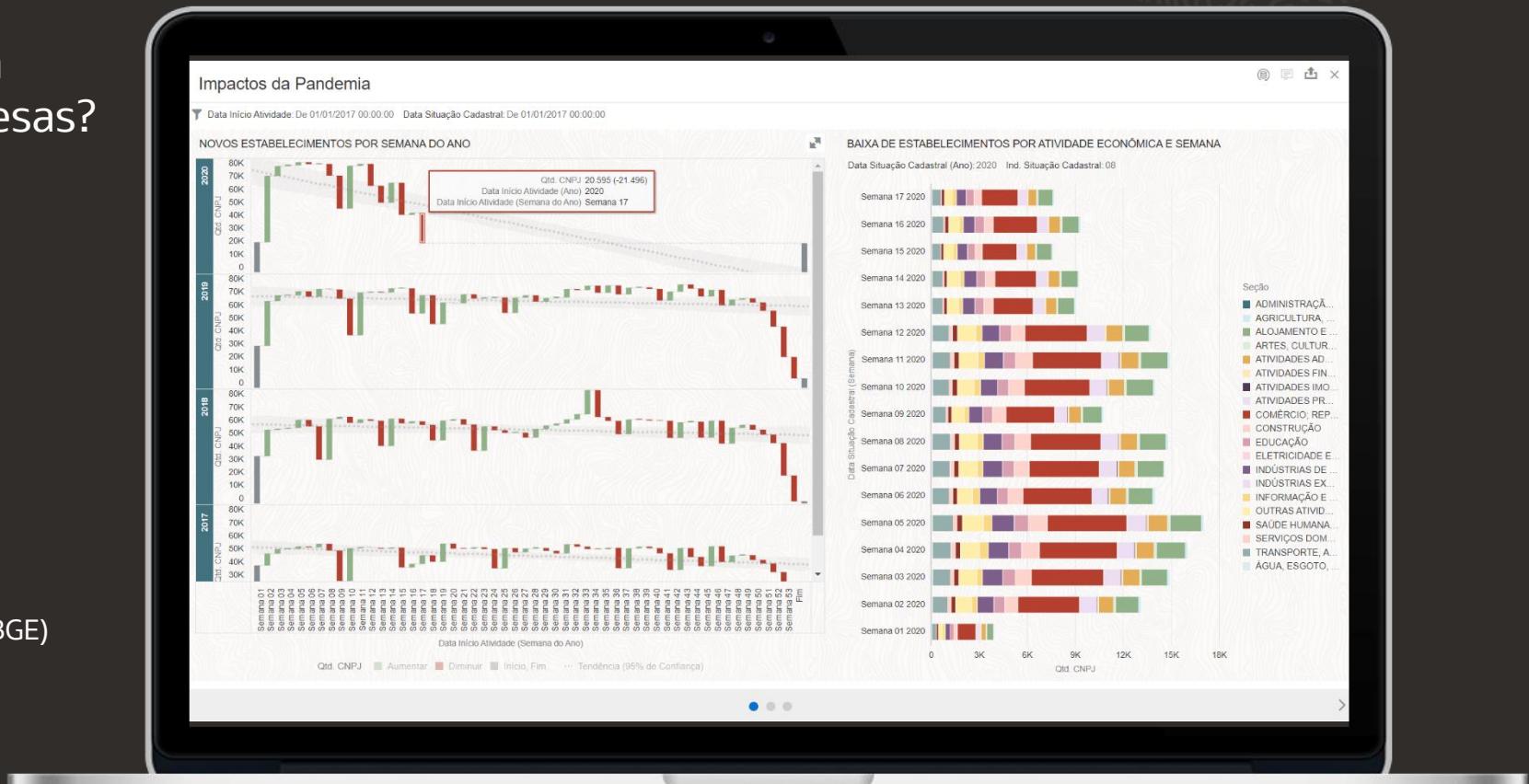
Valor social e comercial



Governança Participativa

Dados públicos do CNPJ

Qual o impacto da pandemia na abertura e baixa de empresas?



Fonte dos dados:

- Tribunal Superior Eleitoral (TSE)
- Instituto Brasileiro de Geografia e Estatística (IBGE)

