

# Using PDF Extraction and Web Scraping Tools to Collect Government Health Insurance Plan Information

**Virginia Gwengi**

**Data Scientist**

**U.S. Census Bureau**

**2021 Federal Computer Assisted Survey Information Collection**

April 13, 2021

*Disclaimer:* Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the agency. All results have been reviewed to ensure that no confidential information is disclosed.

**Shape  
your future  
START HERE >**

United States<sup>®</sup>  
**Census  
2020**

# Overview

- **Summary of Benefits and Coverage Form (SBC)**
- **Motivation**
- **Methods**
- **Benefits**
- **Challenges and next steps**

# Summary of Benefits and Coverage Form

Summary of Benefits and Coverage: What this Plan Covers & What You Pay For Covered Services

Coverage Period: 01/01/2021 - 12/31/2021

The University of Alabama at Huntsville

Coverage For: Individual + Family Plan Type: PPO



The Summary of Benefits and Coverage (SBC) document will help you choose a health plan. The SBC shows you how you and the plan would share the cost for covered health care services. NOTE: Information about the cost of this plan (called the premium) will be provided separately.

This is only a summary. For more information about your coverage, or to get a copy of the complete terms of coverage, call 1-800-239-5772 or visit us at <https://uah.edu/hr/benefits/insurance/health>. For general definitions of common terms, such as [allowed amount](#), [balance billing](#), [coinsurance](#), [copayment](#), [deductible](#), [provider](#), or other underlined terms see the Glossary. You can view the Glossary at [www.bcbosal.org/sbcglossary/](http://www.bcbosal.org/sbcglossary/) or call 1-800-292-8868 to request a copy.

Important Questions	Answers	Why This Matters:
What is the overall deductible?	\$150 individual.	Generally, you must pay all of the costs from providers up to the deductible amount before this plan begins to pay. If you have other family members on the plan, each family member must meet their own individual deductible until the total amount of deductible expenses paid by all family members meets the overall family deductible.
Are there services covered before you meet your deductible?	Yes. Preventive services in-network are covered before you meet your deductible.	This plan covers some items and services even if you haven't yet met the deductible amount. But a copayment or coinsurance may apply. For example, this plan covers certain preventive services without cost-sharing and before you meet your deductible. See a list of covered preventive services at <a href="https://www.healthcare.gov/coverage/preventive-care-benefits/">https://www.healthcare.gov/coverage/preventive-care-benefits/</a> .
Are there other deductibles for specific services?	Yes. \$150 for prescription drug coverage. There are no other specific deductibles.	You must pay all of the costs for these services up to the specific deductible amount before this plan begins to pay for these services.
What is the out-of-pocket limit for this plan?	\$2,500 individual/\$7,150 family. There is a separate \$2,500 individual/\$7,150 family prescription drug out-of-pocket.	The out-of-pocket limit is the most you could pay in a year for covered services. If you have other family members in this plan, they have to meet their own out-of-pocket limit until the overall family out-of-pocket limit has been met.
What is not included in the out-of-pocket limit?	Premiums, balance-billed charges, health care this plan doesn't cover, pre-certification penalties, pharmacy copays and specialty drug coupon program payments.	Even though you pay these expenses, they don't count toward the out-of-pocket limit.
Will you pay less if you use a network provider?	Yes. See <a href="http://AlabamaBlue.com">AlabamaBlue.com</a> or call 1-800-810-BLUE for a list of network providers.	This plan uses a provider network. You will pay less if you use a provider in the plan's network. You will pay the most if you use an out-of-network provider, and you might receive a bill from a provider for the difference between the provider's charge and what your plan pays (balance billing). Be aware your network provider might use an out-of-network provider for some services (such as lab work). Check with your provider before you get services.
Do you need a referral to see a specialist?	No.	You can see the specialist you choose without a referral.

- Mandated as part of Affordable Care Act, the SBC form is the "nutrition facts" of health insurance coverage
- Employers with 50+ employees that offer health insurance must provide SBC forms to employees
- Many SBC forms can also be found online
- Contains information collected in the Medical Expenditure Panel Survey-Insurance Component (MEPS-IC)

# Motivation

- We aim to reduce respondent burden, improve data quality, increase timely responses and reduce survey production costs by taking advantage of existing administrative records (SBC forms)

