



United States
Department of
Agriculture



National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-24-02

October 2024

Comparing Web Scraped Establishment Survey Frames of Industrial Hemp Growers in Seven States: *Costs, Contact Data, and Accuracy of Frame*

Michael Gerling, Chad Garber,
Katherine Vande Pol and Tyler Wilson

Survey Methodology & Technology
Research and Development

The findings and conclusions in this report are those of the authors and should not be construed represent any official USDA or U.S. Government determination or policy.

Acknowledgements

Linda J. Young, Ph.D.

Denise A. Abreu

Robbie Emmett

Darcy Miller

Executive Summary

The United States Department of Agriculture's (USDA's) National Agricultural Statistics Service (NASS) conducts over 300 agricultural surveys to provide official statistics on crops and livestock, explore production practices, and identify economic trends.

The industrial hemp industry has been growing and evolving over the past decade. Most states have passed laws legalizing the growing, processing and transporting of industrial hemp. In the 2018 Farm Bill, industrial hemp was recognized as a legal crop, and the USDA was directed to establish a national regulatory framework for hemp production in the United States. In 2020, NASS was contacted by the USDA's Agricultural Marketing Service (AMS) to conduct a survey on this new agricultural industry.

Three list frames of industrial hemp growers were developed: (1) a NASS web-scraped list frame of hemp growers (Internal frame) (2) a web-scraped list frame of hemp growers developed by a contractor using automated processes (Contractor frame) and (3) a list of registered hemp growers derived from administrative sources (USDA frame). These three list frames were compared in seven of the largest industrial hemp growing states: Colorado, Illinois, Missouri, Montana, Nevada, New York and Tennessee. The Internal and Contractor web scraped frames were designed to assess the under-coverage of the USDA frame.

All three frames were evaluated, focusing on the number of frame records, completeness of the information and the overlap between frames. In six of the seven states, the Contractor's frame had the lowest number of potential and probable industrial hemp operations. In Tennessee, all frames were closely comparable since all three frames utilized mostly the same sources in building their frames. The findings also suggest that utilizing the USDA frame supplemented with the Internal frame would provide the best frame overall.

Recommendations

- Continue to use USDA sources (if available) for the initial development of a new survey frame
 - Employ internal staff to enrich (addresses and phone numbers) and add additional operations to the USDA frame through web crawling and web scraping
- Evaluate whether employing a contractor is necessary for web crawling/scraping
 - Recommend employing a contractor for large web scraping over 100,000 operations to be scraped in less than six months
 - Recommend employing a contractor when there are multiple websites containing over 2,000 operations per website that cannot be easily scraped in-house
 - If a contract is required:
 - Recommend authoring the contract to ensure the ability to make minor modifications without incurring additional costs
 - Internal staff need to hold daily (and sometimes multiple) virtual meetings with the contractor to answer questions and provide technical support
 - Recommend web crawling/scraping in-house first to determine any challenges and how best to resolve those challenges, before entering into a contract
- If NASS decides to incorporate web crawling/scraping practices into its operational frame-building program, then the agency should author a standards document on web crawling/scraping. This document would standardize the key steps for successful frame building using web crawling/scraping and would include the following:
 - 1) how to determine key words
 - 2) best practices for web crawling and web scraping
 - 3) framework for building a list of websites to scrape
 - 4) framework for constructing a database of operations
 - 5) how to find missing addresses and phone numbers
 - 6) data clean-up
 - i. how to find, merge, and denote duplicate records
 - ii. how to find and remove scraped operations not of interest
 - 7) how to conduct a pilot web crawling/scraping in-house to determine any challenges and how best to resolve those before a web crawling/scraping project occurs in a production environment

Table of Contents

Abstract	1
1.0 Introduction	2
2.0 Web Crawling/Scraping in Building Survey Frames	2
3.0 Goals	3
4.0 Industrial Hemp Industry	3
5.0 Methods & Analyses	4
5.1 Internal Frame	4
5.2 Contractor Frame	5
5.3 USDA Frame	5
5.4 Study	5
6.0 Examination of the Seven States Combined	6
7.0 Descriptive Statistics by States & Key Observations	7
7.1 Colorado	7
7.2 Illinois	9
7.3 Missouri	11
7.4 Montana	13
7.5 Nevada	15
7.6 New York	17
7.7 Tennessee	19
7.8 Overall State Specific and Combined Findings	21
8.0 Two and Three-way Overlap/Non-Overlap	21
8.1 Internal Frame – Industrial Hemp Growers Licensing Two and Three-Way Overlap/Non-Overlap	23
9.0 Cost Comparison of Building an Industrial Hemp Growers Frame (50 States)	27
9.1 Internal Frame	27
9.2 Contractor Frame	28
9.3 USDA Frame	28
10.0 Capabilities and Limitations of Internal vs Contractor vs USDA	29
10.1 Internal Frame	29
10.2 Contractor Frame	29
10.3 USDA Frame	30
11.0 Discussion and Lessons Learned	30
12.0 Recommendations	31
13.0 Conclusion	32
14.0 References	32
15.0 Related Articles & Publications	33

Appendices

A	Appendix A - Hemp Industry Daily's 2020 Outlook	A-1
B	Appendix B - 2018 U.S. Hemp Crop Report	B-1

**Comparing Web Scraped Establishment Survey Frames of Industrial Hemp
Growers in Seven States:
*Costs, Contact Data, and Accuracy of Frame***

Michael Gerling, Chad Garber Tyler Wilson and Katherine Vande Pol^{1/}

Abstract

The National Agricultural Statistics Service's (NASS) conducts over 300 agricultural surveys each year to provide official statistics on all facets of U.S. agriculture including crops and livestock, production practices and economic trends.

The industrial hemp industry has been growing and evolving over the past decade. Most states have passed laws legalizing the growing, processing and transporting of industrial hemp. In 2020, NASS was contacted by Agricultural Marketing Service (AMS) to conduct a survey on this new agricultural industry.

Three list frames of industrial hemp growers were developed: (1) a NASS web-scraped list frame of hemp growers (Internal frame) (2) a web-scraped list frame of hemp growers developed by a contractor using automated processes (Contractor frame) and (3) a list of registered hemp growers derived from administrative sources (USDA frame). These three survey frames of industrial hemp growers were compared in seven states (Colorado, Illinois, Missouri, Montana, Nevada, New York and Tennessee).

Results showed that the USDA frame supplemented with the Internal frame would provide the best and most cost-efficient frame overall. In six of the seven states, the Contractor's frame had the least probable industrial hemp growers. In the seventh state (Tennessee) all frames were comparable since each frame utilized the same or similar web sources.