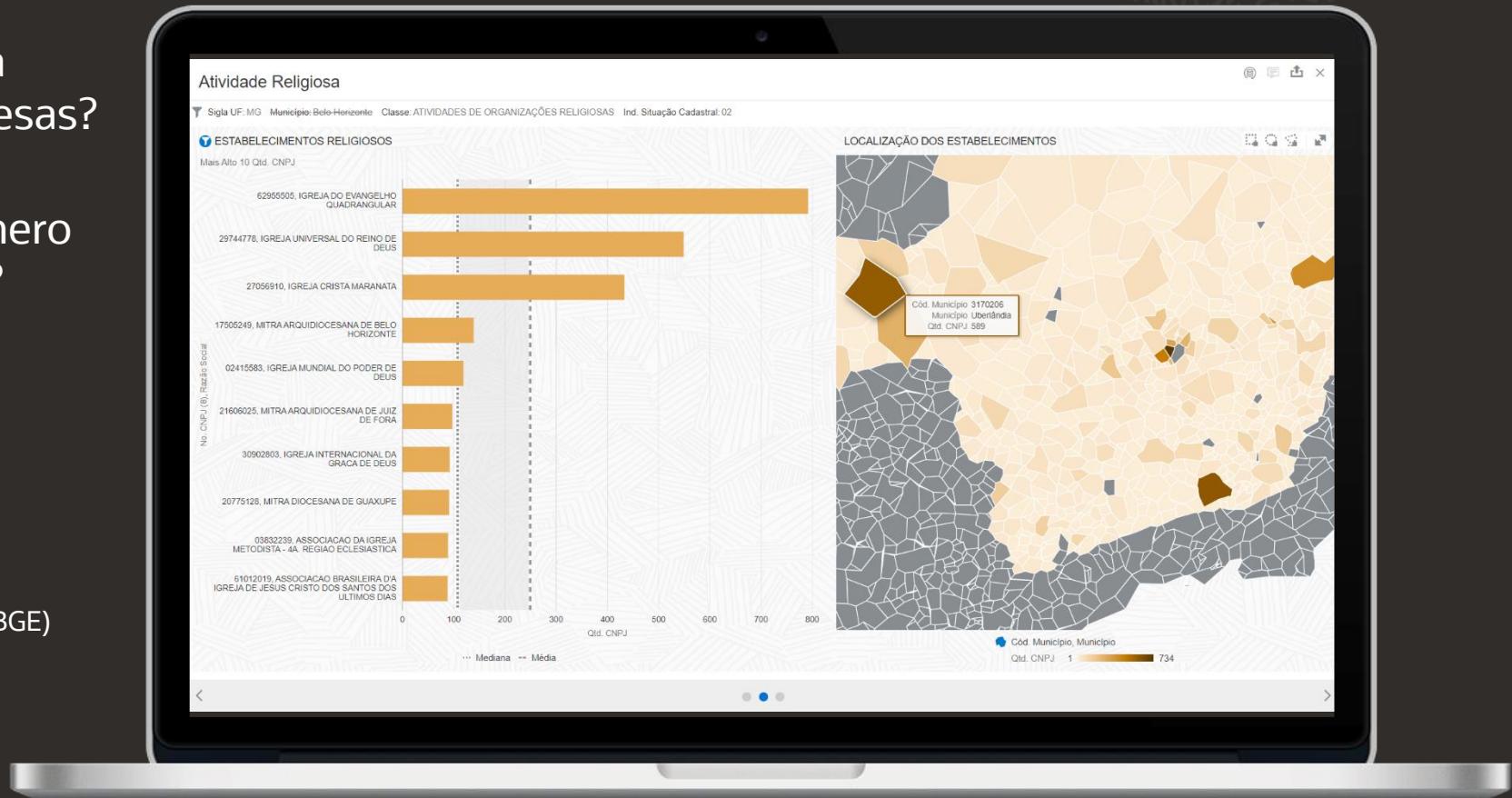
Dados públicos do CNPJ

Qual o impacto da pandemia na abertura e baixa de empresas?

Qual a igreja com maior número de estabelecimentos em MG?

Fonte dos dados:

- Tribunal Superior Eleitoral (TSE)
- Instituto Brasileiro de Geografia e Estatística (IBGE)



Dados públicos do CNPJ

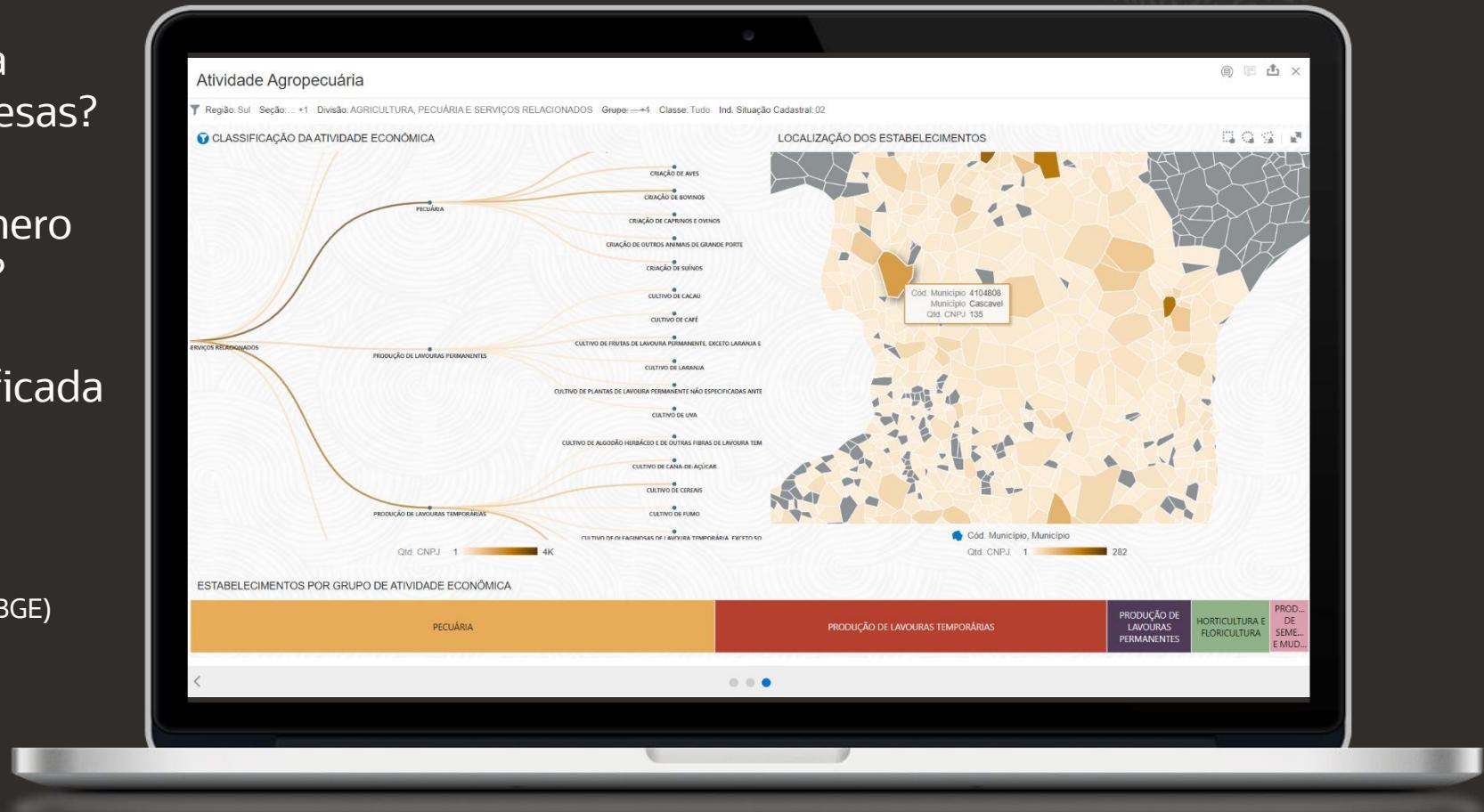
Qual o impacto da pandemia na abertura e baixa de empresas?

