

# FERRAMENTAS, BASES DE DADOS E DICAS PARA CAPTURA DE DADOS - Observatório Universal de Projetos

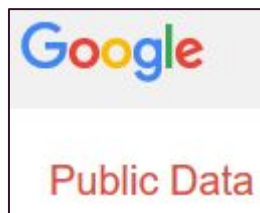
*Álvaro Pinheiro / Bruno Pires*  
*CIn - UFPE*  
*2018*



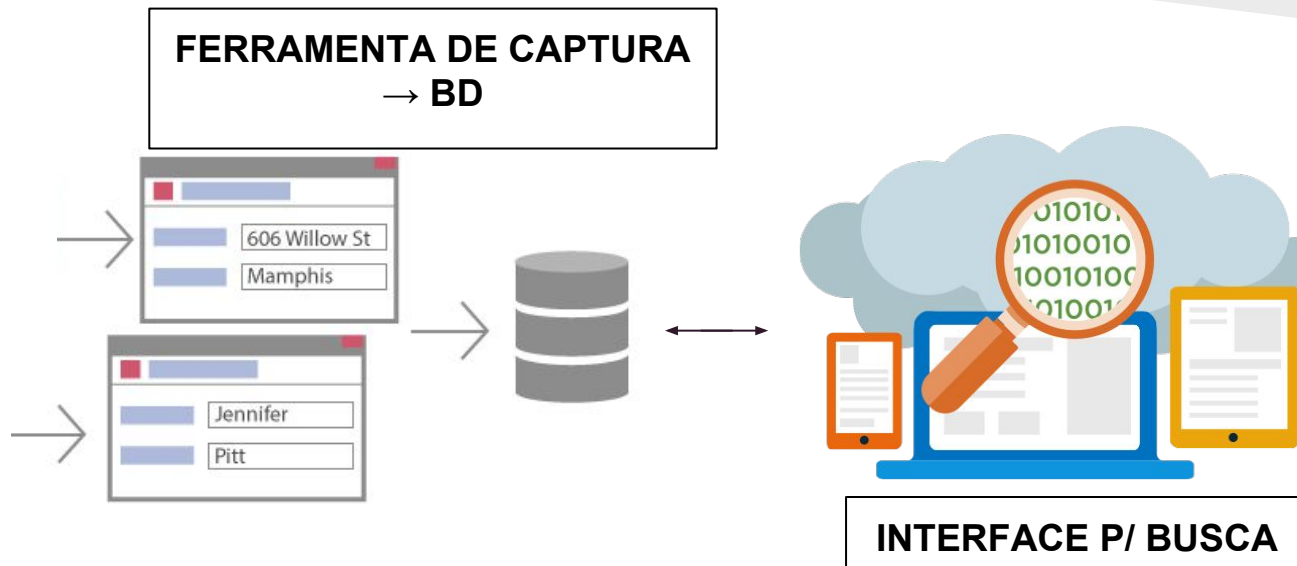
# Agenda

1. Ideia (Brainstorm)
2. Ferramentas (dados):
  - 2.1. Captura
  - 2.2. Conversão
  - 2.3. Extração
3. Bases de dados
  - 3.1. Dados públicos
4. Dicas

# 1. Ideia (brainstorm)



DADOS



Talvez um exemplo...  
<http://govdata.gov.br/>

# 1. Ideia (brainstorm)

## 1. Conjunto Mínimo de Dados - OP?

- a. Definir o que vai ser coletado do projeto (nome, prazo, valor, responsável, etc.)
- b. O que vai ser visualizado? Perfis de visualização?
- c. Tipos de Consulta - filtros?

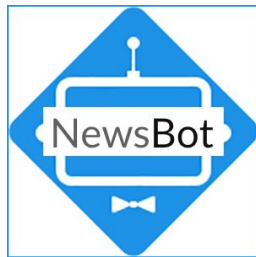
....