Key Words: web scraping, survey frames, agriculture, industrial hemp

^{1/} Michael W. Gerling - Mathematical Statistician, Samuel Garber – Agricultural Statistician, Tyler Wilson – Manager of the Survey Methodology and Technology Section, Katherine Vande Pol – formerly an Agricultural Statistician of the National Agricultural Statistics Service's Research & Development Division located at 1400 Independence Avenue, SW., Washington, DC 20250-2001.

1.0 Introduction

The United States Department of Agriculture's (USDA's) National Agricultural Statistics Service's (NASS) mission is to provide timely, accurate and useful statistics in service to United States agriculture. To meet its mission, each year NASS conducts more than 300 agricultural surveys and releases more than 400 reports on all facets of U.S. agriculture including crops and livestock, production practices and economic trends.

Over the past seven years, the industrial hemp industry has been growing and evolving. Most states have passed laws legalizing the growing, processing and transporting of industrial hemp. In the 2018 Farm Bill, industrial hemp was recognized as a legal crop. Shortly thereafter, NASS was contacted by the USDA Agricultural Marketing Service (AMS) to conduct a survey on this new agricultural industry.

One of the challenges facing NASS was developing an industrial hemp growers list frame under a 2-month time frame. Three list frames of industrial hemp growers were developed simultaneously: (1) a NASS web-scraped list frame of industrial hemp growers (Internal frame) (2) a web-scraped list frame of hemp growers developed by a contractor using automated processes (Contractor frame) and (3) a list of industrial hemp growers derived from various administrative sources, including state departments of agriculture and USDA agencies (USDA frame). The three list frames were compared for seven states of the larger hemp producing states: Colorado, Illinois, Missouri, Montana, Nevada, New York and Tennessee.

2.0 Web Crawling/Scraping in Building Survey Frames

In general terms, web crawling is the process (manual or automated) of identifying websites (URLs or links) of interest by utilizing web crawlers. Web scraping is the process of accessing these websites and copying the information of interest into a database either manually or through automated programs. Hence, the first step is web crawling followed by web scraping.

Web crawling/scraping to build a survey frame has been gaining favor in the survey industry, and literature on this topic is becoming more prevalent. Rhodes et al. (2016) utilized web scraping to build a list frame of Florida's vape stores.

Barcaroli, et al. (2016) used web crawling/scraping to update Italy's listing of agritourism farms, which are agricultural operations where the agricultural activities are integrated with tourism. They describe how a listing of farms and their general characteristics was created and maintained to promote the farming industry.

Bosch, et al. (2018) concluded that web scraping is of growing importance for national statistics institutes. They found that some of the prominent uses of web scraping are to reduce respondent burden and improve statistical production processes. They also discovered that web scraping is being used to explore background variables and to explore characteristics of (sub) populations.

Young, et al. (2018) developed a list of urban agricultural operations within the city limits of Baltimore, Maryland, using a combination of web scraping and satellite imagery. They found that the resolution of the available satellite imagery was not detailed enough to identify specific agricultural characteristics. However, 505 potential agricultural operations were obtained via scraping publicly available web sources.

To produce official statistics of farmers markets, Hyman, et al. (2021) developed a sampling approach that considered the overlap of the NASS list frame and a web-scraped list to improve the estimate of under coverage of the NASS list frame.

Fulton, et al. (2022) used Python-based programs to systematically scan Google Maps and generate a sampling frame of religious congregations in the U.S. They assessed Google Maps' coverage and representativeness by comparing this to two national listings. Google Maps contained approximately 98% of those congregations. They also found using Google Maps to generate sampling frames promises to substantially improve the process for obtaining representative samples for organizational studies by reducing costs, increasing efficiency, and providing greater coverage and representativeness.

3.0 Goals

This study's goals are as follows: (1) examine the benefits and drawbacks of web crawling/scraping to build an open-source frame within NASS versus employing a contractor, (2) descriptively compare the Internal, Contractor and USDA frames and (3) discuss cost and limitations of each frame.

4.0 Industrial Hemp Industry

On AskUSDA's website (2024) industrial hemp is defined as follows: *“Industrial hemp is from the plant species Cannabis sativa and has been used worldwide to produce a variety of industrial and consumer products. Hemp is a source of fiber and oilseed grown in more than 30 nations. In the United States, production is controlled under drug enforcement laws. To produce industrial hemp in the United States, the grower must obtain a permit from the Drug Enforcement Agency (DEA).”*

By definition, industrial hemp is high in fiber and low in active tetrahydrocannabinol (THC), the psychoactive ingredient in marijuana that makes some cannabis varieties a valued drug. Canada and the European Union maintain this distinction by strictly regulating the THC levels of industrial hemp, requiring it to be less than 0.3 percent, compared to THC levels of between 3 to 30 percent in marijuana.

With the passage of the 2018 Farm Bill, the Agricultural Marketing Service (AMS) has been designated as the lead United States Department of Agriculture (USDA) agency to administer the new USDA Hemp Production Program.”

Hemp should not be confused with marijuana, which is also of the cannabis species. Marijuana is a plant that produces resin containing cannabinoids which produces drug like effects in the body.

Two of the many cannabinoids are cannabidiol (CBD) and tetrahydrocannabinol (THC). They are both psychoactive cannabinoids; however, unlike THC, CBD is non-intoxicating and non-euphoric. Different cannabis varieties produce different levels of CBD and THC.

Hemp is a cannabis variety that produces relatively high CBD levels and relatively low THC (less than 0.3%) levels, as noted in USDA's definition.

CBD oil extracted from Hemp leaves and flowers is used for therapeutic purposes. Additionally, the fiber in hemp stalks is used in the manufacturing of clothing, rope, bedding materials, particle boards, ceiling panels and other industrial materials. Hemp seeds are used to produce edible items such as vitamins, flour and milk.

5.0 Methods & Analyses

All three frames were developed and transitioned to NASS's List Frame Section. The List Frame Section then conducted additional business processes before frames transformation (adding these records to NASS's overall frame of agricultural operations). The business processes and frames transformation are excluded from the analyses.

5.1 Internal Frame

Web crawling was conducted in all fifty states. Key words utilized in the web crawling include various combinations of (industrial, hemp, growers, and listing) with and without state name. Two hundred twenty-nine usable websites were found. These websites were then listed in a database showing the website, state(s) covered, number of operations, available information and number of jumps (mouse clicks) to obtain the information.

