Access and Download Publicly Available U.S. Banking Data

Ted Liu*

December 29, 2021

Abstract

This manual describes the main types of U.S. banking data that are publicly available, and the processes in which you can access and download such data for statistical analysis.

Contents

---

* tedliu@ucsc.edu. Please contact me if you have questions or spot any errors in the manual.

1 Introduction

1.1 Types of Banking Data

Names: The two main types of publicly available banking data are Y9C and call reports. The official name for Y9C is "FR Y-9C Consolidated Financial Statements for Bank Holding Companies", and for call report is "Report of Condition and Income." U.S. banks are required by law to file reports of their financial information with regulatory agencies including the Federal Reserve Board of Governors (FRB) and the Federal Deposit Insurance Corporation (FDIC).

Content: Both types of reports summarize the aggregate information of a bank's operations within a given period. In other words, such reports do not provide loan level information, nor identify the recipients or parties involved in the bank's financial transactions. Nevertheless, combining summaries such as income statement and balance sheet, these reports provide detailed items of a bank's operations (e.g., interest-bearing versus non-interest-bearing deposits).

Frequency and Coverage: Both forms are filed quarterly by banks subject to reporting duties. The availability of Y9C data is 1986Q3 to present, and call report is 2001Q1 to present.

1.2 Differences between Call Reports and Y9C

Data Complexity: There is great institutional complexity in the U.S. banking system, which may partly explain the relatively fragmented framework of banking data. While this manual discusses Y9C and call reports, there is a diversity of reports that cover the various segments of the banking system. See works by Goldberg (2014) if you are interested in the full picture of U.S. banking data.

Banks as Institutions: In general, a bank holding company (BHC) is a parent organization of banks and non-bank entities. Under a BHC, there may be one or more commercial bank(s). Note that the data discussed in the manual focus on commercial banks, not credit unions.

Y9C versus Call Reports: In short, Y9C reports the data of a BHC, and call reports provide data of a commercial bank, which is usually owned by a parent BHC.

Example: In relation to banking data, the term/name "Citi" is in fact very imprecise. In the Y9C dataset, Citigroup Inc. is the organization of interest as it is a BHC. Citigroup Inc. is also the parent of Citibank, National Association. Therefore, in the call reports, you can find data for Citibank, but not Citigroup.

2 Look Up Specific Banks

2.1 How to Identify Banks

RSSD: The single most important identifier for banking data is an institution's unique RSSD ID (Research, Statistics, Supervision, Discount Identifier) assigned by the Federal Reserve. RSSD helps you locate not only an institution's financial data, but also its relations to other entities in the corporate hierarchy. Knowing the RSSD is also essential in obtaining qualitative information such as the opening and closing of entities, name changes, mergers and acquisitions, and even address changes.

Repository: The National Information Center (NIC) of the Federal Financial Insti-

tutions Examination Council (FFIEC) is an authoritative repository of bank data and institutional information. It is most useful as a first stop in the data collection process to obtain basic information about the bank you are looking for. You can access the site through https://www.ffiec.gov/npw

Search by Name: Once you are at the NIC page, you can start searching using the `Search Institutions` tab. While this tab offers many data filtering options, the most basic and useful function is to search by `Institution Name or RSSD ID`. Ideally before the search, you already know whether you are looking for a BHC or commercial bank entity.

However, you can start with a general search with a relatively vague search term. Take the example of Citi again, you can start by typing in "Citibank." This yields multiple results. You can verify that entity is indeed what you are looking for by the precise name of the entity, in this case "Citibank, N.A."

Search by RSSD: Suppose you already know precisely the RSSD of an institution, then you can simply type in the number to access the information. For example, by typing in the number 476810 in the search box, you can directly access the unique record of Citibank, as the number is its RSSD.

Additional Search Tools: There are additional data filtering tools:

1. Geographic locations: country, state, city

2. Additional options after clicking on `More Options`

   - The option to search by either `Active` or `Inactive` is particularly useful. Essentially, by selecting `Inactive`, you can also find banks that have closed down or been acquired

3. Organization hierarchy.

   - This is not technically a data filtering tool. Rather, for each institution, NIC provides a summary of what parent entities and subsidiaries the institution has. For example, within the record of Citibank (RSSD: 476810), there is an additional tab called `Organization Hierarchy`. Then you can select the direction of hierarchy, called `Tiering Direction`. With the `Up` option, you can see the parent organization Citicorp. With the `Down` option, you can see a list of child entities under Citibank.

   - This tool is particularly helpful to clear confusions regarding relationships between institutions.

2.2   Download Individual Reports

First Step: Before you proceed to download data, make sure you have the correct RSSD for the institution.

