What Justin Needs: Grant Database Pilot

Overview:

Existing datasets are in two folders: grants and funders.

The grants folder contains searches for previous grant opportunities meeting query parameters.

The funders folder contains a basic funding profile for the Robert Wood Johnson Foundation, one file pulled data on RWJF from the funders api and the other pulled data on RWJF from the transactions api.

What Justin needs to move forward:

1. grants folder:

Justin needs each of the 6 GET responses transformed into a data frame/pivot table. He also needs the total number of results from the meta info (before the name-value pairs start). This info will help him confirm whether each set of query parameters was correct.

2. funders folder:

Justin needs the list of grants from RWJF\_Trans to be in a data frame/pivot table.

3. Justin would LOVE to find an automated way to pull data on annual giving from Part XV, Section 3 of a Form 990. The Premier API does not seem include this data in the response (full query response saved as RWJF\_Financials in funder folder)—but it does return a link to a pdf of the whole Form 990 on Guidestar. Example link: <https://pdf.guidestar.org/PDF_Images/2020/226/029/2020-226029397-202331099349100118-PA.pdf>

NOTE: The annual giving section starts on page 12 of the example PDF.

THOUGHTS/LEADS:

1. Do the PDF versions of Form 990s have structured metadata for screen readers? If so, we should be able to use that as a way to, at least, navigate our data scraper to the correct section. Not a full answer, but perhaps a start. Holly may have ideas about methods we could use to efficiently scrape and structure this data.

2. The IRS has a REST API, but it’s unclear if/how we can use it to pull this data (foundation 990s have to be publicly accessible by law).