

## IT 497 Week 2 Lab

This lab will teach you how to quickly scrape internet data using **Google Sheets**. Google Sheets is a free online spreadsheet app.

For this lab

1. Open Google Sheets. If you do already have the app, go to <https://www.google.com/sheets/about/> to learn how to obtain it.
2. Read the attached exerts from Scraping for Journalists (2nd edition) by Paul Bradshaw. The full version of the book is available from Leanpub at: <https://leanpub.com/scrapingforjournalists>

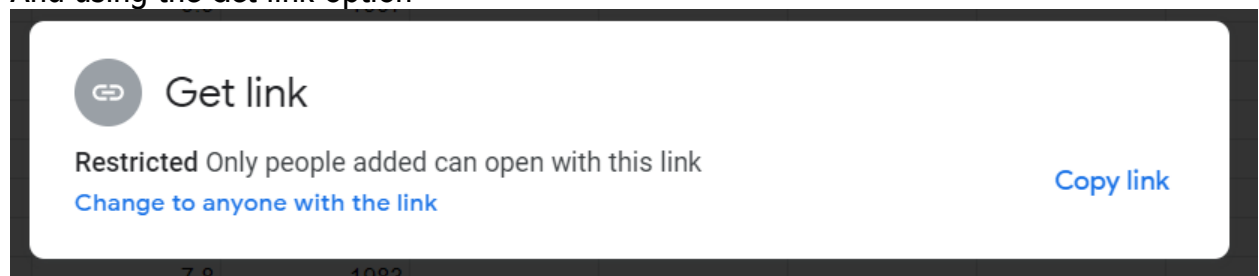
3. Replicate Scraper #1 -- David Bowie's discography.

```
=ImportHTML("https://en.wikipedia.org/wiki/David_Bowie_discography", "table", 2)
```

4. Share the link to this new sheet with me by selecting the share button in the upper right hand corner



And using the Get link option



Copy the link to Regginet

5. Read through the Tests on Page 9 and **complete the last one**. Pick any other musician you want. Create a new Google Sheet. Use the same method to obtain the information about another musician's song/album releases. Share the link to this new sheet with me just as you did in Step 4.
6. Replicate Scraper #2 – the W3 School's XML CD catalog.

```
=IMPORTXML("http://www.w3schools.com/xml/cd_catalog.xml", "CATALOG/CD")
```

Share the link to this new sheet with me via ReggieNet

7. Using the same W3 School's XML CD catalog, create a new google sheet. Change the formula used in 5 to grab ONLY the 24th CD in the list.

Share the link to this new sheet with me via ReggieNet.