**Projeto Querido Diário**

**Ano:** 2019 - atualmente

**Financiamento:** Catarse

**Parcerias:** Digital Ocean, Fundação Lemann, Hivos, Iniciativa Latinoamericana por los Datos Abiertos, Jurema, Purpose.

O Querido Diário é um projeto de código aberto da Open Knowledge Brasil que tem o objetivo de reunir, classificar e contextualizar as informações contidas nos diários oficiais de municípios de todo o Brasil e disponibilizá-los em uma plataforma de acesso livre e fácil.

Dessa forma, a ferramenta busca facilitar o monitoramento e a análise dessas informações.

O projeto inclui três implementações técnicas principais:

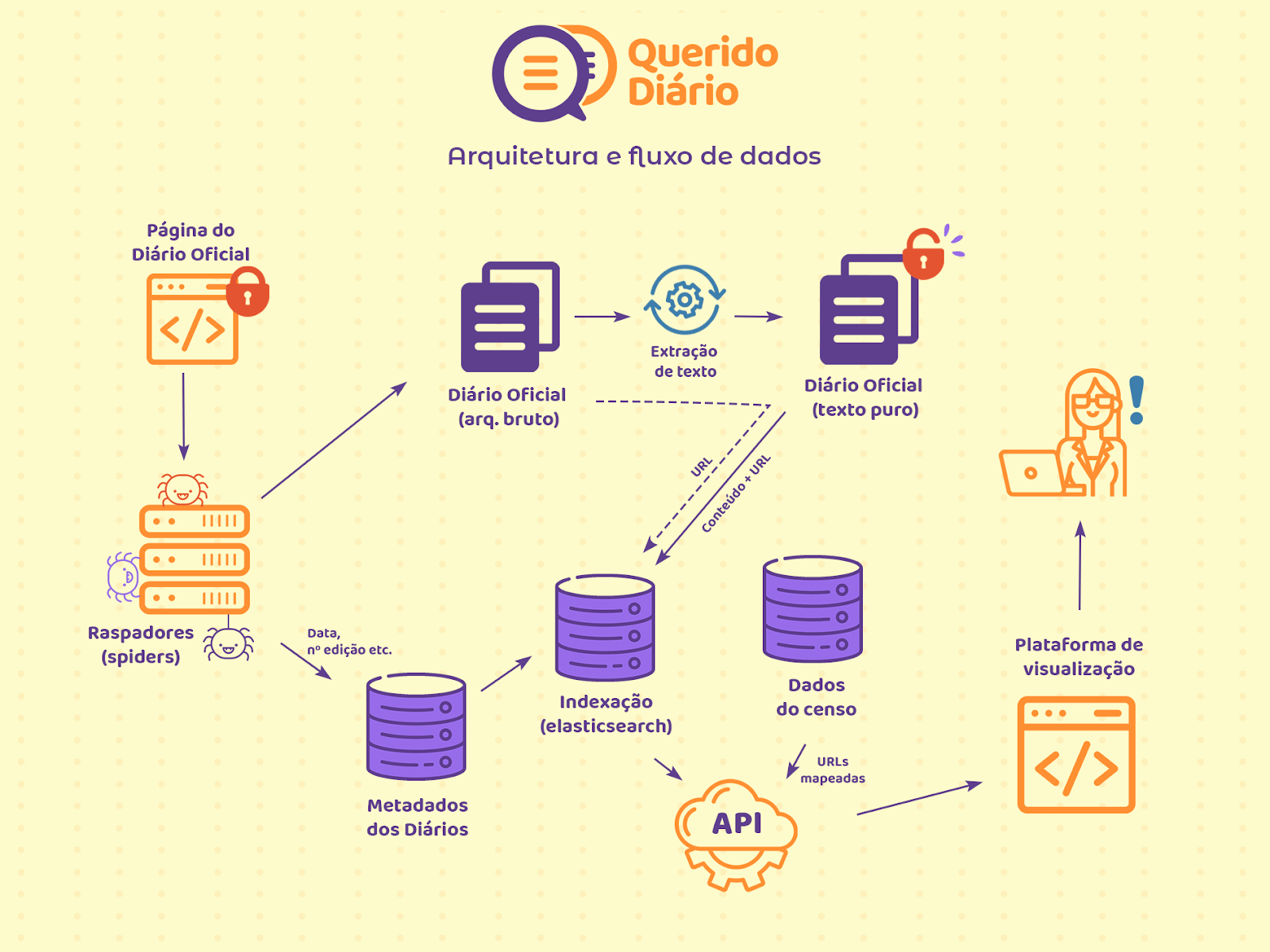
* Algoritmos de inteligência artificial (classificação e associação da informação publicada nos diários oficiais);
* Plataforma para visualização (plataforma em formato aberto);
* API aberta (integração das informações com outros sistemas e aplicações).

**O escopo do Querido Diário:**

Todo ato público, para que seja válido no Brasil, deve ser publicado em um diário oficial (arquivos que devem ser acessíveis pela população). Porém, essas informações publicadas nos diários oficiais dos municípios estão, por muitas vezes, presas em PDFs obscuros de acesso complicado.

O escopo atual do projeto é coletar os atos do poder executivo dos 5.570 municípios brasileiros.

**Arquitetura:**



**Contribua com código:**

Link do GitHub da Open Knowledge Brazil: https://github.com/okfn-brasil

**Processamento de dados:**

É nesta etapa que acontece a mágica de transformar dados fechados em abertos e conectar outras etapas do nosso fluxo de dados. No repositório você encontrará várias tecnologias como:

* *Apache Tika* para extrair texto de documentos de vários formatos;
* *Elasticsearch* para interagir com o indexador de texto dos documentos;
* *Postgresql* para interagir com os dados extraídos das spiders;
* E muito Python puro. 🤠

**Coleta de dados:**

Na estrutura de coleta de arquivos de diários oficiais, todos os scripts são desenvolvidos utilizando o framework Scrapy, ferramenta muito popular para raspagem de dados. Os arquivos (em Python) desenvolvidos pela comunidade para cada município são chamados de spiders, termo utilizado para um script de raspagem de dados no Scrapy.

**Censo:**

Para produzir um spider para um município, é necessário saber em qual site os arquivos do diário oficial estão sendo publicados, e isso não é uma tarefa muito simples.

Por isso foi criado o Censo Querido Diário, que é uma iniciativa de mapeamento das fontes de publicação onde qualquer pessoa pode preencher um formulário e nos enviar a fonte de publicação de diários oficiais de um município. Eventualmente, as fontes de publicação enviadas são validadas e analisadas por um time voluntário.

**API pública:**

A API pública permite a divulgação da base de dados de uma forma acessível e automatizável por máquinas, além de promover um acesso direto aos dados para pessoas desenvolvedoras e pesquisadoras. Além disso, a plataforma consome os dados da API para a disponibilização dos dados para o público geral.

A API foi construída em Python com FastAPI e é integrada ao Swagger para a obtenção da documentação de forma automática.

**Front end:**

Esta plataforma de busca é uma interface visual de interação com a API pública e também uma ferramenta de difusão de conhecimento e incentivo ao uso de diários oficiais.

O seu desenvolvimento é principalmente feito em Angular com Typescript, com responsividade para dispositivos móveis em primeiro lugar.

**Fontes:**

<https://ok.org.br/projetos/querido-diario/>

https://queridodiario.ok.org.br/