Web scraping was a combination of manual and semi-automated processes. During the web scraping process, seven states (Colorado, Illinois, Missouri, Montana, Nevada, New York and Tennessee) became the primary focus of the scraping. With the exception of Nevada, these states were anticipated to be the larger hemp producing states based on (1) Hemp Industry Daily's "*2020 Outlook: Licensed US hemp acreage falls 9% from 2019, but grower numbers increase 27%*" and (2) *2018 U.S. Hemp Crop Report* from Vote Hemp. See Appendices A and B for additional details. Nevada was the first state to be finalized into the database and was selected for its low number of industrial hemp growers, making it a good state to test the overall scraping and database building processes.

All scraped information was compiled into a database by state for all fifty states. Any processors, laboratories, transporters, and medical marijuana growers scraped in the process were included in the database and marked with an appropriate business type. Duplicate records were combined, using various features and functions in Excel and SAS. Next missing addresses and phone numbers were found using specific "people/business finder" websites to obtain this missing information. After the seventh state, Tennessee, was completed, the internal web scraping and data enrichment processes were halted since staff had to assist the contractor with their initial web crawling/scraping processes.

5.2 Contractor Frame

NASS staff held virtual meetings with the contractor and supplied the contractor with key words for web crawling and documentation on the industrial hemp industry. NASS also supplied the contractor with a listing of 229 websites that contain industrial hemp growers covering the fifty states. This saved the contractor several weeks of web crawling for sites to scrape.

The contractor's web scraping process was primarily automated using various software (python and java scripts). Manual scraping was utilized for those websites containing under 20 operations. Various programs were also used to determine and combine duplicate records.

The contractor's scraped entities not containing addresses were excluded since these were deemed unusable. Non-industrial hemp growers (grocery stores, hardware stores, gas stations, sporting goods stores, etc.) were removed from the Contractor's frame as well.

The Contractor frame did include a "Business Type" category, which provided additional insight to whether the business was a grower, processor, transporter, marijuana grower, etc. However, utilizing this category would have conservatively excluded over 64% of the frame. Thus, the "Business Type" category was deemed unreliable, and all of the scraped operations were assumed to be possible industrial hemp growers.

5.3 USDA Frame

The USDA frame was developed from listings provided by state departments of agriculture, and federal agencies, including AMS, Farm Service Agency (FSA) and Risk Management Agency (RMA).

5.4 Study

The seven states, (Colorado, Illinois, Missouri, Montana, Nevada, New York, and Tennessee) were selected for this study because these states (with the exception of Nevada) were anticipated to be the larger hemp producing states and were the states that the Internal frame's staff had completed before being re-allocated to assist the contractor. The descriptive statistics examined include, but are not limited to, number and percentage of operations with a listed operator's/owner's name, operation name, address and phone number.

Both the USDA and Contractor frames contained an undeterminable mix of possible license numbers and operation identification numbers. Thus, the hemp license counts for the USDA and Contractor frames were not calculated, and the hemp license counts across frames were not compared.

In Sections 6 and 7, multiple tables provide information about the frames. Address is broken into multiple categories. General Address is the main mailing address; Address Other is an additional mailing address; and Operation Address is where the operation is physically located. Any Address is counted as having at least one of the above addresses. The same counting technique is applied for phone. General City is the main mailing city and Operation City is the city of the operation's physical location. A similar naming convention is used for zip codes.

6.0 Examination of the Seven States Combined

The Internal frame had the highest number of operations followed by USDA and Contractor frames (see Table 1). The Contractor frame had the lowest percentage of records with a contact person’s name (46.0%) with USDA having the highest (69.3%). The USDA frame did not contain any websites while the Internal frame and the Contractor frame had 17-18%. Any Phone stood out because the Internal frame had almost twice the percentage of phone numbers at 86.7% compared to the USDA and Contractor frames at 48.8% and 34.3%, respectively.

Tennessee accounted for the following proportion in each frame.

- Internal: 3,785 of 9,291 operations (40.7%),
- Contractor: 3,708 of 6,434 operations (57.6%) and
- USDA: 4,050 of 8,582 operations (47.2%).

This equates to Tennessee accounting for 47% of the total scraped operations and is thus significantly influencing the overall comparisons. Hence, each state was examined separately.

Table 1: Seven States (CO, IL, MO, MT, NV, NY, TN) Combined by Frame¹

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations) ²	Internal Not Ind Hemp Percent - (Item/No. of Operations) ²
Number of Operations	9291	6434	8582	6694	2597	-	-	-	-	-
Person Name	5692	2962	5947	4538	1154	61.3%	46.0%	69.3%	67.8%	44.4%
Operation Name	5848	3978	5131	3420	2428	62.9%	61.8%	59.8%	51.1%	93.5%
Any Address	9108	6412	6939	6585	2523	98.0%	99.7%	80.9%	98.4%	97.2%
Any Phone	8058	2205	4185	5969	2089	86.7%	34.3%	48.8%	89.2%	80.4%
Any Email	4101	3390	4409	3395	705	44.1%	52.7%	51.4%	50.7%	27.1%
Website	1625	1143	2	325	1300	17.5%	17.8%	0.0%	4.9%	50.1%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

(2) Industrial Hemp status is based on whether having a grower’s license or not

7.0 Descriptive Statistics by States & Key Observations

Tables 2-8 shows each individual state's descriptive statistics and key observations.

7.1 Colorado

From the information provided in Table 2, the following results are highlighted.

- The Internal frame contained 2,558 operations; the USDA frame had 2,020 operations; and the Contractor frame had 1,666.
- The percentages of operations on a frame with any addresses were very similar, ranging from 98.8% to 99.9%.
- The percentages of operations on a frame with any phones ranged from 82.9% for the Internal frame to 78.2% for the Contractor frame to 97.5% for the USDA frame.
- The percentages of operations on a frame with any emails were 8.5% for the Internal frame, 22.6% for the Contractor frame, and 9.5% for the USDA frame.