# 2. Ferramentas (dados)



**ZAMZAR**

Free online file conversion



**TABULA**

liberate data tables trapped inside PDFs

**twXplorer**

A smarter way to search Twitter

**cometdocs**  
Online Document Management System

**ABBYY**  
**FlexiCapture™**



**import.io**

**Soft Expert**®

Software for Business Excellence

**sabesim!**

# Descrição...

Ferramentas	Tipo	Free?	Descrição
ABBYY FlexiCapture	Captura	N	Software de entrada de dados automatizada.
SabeSim!	Captura	N	Sistema de Criação de Formulários de Coleta de Dados.
SoftExpert	Captura	N	Software que simplifica o tratamento de aspectos como a captura, processamento, validação e gerenciamento de conteúdo.
Cometdocs	Conversão	S	Conversor de PDF para Word, Excel, PowerPoint - diretamente no seu dispositivo móvel ou desktop.
Zamzar	Conversão	S	Conversor online para todos os formatos de arquivos (audio / video / documentos e etc.)
Tabula	Conversão	S	É uma ferramenta para liberar tabelas de dados dentro de arquivos PDF.
Import.io	Conversão	N	Ferramenta para Extrair -> Transformar -> Criar Insights (url -> múltiplas páginas, CSV, etc)
WebScraper	Extração	S	Web Scraper é uma empresa especializada em extração de dados de páginas da web. (Plug.Chr)
twXplorer	Extração	S	Ferramenta online para busca de informações no twitter.

# 2.1. Captura

Ferramenta de captura de dados de formulários

[https://www.abbyy.com/pt-br/processamento do formulario/](https://www.abbyy.com/pt-br/processamento-do-formulario/)

Sistema de criação de Formulários de Coleta de Dados

<http://www.sabesim.com.br/sistema-de-criacao-de-formularios/>

O SE capture é um software que simplifica o tratamento de aspectos como a captura, processamento, validação e gerenciamento de conteúdo

<https://www.softexpert.com.br/produto/captura-documentos/>

## 2.2. Conversão

Transformar seu PDF em XLS (Excel), ODS, TXT ou muitos outros

<https://www.cometdocs.com/>

Transformar PDF em tabelas diferentes, tornando o trabalho de limpeza e organização dos dados ainda maior

<https://www.zamzar.com/>

Transformar PDF e outros formatos em Excel

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

Transformar PDF em CSV inclusive em gráficos

<https://github.com/tabulapdf/tabula>



# 2.3. Extração

Extração de dados de páginas HTML

<https://www.import.io/>

<http://webscraper.io/>

<https://twxplorer.knightlab.com/>

<https://getnewsbot.com/>

# 3. Base de dados

Portal Brasileiro de Dados Abertos

<http://dados.gov.br/>

Gapminder - Essa é uma compilação de dados de diversas fontes

<https://www.gapminder.org/>

Google Public Data - plataforma que tem muitos dados sobre demografia, educação, economia e várias outras categorias

<https://www.google.com.br/publicdata/directory>

# 3. Base de dados

Freebase - São mais de 45 milhões de registros no portal

<https://developers.google.com/freebase/>

- Base de conhecimento colaborativa
- Coleta de dados estruturados online colhidas de várias fontes
- Comprada pela Google em 2010.

UCI Machine Learning Repository - coleção de bancos de dados, teorias de domínio e geradores de dados para análises empíricas

<http://mlr.cs.umass.edu/ml/datasets.html> - Exemplo de Repositório de Dados

# 3.1. Dados públicos

Portal de Transparência nos Recursos Públicos Federais

<http://transparencia.gov.br/>

(pode ser utilizado como consulta de gastos por Ação (projeto))

Portal Transparência do Conselho Nacional de Justiça

[https://paineis.cnj.jus.br/QvAJAXZfc/opendoc.htm?document=qvw\\_l%2FPainelCNJ.qvw&host=QVS%40neodimio03&anonymous=true&sheet=shTRP](https://paineis.cnj.jus.br/QvAJAXZfc/opendoc.htm?document=qvw_l%2FPainelCNJ.qvw&host=QVS%40neodimio03&anonymous=true&sheet=shTRP)

Portal da Transparência do Ministério Público Federal

<http://www.transparencia.mpf.mp.br/>

# 3.1. Dados públicos

Portal da Transparência da Polícia Federal

<http://www.pf.gov.br/acessoainformacao/acesso-a-informacao>

Portal da Transparência do Ministério da Cultura

<http://sistemas.cultura.gov.br/comparar/salicnet/salicnet.php>

Portal de Periódicos Capes - Uma das maiores bases de dados científicos do mundo

<http://www.periodicos.capes.gov.br/>

# 4. Dicas

HairCAPTURE capturar de dados em vídeos

<http://www.linhadecodigo.com.br/artigo/777/haircapture-utilizando-captura-de-video-em-c-semsegredos.aspx>

Capturar dados de um vídeo em tempo real

<https://imasters.com.br/artigo/25167/desenvolvimento/capturando-dados-de-um-video-em-tempo-real/>

Desenvolvendo Robôs com C# para captura de dados

<http://netcoders.com.br/roboscomcsharp/>

# Obrigado!

*Álvaro Pinheiro - [afp@cin.ufpe.br](mailto:afp@cin.ufpe.br)*

*Bruno Pires - [bvpm@cin.ufpe.br](mailto:bvpm@cin.ufpe.br)*