Qual a igreja com maior número de estabelecimentos em MG?

Como a agricultura é diversificada na região sul?

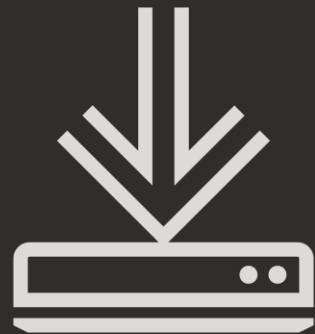
Fonte dos dados:

- Tribunal Superior Eleitoral (TSE)
- Instituto Brasileiro de Geografia e Estatística (IBGE)



Passo a passo

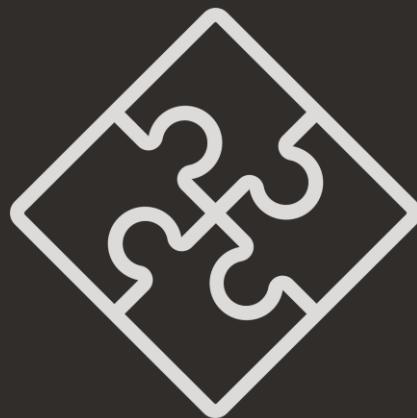
Visão geral



1. Coletar



2. Armazenar



3. Preparar



4. Analisar

Passo a passo

Origem dos dados



Receita Federal



Instituto Brasileiro
de Geografia e Estatística

Passo a passo

Linguagens



Python
(Anaconda)



SQL

Passo a passo

Opção 01: Ferramentas



JupyterLab



Object
Storage



Autonomous
Data Warehouse



Analytics
Desktop

Passo a passo

Opção 02: Ferramentas



Data Flow
(Apache Spark)