Table 2: Colorado’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	2558	1666	2020	1189	1369	-	-	-	-	-
Name										
Person Name	600	81	562	315	285	23.5%	4.9%	27.8%	26.5%	20.8%
Operation Name	2358	1604	1551	989	1369	92.2%	96.3%	76.8%	83.2%	100.0%
Address										
General Address	2527	1649	2018	1187	1340	98.8%	99.0%	99.9%	99.8%	97.9%
Address Other	13	12	38	3	10	0.5%	0.7%	1.9%	0.3%	0.7%
Operation Address	0	88	2	0	0	0.0%	5.3%	0.1%	0.0%	0.0%
Any Address	2527	1649	2018	1187	1340	98.8%	99.0%	99.9%	99.8%	97.9%
Phone										
General Phone	2121	1303	1970	1164	957	82.9%	78.2%	97.5%	97.9%	69.9%
Phone Other	109	81	270	51	58	4.3%	4.9%	13.4%	4.3%	4.2%
Operation Phone	13	2	53	7	6	0.5%	0.1%	2.6%	0.6%	0.4%
Any Phone	2121	1303	1970	1164	957	82.9%	78.2%	97.5%	97.9%	69.9%
City										
General City	2558	1666	2020	1189	1369	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	88	2	0	0	0.0%	5.3%	0.1%	0.0%	0.0%
Zip										
General Zip	2558	1666	2020	1189	1369	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	88	2	0	0	0.0%	5.3%	0.1%	0.0%	0.0%
Email										
General Email	218	377	32	36	182	8.5%	22.6%	1.6%	3.0%	13.3%
Operation Email	0	6	178	0	0	0.0%	0.4%	8.8%	0.0%	0.0%
Any Email	218	377	191	36	182	8.5%	22.6%	9.5%	3.0%	13.3%
Other										
Website	943	700	0	72	871	36.9%	42.0%	0.0%	6.1%	63.6%
Operation County	2517	1656	376	1186	1331	98.4%	99.4%	18.6%	99.7%	97.2%
Hemp License Number	2463	N/A	N/A	1174	1289	96.3%	N/A	N/A	98.7%	94.2%

Notes:

- (1) All values exclude operations with an operation state other than the state of interest and all duplicate operations
- (2) Industrial Hemp status is based on whether having a grower's license or not
- (3) Color Coding of Main Contact Info

Green = Highest value in column set and row
 Orange = Intermediate value in column set and row
 Red = Lowest value in column set and row

7.2 Illinois

From the information provided in Table 3, the following results are highlighted.

- The Internal frame had 682 operations compared to Contractor frame at 127 and USDA frame with 964.
 - Internal frame had 530 operations with grower licenses.
 - One theory, for the low number of operations on the Contractor frame, is that the websites scraped by the contractor did not always include addresses and were thus excluded from the final deliverables per the contractual agreement.
 - Staff working on the Internal frame did have to scrape additional websites for addresses and phone numbers.
- Internal and Contractor frames had 97.5% and 100.0% of its operations with any address whereas USDA had only 30%.
 - This difference is due to the Internal frame's operations were data enriched from multiple websites and the contractor's deliverables had to have addresses.
- Internal and Contractor frames had 95.2% and 74.8% of its operations with any phone number whereas the USDA frame had only 26.5%.
- Both Internal and USDA frames had a high percentage of person names at 98.4% and 79.4% respectively compared to the Contractor's frame at 5.5%.

Table 3: Illinois’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	682	127	964	530	152	-	-	-	-	-
Name										
Person Name	671	7	765	529	142	98.4%	5.5%	79.4%	99.8%	93.4%
Operation Name	200	127	728	51	149	29.3%	100.0%	75.5%	9.6%	98.0%
Address										
General Address	665	127	289	524	141	97.5%	100.0%	30.0%	98.9%	92.8%
Address Other	2	0	2	1	1	0.3%	0.0%	0.2%	0.2%	0.7%
Operation Address	0	2	5	0	0	0.0%	1.6%	0.5%	0.0%	0.0%
Any Address	665	127	289	524	141	97.5%	100.0%	30.0%	98.9%	92.8%
Phone										
General Phone	649	95	255	513	136	95.2%	74.8%	26.5%	96.8%	89.5%
Phone Other	285	2	112	255	30	41.8%	1.6%	11.6%	48.1%	19.7%
Operation Phone	3	2	17	3	0	0.4%	1.6%	1.8%	0.6%	0.0%
Any Phone	649	95	255	513	136	95.2%	74.8%	26.5%	96.8%	89.5%
City										
General City	682	127	964	530	152	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	0	5	0	0	0.0%	0.0%	0.5%	0.0%	0.0%
Zip										
General Zip	682	127	964	530	152	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	2	5	0	0	0.0%	1.6%	0.5%	0.0%	0.0%
Email										
General Email	219	22	17	177	42	32.1%	17.3%	1.8%	33.4%	27.6%
Operation Email	1	0	187	1	0	0.1%	0.0%	19.4%	0.2%	0.0%
Any Email	219	22	188	177	42	32.1%	17.3%	19.5%	33.4%	27.6%
Other										
Website	95	86	1	20	75	13.9%	67.7%	0.1%	3.8%	49.3%
Operation County	608	112	603	472	136	89.1%	88.2%	62.6%	89.1%	89.5%
Hemp License Number	3	N/A	N/A	2	1	0.4%	N/A	N/A	0.4%	0.7%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

(2) Industrial Hemp status is based on whether having a grower's license or not

(3) Color Coding of Main Contact Info

Green = Highest value in column set and row

Orange = Intermediate value in column set and row

Red = Lowest value in column set and row

7.3 Missouri

From the information provided in Table 4, the following results are highlighted.

- Internal and Contractor frames had twice the number of operations at 514 and 530, respectively compared to USDA's 228.
- All three frames had 99-100% with any address.
- USDA frame had the highest percentage with any phones at 96.5%, versus Internal 82.1% and Contractor at 60.9%.
- Websites collected, Contractor had 9.4% and Internal 10.3%, while USDA had none.

Table 4: Missouri's Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	514	530	228	195	319	-	-	-	-	-
Name										
Person Name	389	337	217	90	299	75.7%	63.6%	95.2%	46.2%	93.7%
Operation Name	464	490	166	145	319	90.3%	92.5%	72.8%	74.4%	100.0%
Address										
General Address	510	527	228	193	317	99.2%	99.4%	100.0%	99.0%	99.4%
Address Other	1	2	10	0	1	0.2%	0.4%	4.4%	0.0%	0.3%
Operation Address	1	0	3	1	0	0.2%	0.0%	1.3%	0.5%	0.0%
Any Address	510	527	228	193	317	99.2%	99.4%	100.0%	99.0%	99.4%
Phone										
General Phone	422	323	220	132	290	82.1%	60.9%	96.5%	67.7%	90.9%
Phone Other	35	26	3	24	11	6.8%	4.9%	1.3%	12.3%	3.4%
Operation Phone	8	2	1	7	1	1.6%	0.4%	0.4%	3.6%	0.3%
Any Phone	422	323	220	132	290	82.1%	60.9%	96.5%	67.7%	90.9%
City										
General City	514	530	228	195	319	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	0	3	0	0	0.0%	0.0%	1.3%	0.0%	0.0%
Zip										
General Zip	514	530	228	195	319	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	0	1	0	0	0.0%	0.0%	0.4%	0.0%	0.0%
Email										
General Email	343	298	23	43	300	66.7%	56.2%	10.1%	22.1%	94.0%
Operation Email	1	2	190	0	1	0.2%	0.4%	83.3%	0.0%	0.3%
Any Email	343	298	208	43	300	66.7%	56.2%	91.2%	22.1%	94.0%
Other										
Website	53	50	0	34	19	10.3%	9.4%	0.0%	17.4%	6.0%
Operation County	388	407	2	168	220	75.5%	76.8%	0.9%	86.2%	69.0%
Hemp License Number	495	N/A	N/A	192	303	96.3%	N/A	N/A	98.5%	95.0%

Notes:

- (1) All values exclude operations with an operation state other than the state of interest and all duplicate operations
- (2) Industrial Hemp status is based on whether having a grower's license or not
- (3) Color Coding of Main Contact Info

Green = Highest value in column set and row
 Orange = Intermediate value in column set and row
 Red = Lowest value in column set and row

7.4 Montana

From the information provided in Table 5, the following results are highlighted.

