**Socrata API**

* O que é:
  + API para acessar dados disponibilizados publicamente.
* Como funciona:
  + Precisa de um token.
  + Limita o número de requisições. Usando um token, podem ser feitas até 1000 requisições por hora.
  + Permite a publicação de dados via ferramenta web. Não precisa usar a API.
    - Recomenda usa-la principalmente se o conjunto de dados será pouco atualizado.
  + A API segue o paradigma RESTful, ou seja, todas as requisições são feitas via HTTP e as respostas obtidas são no formato JSON, XML ou CSV.
    - Isso foi obtido num tutorial para acessar dados da cidade de Nova York.
    - Na cidade de Chicago isto também é valido.
    - Tem que verificar se é coisa da API ou é funcionalidade extra destes lugares. (É COISA DA API, NÃO DESTES DATASETS ESPECIFICOS).
  + Obviamente que, se os dados forem acessados via linguagem de programação, eles são retornados para dentro do programa, sem necessidade de se ler o JSON e etc.
    - É possível que a API “faça isso por trás dos panos”.
    - A autenticação do token é feita via HTTP
  + Acessa os datasets através do seu endpoint.
    - Endpoint é uma URL única que representa o dataset.
    - Geralmente, o endpoint é o que fica depois do /resources/ no link
    - Por exemplo: <https://data.cityofchicago.org/resource/ydr8-5enu.json?> É um dataset cujo endpoint é ‘ydr8-5enu’.
* Como guarda os dados:
  + Segundo eles, no ecossistema de open data temos dois tipos de sistemas: catálogos e data stores.
  + O SODA é uma data store, ou seja, ele armazena os dados.
  + Usa dois bancos de dados
    - Um é para os dados “duráveis”
    - Outro é quase um backup.
  + Basicamente, quando existe uma atualização de dados, quem faz o upload acessa via API, enviando os dados para o servidor.
  + Os dados são upados para o banco de dados durável.
  + É chamado um programa para verificar se tudo foi upado direito.
  + Só então é feito o sincronismo com o backup. Neste caso, somente o que foi alterado será carregador no backup.
  + Em caso de falha, é possível fazer o sincronismo total entre os bancos de dados.
* SoQL:
  + Socrata Query Language
* CKAN:
  + “Concorrente” do Socrata.
  + É de graça, mas tem uma versão paga para publicação.
  + Token são necessários na publicação dos dados.
    - Não precisa de Token na hora de acessa-los.
  + Aparentemente, o governo de Alagoas disponibiliza dados por esta ferramenta.
    - <http://dados.al.gov.br/guia-do-desenvolvedor/#/ferramentas-para-api>
  + Definição do FAQ do CKAN: Quem usa CKAN?
    - “CKAN é usado para “empoderar” portais oficiais de dados de governos nacionais e regionais no Reino Unido, Brasil, Holanda, Áustria, Estados Unidos e outros lugares...”
  + CKAN se mostrou confuso demais. Não é bem explicado.
  + Só tem uma API oficial, para Python.
  + Tem RESTFUL, ou seja, é acessível via HTTP.
  + É difícil achar o link correto.
  + De acordo com Ben McInnis, do Socrata, a comparação entre CKAN e Socrata é:
    - “*CKAN is an open source project, developed by the Open Knowledge Foundation, that lets users provision open data catalogs and, in some cases, visualizations and APIs.    
        
      Socrata is a software-as-a-service platform that provides a cloud-based solution for open data publishing and visualization.  All Socrata datasets are API-enabled and Socrata Open Data API (SODA) developers worldwide can use SODA to create apps, analyses, and complex visualizations atop any Socrata dataset.  The SODA server has been open sourced and could be self-provisioned as well.  The New York City open data site is a great example of a Socrata site:* [*NYC Open Data*](https://data.cityofnewyork.us/) *The most significant differences between Socrata and CKAN are the delivery model (self-provisioned vs. cloud) and the extent of API-enablement.  Socrata and CKAN datasets are interoperable via federation and both organizations share the goal of advancing the open data movement globally.“*
  + Outro funcionário do Socrata diz ver ambas ferramentas como complementares.
* Como descobrir o dominínios e datasets disponíveis via socrata:

Utilizando a API de pesquisa (*Catalog Search API*), disponibilizada pelo Socrata Labs, é possível achar (supostamente) os domínios disponíveis via Socrata.

A API é RESTful, retornando um arquivo JSON via acesso HTTP.

O primeiro passo é sempre achar o domínio. É possível achar uma lista dos domínios atráves do link <http://api.us.socrata.com/api/catalog/v1/domains>. Repare que este link retorna dois campos: o campo “domain”, com o link do domínio e um campo “count”. Testando aqui, acredito que este count equivalha ao número de datasets disponibilizados, via Socrata, pelo domínio. Vale ressaltar que isso não significa que só tem aquele número de datasets. É possível que tenham outros datasets sendo disponibilizados por outros meios.

Essa lista tem alguns links que não tem cara de domínio. Porém, tais links tem count = 0. Seguindo a minha teoria de que o count representa o número de domínios disponíveis via Socrata, nós não precisamos nos preocupar com estes links, se fizermos uma exclusão pelo valor do count.

Uma vez com o domínio, é possível achar a lista de datasets através de [http://api.us.socrata.com/api/catalog/v1?domains={dominio}](http://api.us.socrata.com/api/catalog/v1?domains=%7bdominio%7d). Por exemplo, o link <http://api.us.socrata.com/api/catalog/v1?domains=data.cityofchicago.org> nos retorna um JSON com todos os datasets da cidade de Chicago.

OBS: o endpoint de cada dataset está no campo “nbe\_fxf” (só Deus sabe o que isso significa).

* Links Úteis:
  + https://dev.socrata.com/consumers/getting-started.html
  + <http://www.thegreatcodeadventure.com/nyc-open-data-a-socrata-api/>
  + <https://stackoverflow.com/search?tab=newest&q=soda%20OR%20socrata>
  + <https://dev.socrata.com/docs/queries/>
  + <http://open-source.socrata.com/architecture/>
  + <http://labs.socrata.com/docs/search.html>