Object
Storage



Autonomous
Data Warehouse

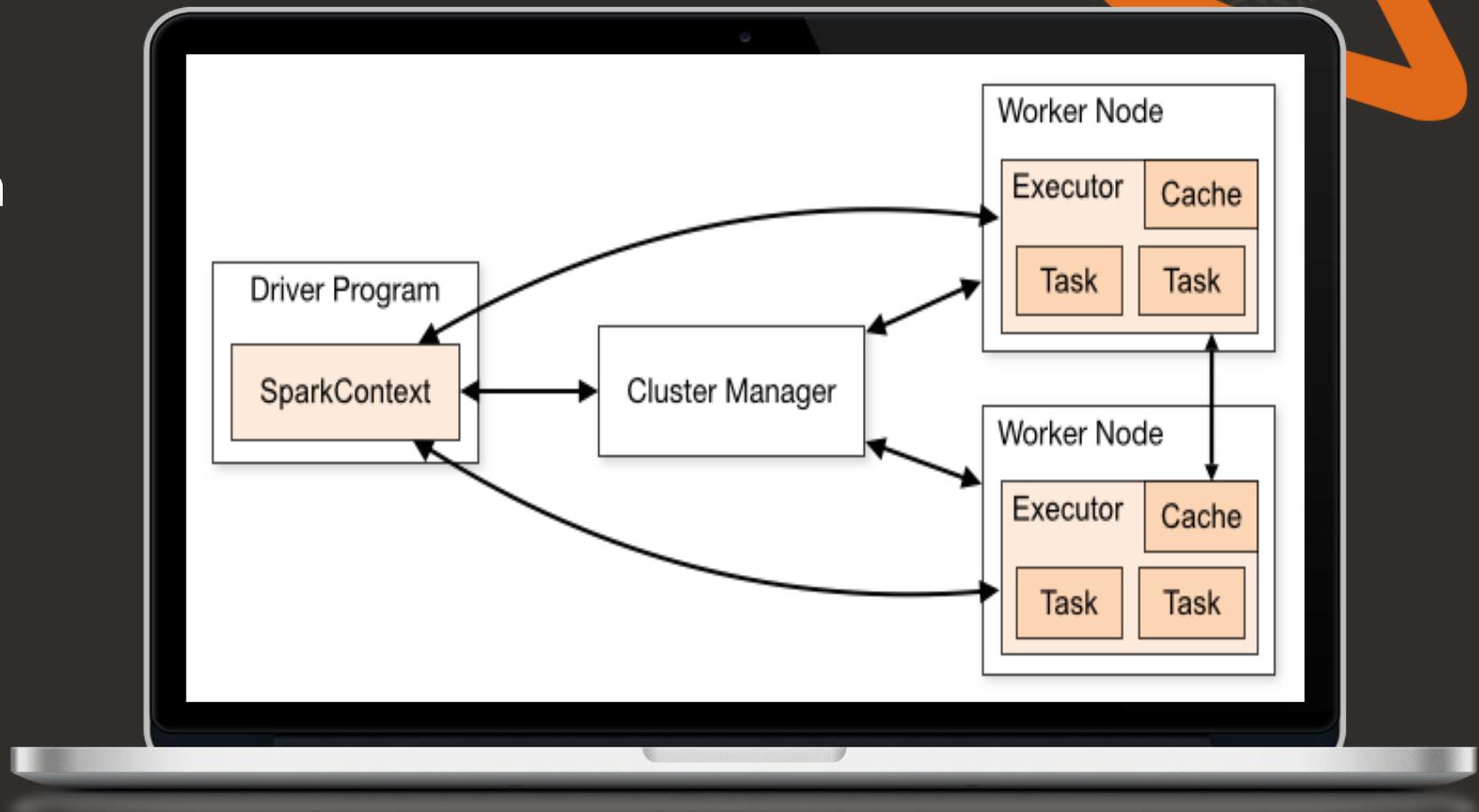


Analytics
Desktop

Passo a passo

Processamento de grandes volumes

“Apache Spark é um motor de análise unificado para processamento de dados em grande escala.”



Fonte: [Apache Spark](#)



Vamos trilhar este
caminho juntos?



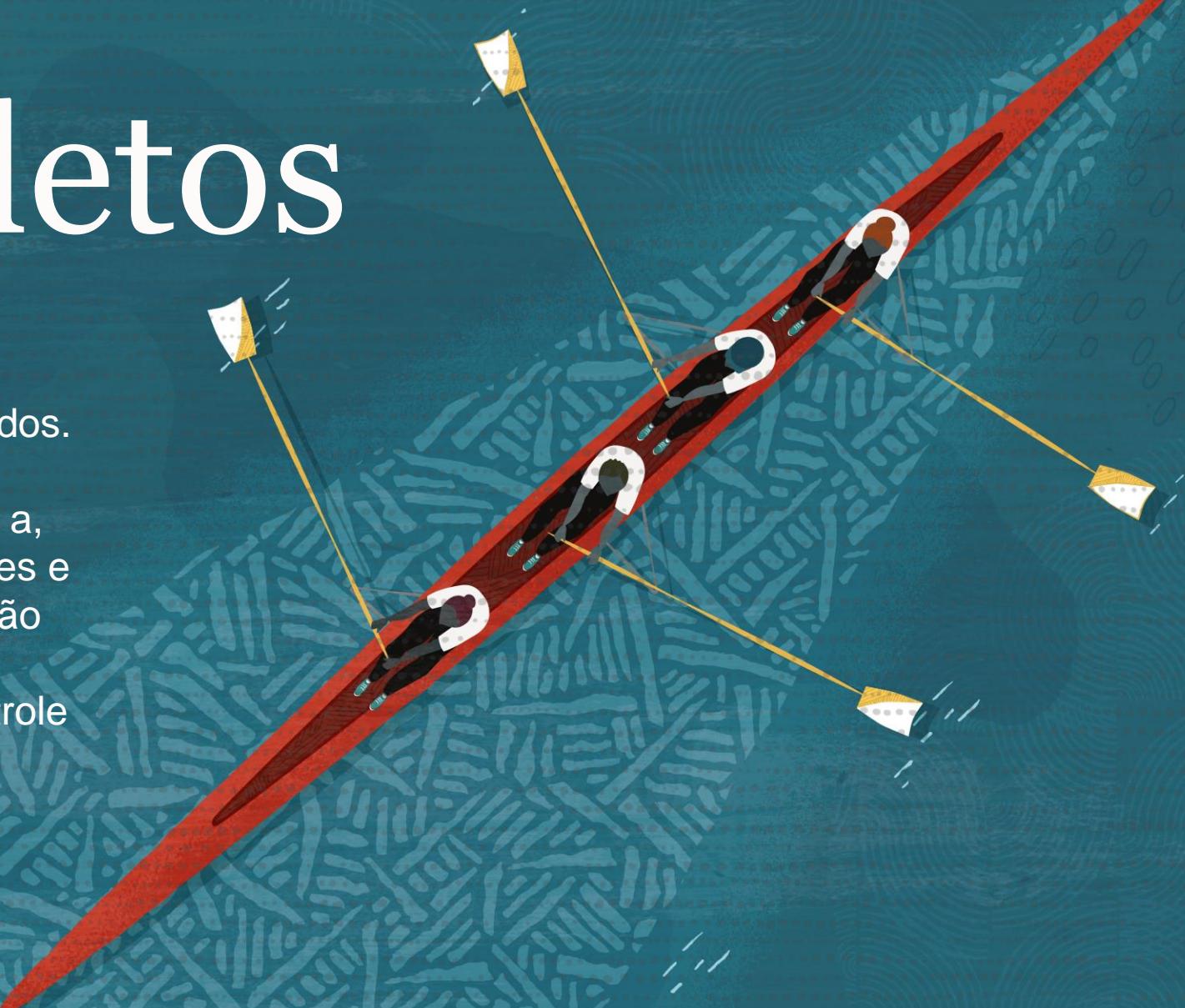
Parabéns!