- Internal frame had the highest number of operations at 654, followed by USDA at 354 and Contractor at 244. However, the Internal frame only had 271 with known grower licenses.
- Internal and Contractor frames had over 90% with any address and any phones while USDA fluctuated between 55.4% and 58.2%.
- USDA frame had 94.9%, with person name, followed by Internal frame at 84.7% and the Contractor frame at 0%.
- Contractor frame had 100% with operation name, followed by Internal frame with 68.0% and USDA with 66.9% .
- Contractor frame had the highest percentage of websites collected at 71.3% while Internal and USDA frames were at 19.7% and 0% respectively.

Table 5: Montana’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	654	244	354	271	383	-	-	-	-	-
Name										
Person Name	554	0	336	270	284	84.7%	0.0%	94.9%	99.6%	74.2%
Operation Name	445	244	237	188	257	68.0%	100.0%	66.9%	69.4%	67.1%
Address										
General Address	635	244	206	265	370	97.1%	100.0%	58.2%	97.8%	96.6%
Address Other	3	1	39	1	2	0.5%	0.4%	11.0%	0.4%	0.5%
Operation Address	0	0	4	0	0	0.0%	0.0%	1.1%	0.0%	0.0%
Any Address	635	244	206	265	370	97.1%	100.0%	58.2%	97.8%	96.6%
Phone										
General Phone	623	223	196	246	377	95.3%	91.4%	55.4%	90.8%	98.4%
Phone Other	21	5	79	10	11	3.2%	2.0%	22.3%	3.7%	2.9%
Operation Phone	0	0	12	0	0	0.0%	0.0%	3.4%	0.0%	0.0%
Any Phone	623	223	196	246	377	95.3%	91.4%	55.4%	90.8%	98.4%
City										
General City	654	244	354	271	383	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	0	4	0	0	0.0%	0.0%	1.1%	0.0%	0.0%
Zip										
General Zip	654	244	354	271	383	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	0	82	0	0	0.0%	0.0%	23.2%	0.0%	0.0%
Email										
General Email	44	57	20	10	34	6.7%	23.4%	5.6%	3.7%	8.9%
Operation Email	0	0	154	0	0	0.0%	0.0%	43.5%	0.0%	0.0%
Any Email	44	57	155	10	34	6.7%	23.4%	43.8%	3.7%	8.9%
Other										
Website	129	174	0	16	113	19.7%	71.3%	0.0%	5.9%	29.5%
Operation County	626	242	92	264	362	95.7%	99.2%	26.0%	97.4%	94.5%
Hemp License Number	0	N/A	N/A	0	0	0.0%	N/A	N/A	0.0%	0.0%

Notes:

- (1) All values exclude operations with an operation state other than the state of interest and all duplicate operations
- (2) Industrial Hemp status is based on whether having a grower's license or not
- (3) Color Coding of Main Contact Info

Green = Highest value in column set and row
 Orange = Intermediate value in column set and row
 Red = Lowest value in column set and row

7.5 Nevada

From the information provided in Table 6, the following results are highlighted.

- Internal frame had the most operations at 361. Internal frame's operations, with grower licenses (230), were still more than USDA's frame at 173 and the Contractor's frame at 104.
- USDA frame contained the lowest percentages with any address at 13.9% and with any phone at 10.4%.
- Internal frame had the highest percentage of person names at 59.0%.
- Contractor's primary contact information was at or above 90% for any addresses and for any phone.

Table 6: Nevada’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	361	104	173	230	131	-	-	-	-	-
Name										
Person Name	213	1	32	166	47	59.0%	1.0%	18.5%	72.2%	35.9%
Operation Name	346	104	151	217	129	95.8%	100.0%	87.3%	94.3%	98.5%
Address										
General Address	311	103	24	192	119	86.1%	99.0%	13.9%	83.5%	90.8%
Address Other	0	0	1	0	0	0.0%	0.0%	0.6%	0.0%	0.0%
Operation Address	0	0	0	0	0	0.0%	0.0%	0.0%	0.0%	0.0%
Any Address	311	103	24	192	119	86.1%	99.0%	13.9%	83.5%	90.8%
Phone										
General Phone	263	94	18	153	110	72.9%	90.4%	10.4%	66.5%	84.0%
Phone Other	23	8	0	20	3	6.4%	7.7%	0.0%	8.7%	2.3%
Operation Phone	0	0	0	0	0	0.0%	0.0%	0.0%	0.0%	0.0%
Any Phone	263	94	18	153	110	72.9%	90.4%	10.4%	66.5%	84.0%
City										
General City	361	104	173	230	131	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	0	0	0	0	0.0%	0.0%	0.0%	0.0%	0.0%
Zip										
General Zip	361	104	173	230	131	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	0	0	0	0	0.0%	0.0%	0.0%	0.0%	0.0%
Email										
General Email	163	22	82	131	32	45.2%	21.2%	47.4%	57.0%	24.4%
Operation Email	1	0	41	1	0	0.3%	0.0%	23.7%	0.4%	0.0%
Any Email	164	22	123	131	32	45.4%	21.2%	71.1%	57.0%	24.4%
Other										
Website	140	87	1	51	89	38.8%	83.7%	0.6%	22.2%	67.9%
Operation County	338	96	153	223	115	93.6%	92.3%	88.4%	97.0%	87.8%
Hemp License Number	4	N/A	N/A	2	2	1.1%	N/A	N/A	0.9%	1.5%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

(2) Industrial Hemp status is based on whether having a grower's license or not

(3) Color Coding of Main Contact Info

Green = Highest value in column set and row

Orange = Intermediate value in column set and row

Red = Lowest value in column set and row

7.6 New York

From the information provided in Table 7, the following results are highlighted.