## PDF Extraction Tool

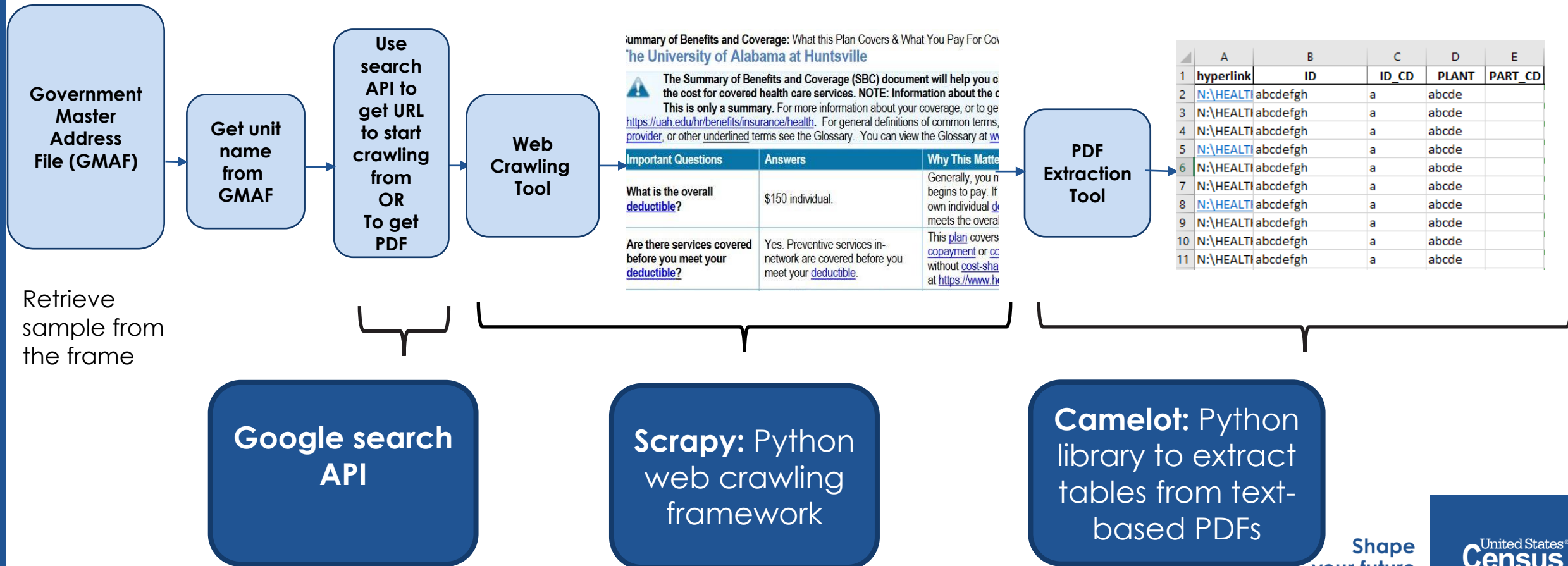
- Analysts manually extract information from SBC forms and hand enter into excel spreadsheets
- Data from SBC forms is not exhausted

## Web scraping Tool

To collect SBC forms:

- Analysts manually search for forms through web searches
- Respondents can upload forms into data collection tool

# Methods



Summary of Benefits and Coverage: What this Plan Covers & What You Pay For Coverage  
The University of Alabama at Huntsville

**The Summary of Benefits and Coverage (SBC) document will help you understand the cost for covered health care services. NOTE: Information about the cost of services is not included in this summary. For more information about your coverage, or to get a copy of the SBC, visit <https://uah.edu/hr/benefits/insurance/health>. For general definitions of common terms, visit [provider](#), or other underlined terms see the Glossary. You can view the Glossary at [www.uah.edu/hr/benefits/insurance/health/glossary](#).**

Important Questions	Answers	Why This Matters
What is the overall deductible?	\$150 individual.	Generally, you begin to pay. If your own individual deductible meets the overall deductible.
Are there services covered before you meet your deductible?	Yes. Preventive services in-network are covered before you meet your deductible.	This plan covers copayment or cost-sharing without cost-sharing at <a href="https://www.uah.edu/hr/benefits/insurance/health">https://www.uah.edu/hr/benefits/insurance/health</a> .

	A	B	C	D	E
1	hyperlink	ID	ID_CD	PLANT	PART_CD
2	<a href="#">N:\HEALTH</a>	abcdefgh	a	abcde	
3	N:\HEALTH	abcdefgh	a	abcde	
4	N:\HEALTH	abcdefgh	a	abcde	
5	<a href="#">N:\HEALTH</a>	abcdefgh	a	abcde	
6	N:\HEALTH	abcdefgh	a	abcde	
7	N:\HEALTH	abcdefgh	a	abcde	
8	<a href="#">N:\HEALTH</a>	abcdefgh	a	abcde	
9	N:\HEALTH	abcdefgh	a	abcde	
10	N:\HEALTH	abcdefgh	a	abcde	
11	N:\HEALTH	abcdefgh	a	abcde	

# Benefits

## PDF Extraction Tool

- Tool identifies and extracts SBC form in 15 seconds compared to 2-4 minutes in the manual process
- Extraction tool can provide additional plan-level information without additional cost
- Identifies PDFs to scrape from the web

## Web scraping tool and search API

- Quickly identifies SBC forms
- Allows for early data collection, as soon as sample is available (April), as opposed to waiting until the survey collection begins (June)

# Challenges and next steps

## Challenges

### ***PDF Extraction Tool***

- Not all units will have SBC forms (those with fewer than 50 employees, noncompliant employers)
- Can only extract SBC forms saved as readable PDFs (not scanned or in Word or plan brochures)
- Some forms are trickier than others to extract: inconsistent number of columns, varying patterns of text

### ***Web scraping Tool***

- Cannot scrape SBC forms on websites that require login information to retrieve forms
- Not all units may have web presence

## Next Steps

- Employ named entity recognition (NER) to increase the accuracy of the extraction tool



# Thank you

**Virginia Gwengi**

martha.v.gwengi@census.gov