8 Princípios dos Dados Abertos Governamentais



1 Completos

Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a, documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.



Fonte: Portal Brasileiro de Dados Abertos

2 Primários

Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.



Fonte: Portal Brasileiro de Dados Abertos

3 Atuais

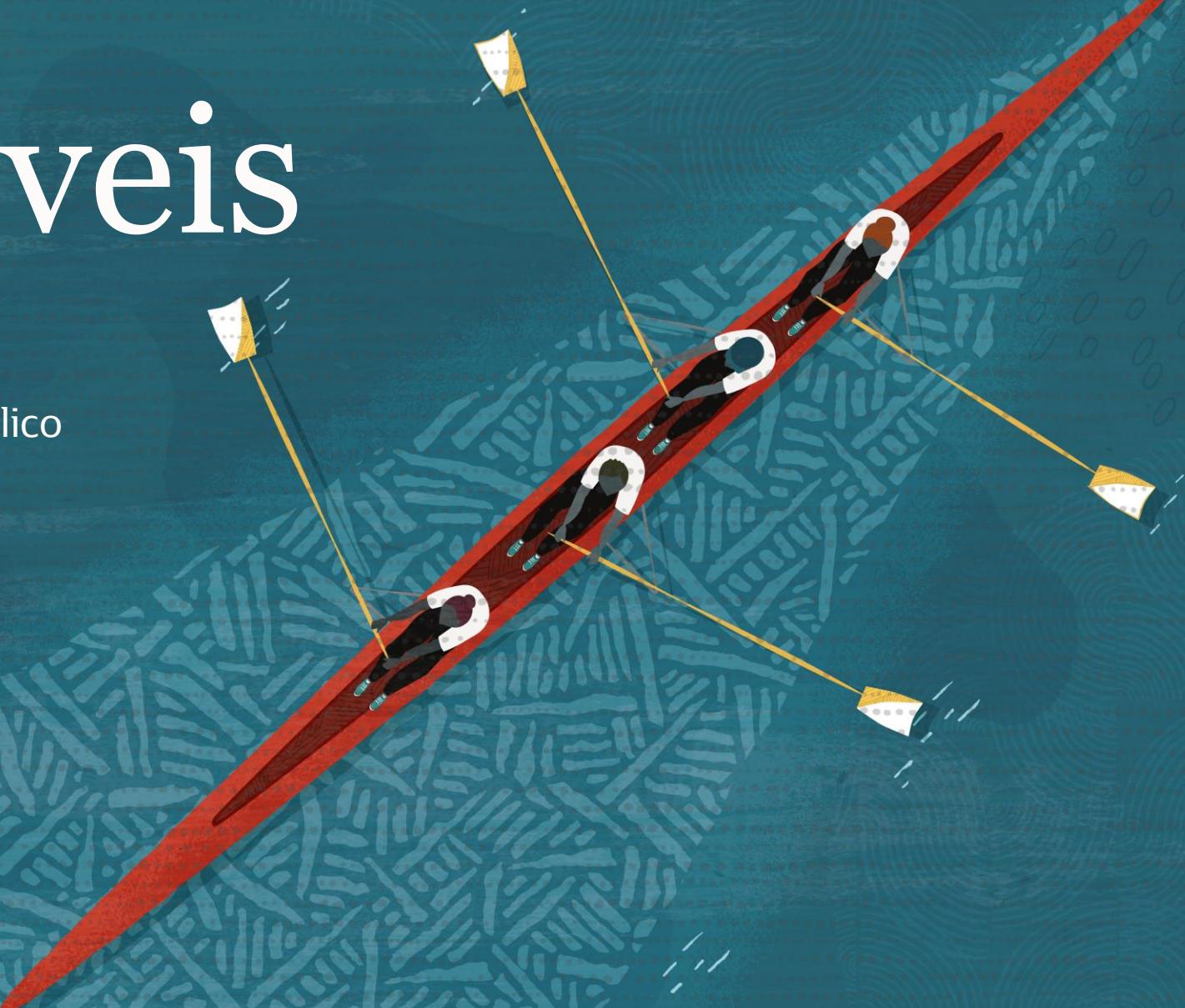
Os dados são disponibilizados o quanto rapidamente seja necessário para preservar o seu valor.