- Contractor frame had only 55 operations total compared to Internal frame's 737 and USDA frame's 793. Internal frame had 649 licensed growers.
- USDA frame was the lowest in typical contact information, any address 15.6% and any phone number (11.1%) .
- Contractor frame had the greatest percentage of any phones at 96.4% followed by Internal and USDA at 60.1% and 11.1%, respectively.
- USDA frame had the highest percentage of email addresses at 91.7% followed closely by Internal frame at 90.5% and by the Contractor frame at 81.8%.

Table 7: New York’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	737	55	793	649	88	-	-	-	-	-
Name										
Person Name	702	11	774	644	58	95.3%	20.0%	97.6%	99.2%	65.9%
Operation Name	666	53	689	579	87	90.4%	96.4%	86.9%	89.2%	98.9%
Address										
General Address	676	55	124	594	82	91.7%	100.0%	15.6%	91.5%	93.2%
Address Other	0	0	5	0	0	0.0%	0.0%	0.6%	0.0%	0.0%
Operation Address	0	1	0	0	0	0.0%	1.8%	0.0%	0.0%	0.0%
Any Address	676	55	124	594	82	91.7%	100.0%	15.6%	91.5%	93.2%
Phone										
General Phone	443	53	88	377	66	60.1%	96.4%	11.1%	58.1%	75.0%
Phone Other	16	7	30	12	4	2.2%	12.7%	3.8%	1.8%	4.5%
Operation Phone	1	0	9	1	0	0.1%	0.0%	1.1%	0.2%	0.0%
Any Phone	443	53	88	377	66	60.1%	96.4%	11.1%	58.1%	75.0%
City										
General City	737	55	793	649	88	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	1	0	0	0	0.0%	1.8%	0.0%	0.0%	0.0%
Zip										
General Zip	737	55	793	649	88	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	1	0	0	0	0.0%	1.8%	0.0%	0.0%	0.0%
Email										
General Email	667	45	654	617	50	90.5%	81.8%	82.5%	95.1%	56.8%
Operation Email	0	3	102	0	0	0.0%	5.5%	12.9%	0.0%	0.0%
Any Email	667	45	727	617	50	90.5%	81.8%	91.7%	95.1%	56.8%
Other										
Website	181	42	0	122	59	24.6%	76.4%	0.0%	18.8%	67.0%
Operation County	719	53	702	632	87	97.6%	96.4%	88.5%	97.4%	98.9%
Hemp License Number	671	N/A	N/A	621	50	91.0%	N/A	N/A	95.7%	56.8%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

(2) Industrial Hemp status is based on whether having a grower’s license or not

(3) Color Coding of Main Contact Info

- Green = Highest value in column set and row
- Orange = Intermediate value in column set and row
- Red = Lowest value in column set and row

7.7 Tennessee

From the information provided in Table 8, the following results are highlighted.

- Overall, all three frames were fairly comparable.
 - USDA frame had the highest number of operations at 4,050 followed by Internal frame at 3,785 and Contractor frame at 3,708.
 - All three frames had 100% with addresses.
 - All three frames had a dismal number of websites with Internal having the highest percentage at 2.2%.
- The three frames differed substantially on any phone. The Internal frame had 93.4% followed by USDA at 35.5 % and Contractor at 3.1%.

Table 8: Tennessee’s Descriptive Statistics by Frame^{1,2,3}

Item	Internal	Contractor	USDA	Internal - Ind Hemp	Internal - Not Ind Hemp	Internal Percent - (Item/No. of Operations)	Contractor Percent - (Item/No. of Operations)	USDA Percent - (Item/No. of Operations)	Internal Ind Hemp Percent - (Item/No. of Operations)	Internal Not Ind Hemp Percent - (Item/No. of Operations)
Number of Operations	3785	3708	4050	3630	155	-	-	-	-	-
Name										
Person Name	2563	2525	3261	2524	39	67.7%	68.1%	80.5%	69.5%	25.2%
Operation Name	1369	1356	1609	1251	118	36.2%	36.6%	39.7%	34.5%	76.1%
Address										
General Address	3784	3707	4050	3630	154	100.0%	100.0%	100.0%	100.0%	99.4%
Address Other	18	16	278	17	1	0.5%	0.4%	6.9%	0.5%	0.6%
Operation Address	1	6	178	1	0	0.0%	0.2%	4.4%	0.0%	0.0%
Any Address	3784	3707	4050	3630	154	100.0%	100.0%	100.0%	100.0%	99.4%
Phone										
General Phone	3537	114	1438	3384	153	93.4%	3.1%	35.5%	93.2%	98.7%
Phone Other	2826	13	205	2809	17	74.7%	0.4%	5.1%	77.4%	11.0%
Operation Phone	2332	0	44	2323	9	61.6%	0.0%	1.1%	64.0%	5.8%
Any Phone	3537	114	1438	3384	153	93.4%	3.1%	35.5%	93.2%	98.7%
City										
General City	3785	3708	4050	3630	155	100.0%	100.0%	100.0%	100.0%	100.0%
Operation City	0	6	177	0	0	0.0%	0.2%	4.4%	0.0%	0.0%
Zip										
General Zip	3785	3708	4050	3630	155	100.0%	100.0%	100.0%	100.0%	100.0%
Operation Zip	0	6	175	0	0	0.0%	0.2%	4.3%	0.0%	0.0%
Email										
General Email	2446	2566	2217	2381	65	64.6%	69.2%	54.7%	65.6%	41.9%
Operation Email	22	68	744	19	3	0.6%	1.8%	18.4%	0.5%	1.9%
Any Email	2446	2569	2817	2381	65	64.6%	69.3%	69.6%	65.6%	41.9%
Other										
Website	84	4	0	10	74	2.2%	0.1%	0.0%	0.3%	47.7%
Operation County	2378	3677	2541	2373	5	62.8%	99.2%	62.7%	65.4%	3.2%
Hemp License Number	1566	N/A	N/A	1413	153	41.4%	N/A	N/A	38.9%	98.7%

Notes:

- (1) All values exclude operations with an operation state other than the state of interest and all duplicate operations
- (2) Industrial Hemp status is based on whether having a grower's license or not
- (3) Color Coding of Main Contact Info

Green = Highest value in column set and row

Orange = Intermediate value in column set and row

Red = Lowest value in column set and row

7.8 Overall State Specific and Combined Findings

Examining the combined states table, the Internal frame had the highest number of operations at 9,291 followed by USDA and the Contractor frame at 6,434 and 8,582, respectively. The Contractor frame had the highest percentage with any address. However, this was one of the requirements placed upon the contractor for having a completed deliverable industrial hemp grower. The USDA frame was ranked last in having any address and did not contain any websites. Internal frame had the highest percentage of any phone at 86.7%, followed by the USDA frame (48.8%) and the Contractor frame (34.3%).

