



Chapter 9

Scraping Sites That Use JavaScript and AJAX

As of August 2017, the website used for this tutorial had been archived by the Sudbury and District Health Unit, and was soon to be replaced. This tutorial has been updated to use the embedded Firebox developer tools. With the coming phaseout of the Sudbury food site, this tutorial will be replaced by a revised tutorial scraping a different live site. Once complete, the new tutorial will be posted to thedatajournalist.ca/newtutorials and will be posted on the Oxford University Press ARC site during the next refresh in summer 2018.

Skills you will learn: How to make GET requests to obtain data that updates a web page using Ajax; building a more complex multi-layer Ajax scrape; creating functions to reduce duplication in your code; using the JSON module to parse JSON data; Python dictionaries; Python error handling; how to use Firebug as part of the development of a more difficult scraping project.

Getting started

Like many communities throughout Canada and the U.S., the Sudbury and District Health Unit provides basic details about health inspections to the public via a website. You can go to the site at <https://www.sdhu.com/inspection-results#/>

If you click on the Inspection Results icon you'll see that the page soon populates with basic results of inspections.

Purchasewww.google.com