Fonte: Portal Brasileiro de Dados Abertos

4 Acessíveis

Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

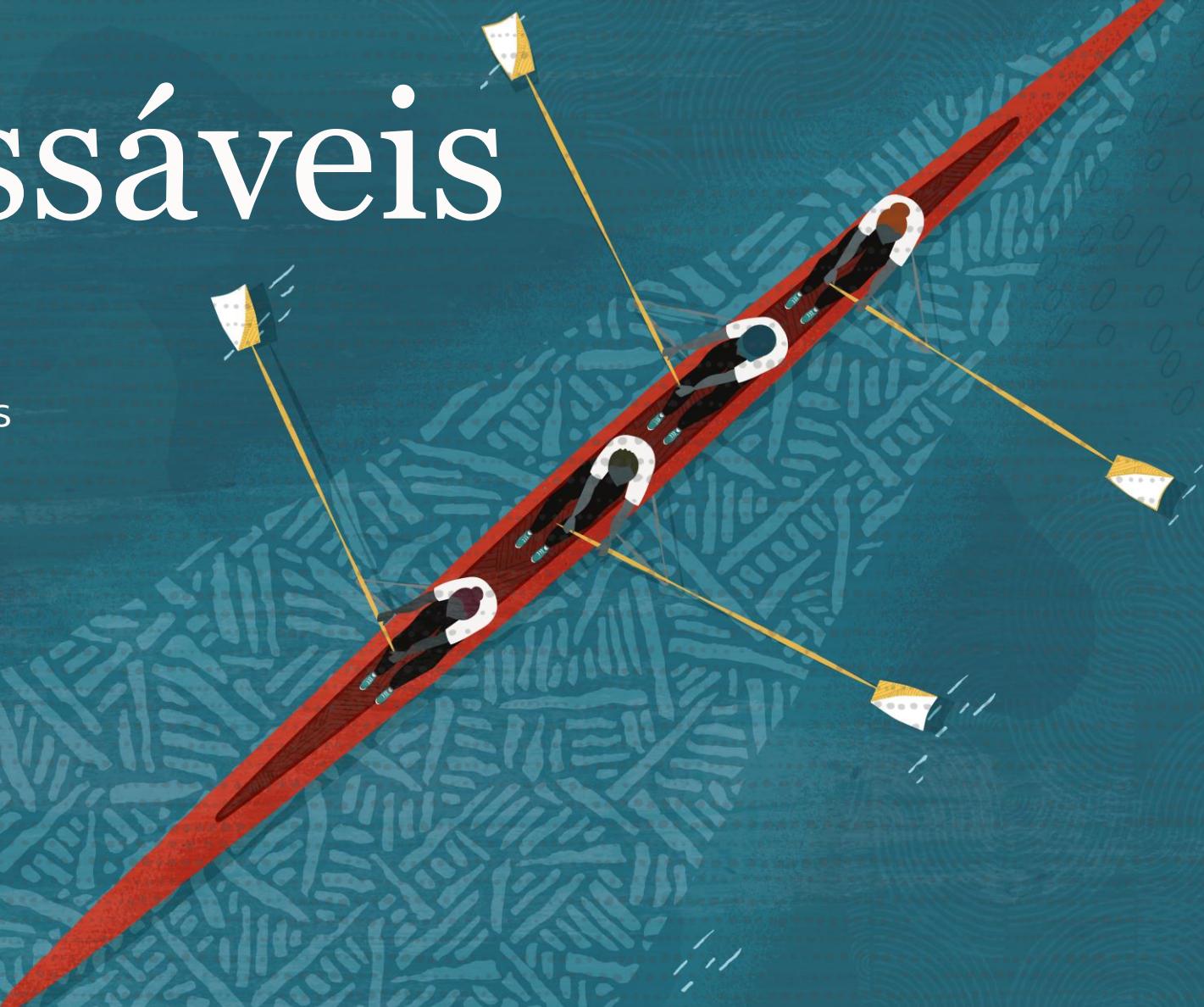


Fonte: Portal Brasileiro de Dados Abertos

5

Processáveis por máquina

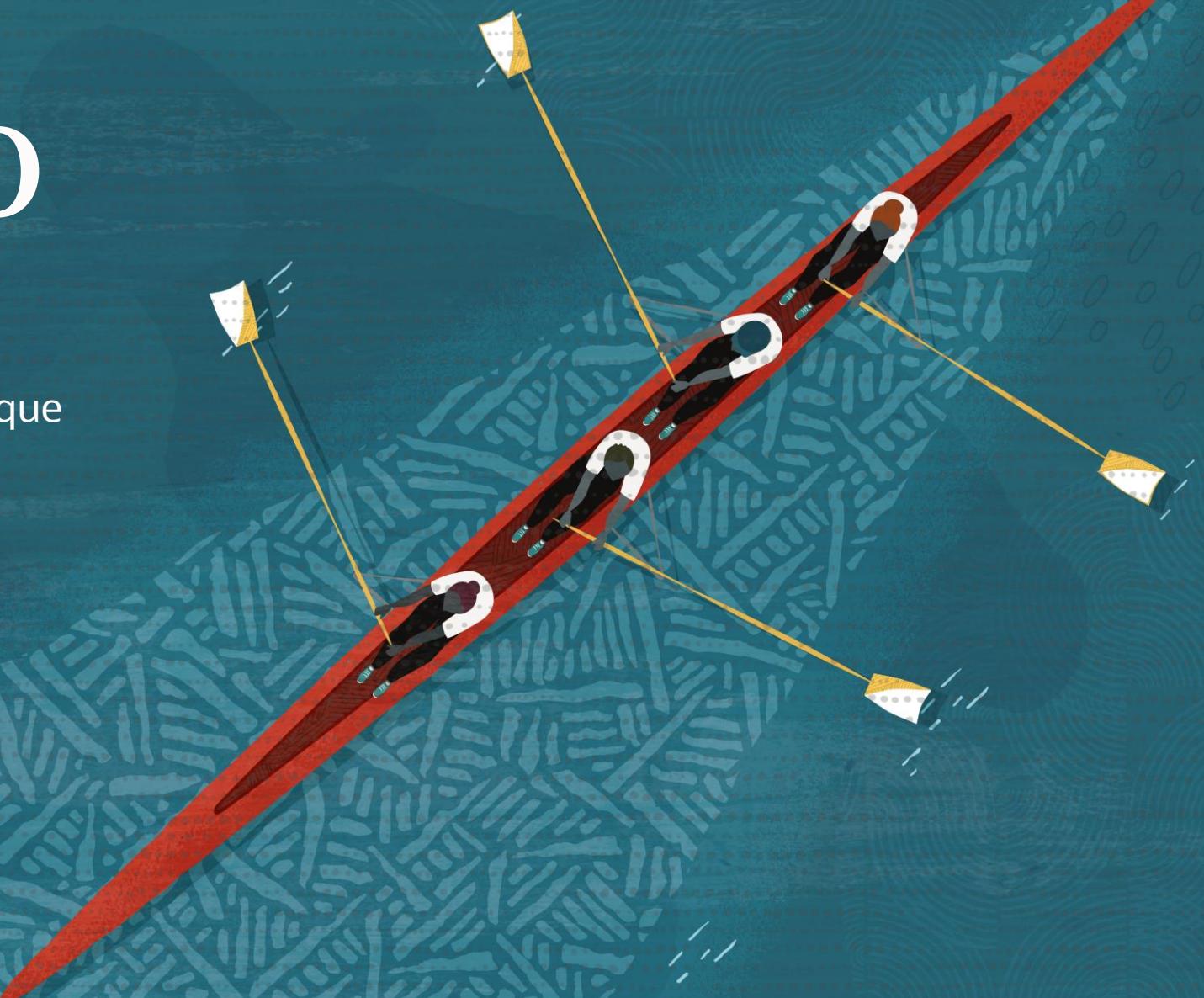
Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.



Fonte: Portal Brasileiro de Dados Abertos

6 Acesso não discriminatório

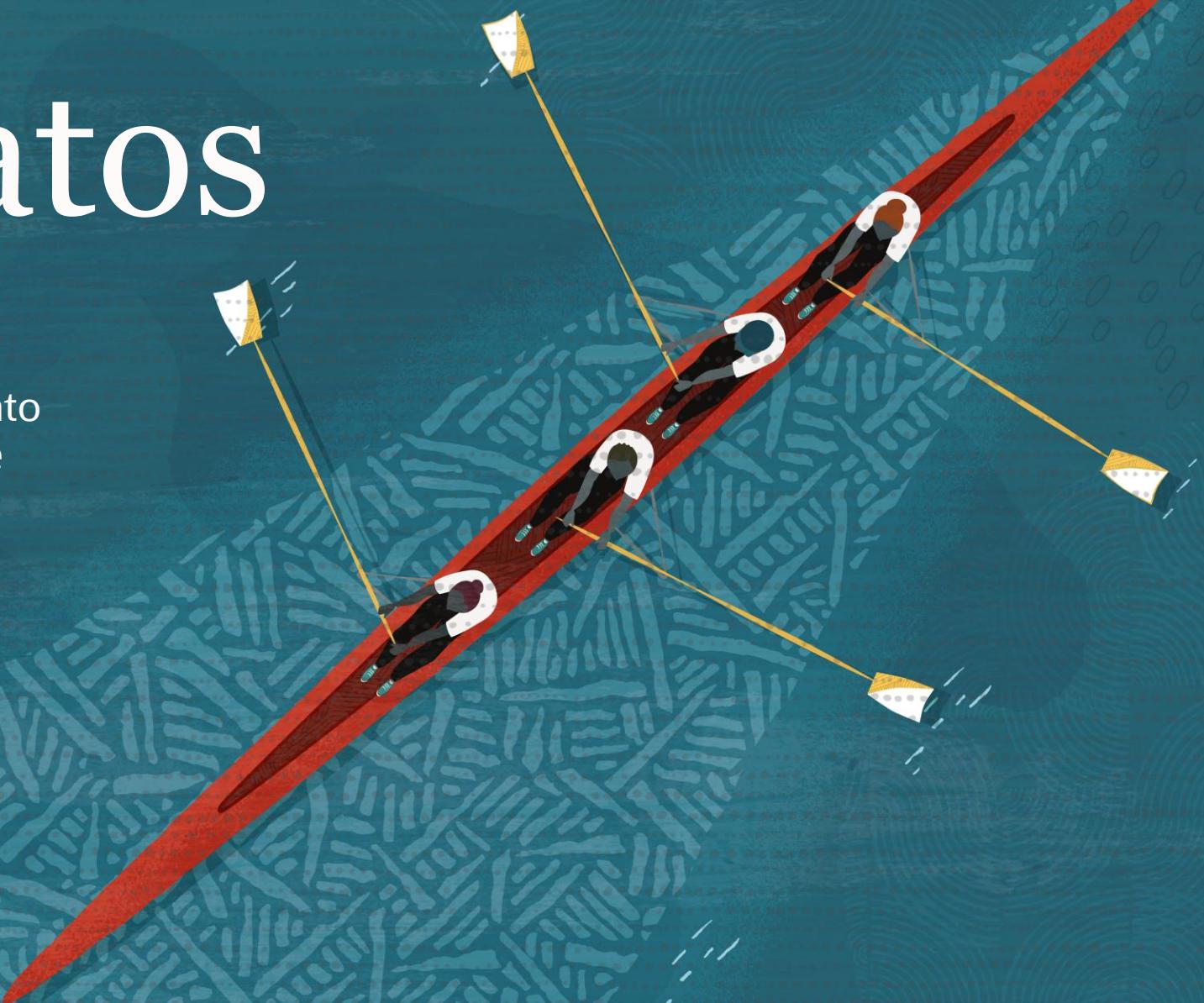
Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.