Y9C: The example of Citigroup Inc. (RSSD: 1951350)

1. Use https://www.ffiec.gov/npw

2. Type in 1951350 in search box `Institution Name or RSSD ID`

3. Click on result `CITIGROUP INC.`

4. Under the tab `Financial Data`, you will see the section `Consolidated Financial Statements for BHCs (FR Y-9C)`

5. You can download quarterly data either as CSV or PDF

6. Note the NIC website offers limited year coverage of Y9C data. See section 3 to learn how to access older data

Call report: The example of Citibank, NA (RSSD: 476810)

1. Use FFIEC's Central Data Repository Public Data Distribution website: https://cdr.ffiec.gov/public/

2. Select `Call Report` in the dropdown menu

3. In the dropdown menu `Report Date`, select reporting quarter of interest (or a range of quarters)

4. In the search box next to `Unique Identifier`, type in 476810

5. Click on `Generate`

6. A new window will open with the PDF of the call report. You can also download the data as XBRL or SDF file

3 Download (Bulk) and Analyze Data

3.1 When to Bulk Download

Advantages:

1. Time-saving if your project requires a large sample of banks

2. Not difficult to automate with Python

3. Downloaded files readily usable as CSV or format of your choice (instead of PDFs)

Disadvantages:

1. Bulk download is not necessarily efficient if you are only looking for a handful or banks or quarters

2. It is not costless timewise to slice the bulk dataset. There are hundreds of variables within the Y9C and call report data. You should first have a clear idea of what variables you are looking for before working with a bulk dataset

3.2 Metadata

Data dictionary: It can be overwhelming to understand the definitions of all the variables in the banking data. The Federal Reserve Board provides an online data dictionary:
https://www.federalreserve.gov/apps/mdrm/data-dictionary

- For example, suppose you want to understand the definition of `Total assets` in the call report.

- In the call report, the item code for `Total assets` is RCFD2170

- In the data dictionary,

  1. Select `Item Number` under `Search By`

  2. Under `for Series`, select RCFD

  3. In the blank box `Search for`, type in 2170

  4. Click on `Search`

List of BHCs:	Complete list of BHCs available here: https://www.chicagofed.org/banking/financial-institution-reports/bhc-name-list Only the PDF file is available, but you can use online resources or Adobe Acrobat to convert to to an Excel file. The RSSD information in the file is particularly useful.
Reporting Form Change:	The reporting guidelines for Y9C and call reports have changed over the years. This sometimes poses challenges for reconciling bank variables across time. The Federal Reserve Bank of New York provides resources documenting the key changes:

- Y9C: Document `Historic Reporting Changes` in
  https://www.newyorkfed.org/banking/reportingforms/FR_Y_9C.html

- Call report: Document `Historic Reporting Changes` in
  https://www.newyorkfed.org/banking/reportingforms/FFIEC_031.html

3.3 Y9C

Repository:	To download Y9C in bulk, use Chicago Federal Reserve's Distribution website: https://www.chicagofed.org/banking/financial-institution-reports/bhc-data **Note**: starting in 2021Q2, bulk Y9C data are available only at https://www.ffiec.gov/npw/FinancialReport/FinancialDataDownload
Download process:	

- Y9C data (for all BHCs) are available quarter by quarter as individual CSV files under `Historical quarterly datasets` of the Chicago Fed website. You can choose to download files one by one.

- An easier way is to automate the download process using Python. The URLs of the Y9C data files follow a sequential pattern. You can choose write a loop of URLs to download, or use `Beautiful Soup` to find and scrape all the CSV files on the website.

- An alternative approach is less elegant but works just as well: create a TXT file by concatenating the URL strings (easily doable in Excel), and use this file to download the data. For example of the TXT file, see
  https://github.com/tengtedliu/us_banking_data/blob/main/y9c_links.txt

Sample code of downloading using TXT file:

```
import os
os.getcwd()
os.chdir('/Users/ted/Downloads')

from requests import get

def download(url, filename):
with open(filename, "wb") as file:
response = get(url.replace("\n", ""), stream = True)
file.write(response.content)

f = open('y9c_links.txt','r')
for link in f:
url = link
split_url = url.split("/")
filename = split_url[-1]
filename = filename.replace("\n", "")
download(url,filename)
```

- Another option is to automate the download process with R. For an example of the R script, see
https://github.com/tengtedliu/us_banking_data/blob/main/y9c_download.R
Note: this script worked as of January 2021.

3.4 Call Report

Repository: Use FFIEC's Public Data Distribution Site:
https://cdr.ffiec.gov/public/PWS/DownloadBulkData.aspx

Download process:

- Brute force:
  1. In the repository, choose `Call Reports -- Single Period`
  2. Choose time periods under `Reporting Period End Date or Year (Four Periods in a Calendar Year)`. Select the file format and then click on `Download`

- Webscrapping: You can also use `Selenium` to scrape the data. Essentially you automate the 'click and download' process. Make sure you use Google Chrome, ChromeDriver, and that the these two are of the same versions. For an example of code, see
https://github.com/tengtedliu/us_banking_data/blob/main/call_report_download.ipynb

3.5 Processing and Analyzing Data

Data Cleaning: The first step after downloading the reports is to append and clean the data. This step really depends on your projects and what banks and variables you need. For examples of how to clean and filter data in Stata, see these two DO files, for Y9C and call reports respectively:

- Y9C: https://github.com/tengtedliu/us_banking_data/blob/main/y9c_dataclean.do

- Call report: https://github.com/tengtedliu/us_banking_data/blob/main/call_dataclean.do