Looking at individual states, the USDA and Internal frames had significantly more operations than the Contractor frame for all states except Missouri. In New York, the Contractor frame had only 55 operations total compared to Internal frame's 737 and USDA frame's 793. Internal frame had 649 licensed growers. Meaning, that both the Internal and USDA frames had over 13 times the number of operations than the Contractor frame. In Tennessee, all three frames were fairly comparable. The USDA frame had the highest number of operations at 4,050 followed by the Internal and Contractor frames at 3,785 and 3,708, respectively.

Although interesting, this still did not answer the question on which frame had the most industrial hemp growers. So, the next step was to examine the number of operations overlapping across the three frames.

8.0 Two and Three-way Overlap/Non-Overlap

The numbers of two- and three-way matches across the three frames as well as those operations unique to a frame were determined (see Table 9) . Tennessee is the outlier having the highest percentage of matches across all three frames at 89.5%, followed by Missouri at a distant 33.5%. This makes sense because all three frames used mostly the same sources for acquiring industrial hemp operations for Tennessee. The more telling picture is shown in Tables 10 and 11 where matches were conducted based on Internal frame's operations having an industrial hemp grower's license or not.

Table 9: Number and Percent of Operations Matched Across Two or More Frames¹

State	Total Operations	Internal to Contractor Matches			Internal to USDA Matches			Contractor to USDA Matches			Operations Matching Across All Three Frames					
		Operation Count	Percent of Internal	Percent of Contractor	Operation Count	Percent of Internal	Percent of USDA	Operation Count	Percent of Contractor	Percent of USDA	Operation Count Per Frame	Percent of Internal	Percent of Contractor	Percent of USDA	Operation Count of All 3 Frames	Percent of All Operations
Colorado	6244	1342	52.5%	80.6%	1034	40.4%	51.2%	345	20.7%	17.1%	268	10.5%	16.1%	13.3%	804	12.9%
Illinois	1773	24	3.5%	18.9%	294	43.1%	30.5%	2	1.6%	0.2%	0	0.0%	0.0%	0.0%	0	0.0%
Missouri	1272	411	80.0%	77.5%	150	29.2%	65.8%	171	32.3%	75.0%	142	27.6%	26.8%	62.3%	426	33.5%
Montana	1252	142	21.7%	58.2%	217	33.2%	61.3%	3	1.2%	0.8%	2	0.3%	0.8%	0.6%	6	0.5%
Nevada	638	75	20.8%	72.1%	90	24.9%	52.0%	4	3.8%	2.3%	3	0.8%	2.9%	1.7%	9	1.4%
New York	1585	46	6.2%	83.6%	628	85.2%	79.2%	8	14.5%	1.0%	8	1.1%	14.5%	1.0%	24	1.5%
Tennessee	11543	3529	93.2%	95.2%	3591	94.9%	88.7%	3604	97.2%	89.0%	3445	91.0%	92.9%	85.1%	10335	89.5%
Seven States	24307	5569	59.9%	86.6%	6004	64.6%	70.0%	4137	64.3%	48.2%	3868	41.6%	60.1%	45.1%	11604	47.7%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

8.1 Internal Frames – Industrial Hemp Growers Licensing Two and Three-way Overlap/Non-Overlap

Next, the Internal frame was split into two groups: (1) having an industrial hemp grower's license, and (2) those not having a grower hemp license (processor, transporter, etc.) and/or being non-hemp businesses (medical marijuana, etc.). The Internal frame's industrial hemp growers had very few matches to the Contractor's frame, but the Internal frame's non-industrial hemp grower list had a high percentage of matches with the Contractor. This likely occurred because the Contractor also scraped marijuana growers and other non-industrial hemp grower entities. In Section 5.2, the Contractor frame having each operation denoted with a "Business Type" (industrial hemp grower, processor, marijuana distributor, shipper, etc.) was mentioned. Thus, this high match to the Internal frame's non-industrial hemp growers adds credence to the validity of the "Business Type" category, which was originally deemed unreliable.

Contractor frame matched very few Internal (Industrial Hemp) operations by state (except for Tennessee) as shown in Table 10.

Table 10: Number and Percent of Operations Matched Across Two or More Frames Focusing on Internal Frame’s Industrial Hemp (Licensed for Growing) Operations¹

State	Internal (Ind Hemp) to Contractor				Internal (Ind Hemp) to USDA				Operations Matching Across All Three Frames						
	Operation Count	Percent of Internal (Ind Hemp)	Percent of Contractor	Percent of All Internal to Contractor Matches	Operation Count	Percent of Internal (Ind Hemp)	Percent of USDA	Percent of All Internal to USDA Matches	Operation Count Per Frame	Percent of Internal (Ind Hemp)	Percent of Contractor	Percent of USDA	Operation Count of All 3 Frames	Percent of All Operations	Percent of All 3-Way Matches
Colorado	332	27.9%	19.9%	24.7%	1029	86.5%	50.9%	99.5%	267	22.5%	16.0%	13.2%	801	12.8%	99.6%
Illinois	1	0.2%	0.8%	4.2%	272	51.3%	28.2%	92.5%	0	0.0%	0.0%	0.0%	0	0.0%	-
Missouri	149	76.4%	28.1%	36.3%	149	76.4%	65.4%	99.3%	141	72.3%	26.6%	61.8%	423	33.3%	99.3%
Montana	0	0.0%	0.0%	0.0%	200	73.8%	56.5%	92.2%	2	0.7%	0.8%	0.6%	6	0.5%	100.0%
Nevada	2	0.9%	1.9%	2.7%	62	27.0%	35.8%	68.9%	2	0.9%	1.9%	1.2%	6	0.9%	66.7%
New York	8	1.2%	14.5%	17.4%	624	96.1%	78.7%	99.4%	8	1.2%	14.5%	1.0%	24	1.5%	100.0%
Tennessee	3474	95.7%	93.7%	98.4%	3569	98.3%	88.1%	99.4%	3426	94.4%	92.4%	84.6%	10278	89.0%	99.4%
Total	3966	59.2%	61.6%	71.2%	5905	88.2%	68.8%	98.4%	3846	57.5%	59.8%	44.8%	11538	47.5%	99.4%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

Excluding Tennessee, the Contractor frame matched a high percentage of Internal Frames’s Non-Industrial Hemp operations. Also, the table shows that only 97 operations matched between USDA and Internal Frames’s Non-Industrial Hemp operations. See Table 11.