Fonte: Portal Brasileiro de Dados Abertos

7 Formatos não proprietários

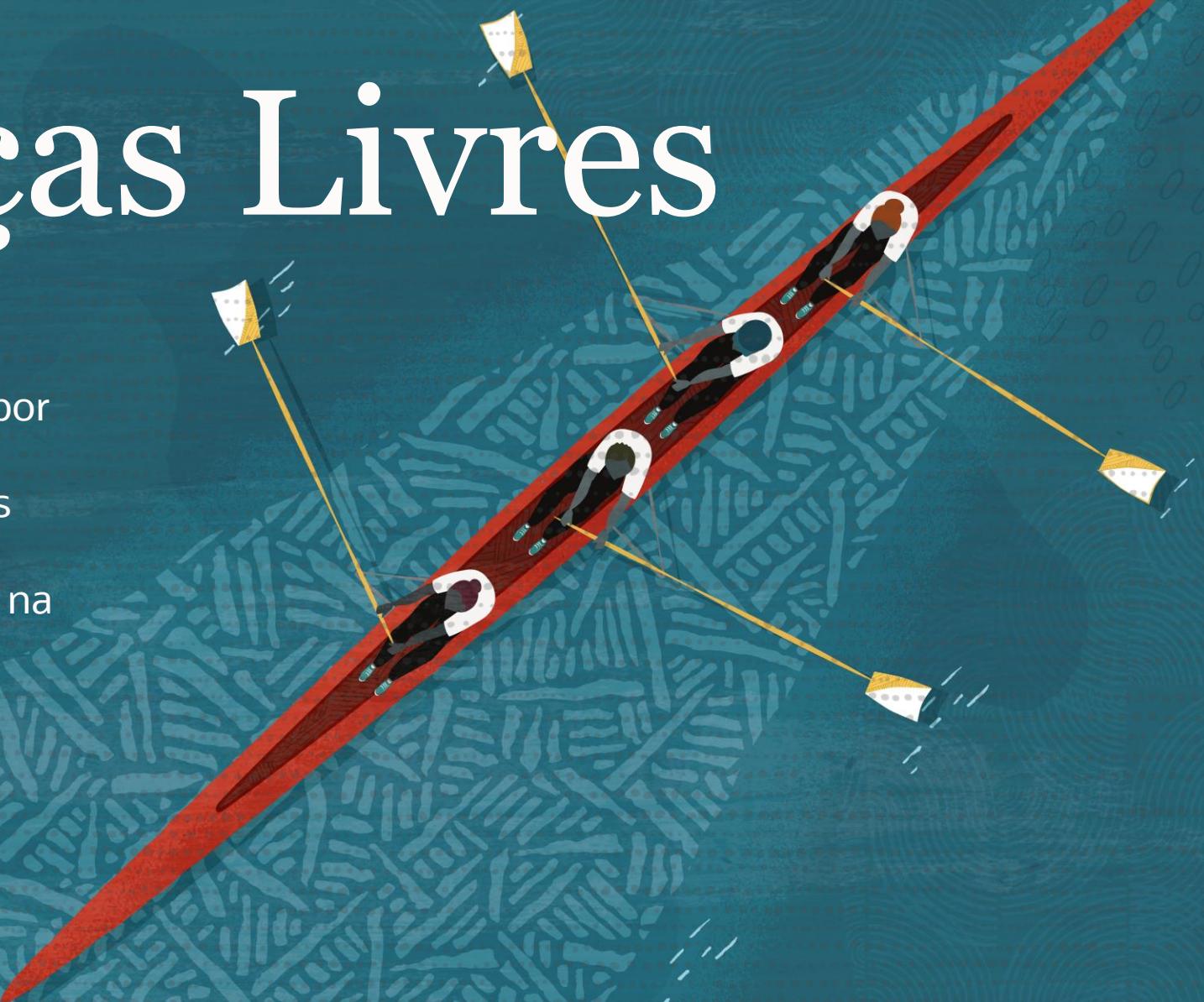
Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.



Fonte: Portal Brasileiro de Dados Abertos

8 Licenças Livres

Os dados não estão sujeitos a restrições por regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.



Fonte: Portal Brasileiro de Dados Abertos



FaLaBR

Nós podemos ajudar!

**Plataforma Integrada de Ouvidoria
e Acesso à Informação**

falabr.cgu.gov.br



github.com/dspanizzo



linkedin.com/in/dspanizzo

Obrigado!



Vamos juntos nesta trilha!



linkedin.com/groups/8984009



youtube.com/c/Inovaçãocomdadosemnuvem



anchor.fm/inova-dados-nuvem



github.com/taborda-cbip/inovacao-com-dados-em-nuvem



Inovação com dados em nuvem

**ANALISANDO GRANDES
VOLUMES DE DADOS
ABERTOS**



ORACLE

O