Table 11: Number and Percent of Operations Matched Across Two or More Frames Focusing on Internal Frames's Non-Industrial Hemp (Non-Licensed for Growing) Operations

State	Internal (Non Ind Hemp) to Contractor				Internal (Non Ind Hemp) to USDA/NASS				Operations Matching Across All Three Frames						
	Operation Count	Percent of Internal (NON - Ind Hemp)	Percent of Contractor	Percent of All Internal to Contractor Matches	Operation Count	Percent of Internal (NON - Ind Hemp)	Percent of USDA	Percent of All Internal to USDA Matches	Operation Count Per Frame	Percent of Internal (NON - Ind Hemp)	Percent of Contractor	Percent of USDA	Operation Count of All 3 Frames	Percent of All Operations	Percent of All 3-Way Matches
Colorado	1010	73.8%	60.6%	75.3%	5	0.4%	0.2%	0.5%	1	0.1%	0.1%	0.0%	3	0.0%	0.4%
Illinois	23	15.1%	18.1%	95.8%	22	14.5%	2.3%	7.5%	0	0.0%	0.0%	0.0%	0	0.0%	-
Missouri	261	81.8%	49.2%	63.5%	1	0.3%	0.4%	0.7%	1	0.3%	0.2%	0.4%	3	0.2%	0.7%
Montana	140	36.6%	57.4%	98.6%	17	4.4%	4.8%	7.8%	0	0.0%	0.0%	0.0%	0	0.0%	0.0%
Nevada	73	55.7%	70.2%	97.3%	28	21.4%	16.2%	31.1%	1	0.8%	1.0%	0.6%	3	0.5%	33.3%
New York	37	42.0%	67.3%	80.4%	2	2.3%	0.3%	0.3%	0	0.0%	0.0%	0.0%	0	0.0%	0.0%
Tennessee	55	35.5%	1.5%	1.6%	22	14.2%	0.5%	0.6%	19	12.3%	0.5%	0.5%	57	0.5%	0.6%
Total	1599	61.6%	24.9%	28.7%	97	3.7%	1.1%	1.6%	22	0.8%	0.3%	0.3%	66	0.3%	0.6%

Notes:

(1) All values exclude operations with an operation state other than the state of interest and all duplicate operations

The USDA frame had more unique operations (“*unique*” meaning found on *only one* frame), across the three frames compared to the Internal and the Contractor frames. This means web scraping did not account for all of the operations listed on USDA’s sources. Further research is needed to better understand this finding. See Table 12.

Table 12: Number and Percent of Matching and Unique Operations Across Two or More Frames^{1,2}

State	Internal		Contractor		USDA		All		Internal - Ind Hemp		Internal - Not Ind Hemp	
	Count	Percent of Total	Count	Percent of Total	Count	Percent of Total	Count	Percent of Total	Count	Percent of Total	Count	Percent of Total
Colorado												
Unique operations	450	17.6%	246	14.8%	909	45.0%	1605	25.7%	95	8.0%	355	25.9%
Matched operations (any match)	2108	82.4%	1420	85.2%	1111	55.0%	4639	74.3%	1094	92.0%	1014	74.1%
Total operations	2558	-	1666	-	2020	-	6244	-	1189	-	1369	-
Illinois												
Unique operations	364	53.4%	100	78.7%	668	69.3%	1132	63.8%	257	48.5%	107	70.4%
Matched operations (any match)	318	46.6%	27	21.3%	296	30.7%	641	36.2%	273	51.5%	45	29.6%
Total operations	682	-	127	-	964	-	1773	-	530	-	152	-
Missouri												
Unique operations	95	18.5%	90	17.0%	49	21.5%	234	18.4%	38	19.5%	57	17.9%
Matched operations (any match)	419	81.5%	440	83.0%	179	78.5%	1038	81.6%	157	80.5%	262	82.1%
Total operations	514	-	530	-	228	-	1272	-	195	-	319	-
Montana												
Unique operations	297	45.4%	101	41.4%	136	38.4%	534	42.7%	71	26.2%	226	59.0%
Matched operations (any match)	357	54.6%	143	58.6%	218	61.6%	718	57.3%	200	73.8%	157	41.0%
Total operations	654	-	244	-	354	-	1252	-	271	-	383	-
Nevada												
Unique operations	199	55.1%	28	26.9%	87	50.3%	314	49.2%	168	73.0%	31	23.7%
Matched operations (any match)	162	44.9%	76	73.1%	86	49.7%	324	50.8%	62	27.0%	100	76.3%
Total operations	361	-	104	-	173	-	638	-	230	-	131	-
New York												
Unique operations	74	10.0%	10	18.2%	165	20.8%	249	15.7%	25	3.9%	49	55.7%
Matched operations (any match)	663	90.0%	45	81.8%	628	79.2%	1336	84.3%	624	96.1%	39	44.3%
Total operations	737	-	55	-	793	-	1585	-	649	-	88	-
Tennessee												
Unique operations	110	2.9%	20	0.5%	300	7.4%	430	3.7%	13	0.4%	97	62.6%
Matched operations (any match)	3675	97.1%	3688	99.5%	3750	92.6%	11113	96.3%	3617	99.6%	58	37.4%
Total operations	3785	-	3708	-	4050	-	11543	-	3630	-	155	-
Total												
Unique operations	1589	17.1%	595	9.2%	2314	27.0%	4498	18.5%	667	10.0%	922	35.5%
Matched operations (any match)	7702	82.9%	5839	90.8%	6268	73.0%	19809	81.5%	6027	90.0%	1675	64.5%
Total operations	9291	-	6434	-	8582	-	24307	-	6694	-	2597	-

Notes:

- (1) All values exclude operations with an operation state other than the state of interest and all duplicate operations
- (2) Industrial Hemp status is based on whether having a grower's license or not

9. Cost Comparison of Building an Industrial Hemp Growers Frame (50 States)

USDA frame was the least expensive to build at \$66,354. The Contractor frame was the most expensive at \$461,933. The explanation, on how these estimated costs were derived for each frame, are provided below.

Table 13: Cost Comparison of Developing and Cleaning Each Frame (Extrapolating for 50 states)

Frame	Cash Expenses	Estimated NASS Staff Days	Estimated Time Frame	Estimated Total Cost
Internal	\$0	225	1 year	\$185,800
Contractor	\$400,000	75	9+ months	\$461,933
USDA	\$0	78	3+ months	\$66,354

9.1 Internal Frame

We estimated the cost of building a listing of industrial hemp growers for all fifty states internally at \$185,800 which includes salary and benefits.

This total was based on two NASS staff at the GS-13 Step 10 who were responsible for web crawling/scraping of industrial hemp growers into a database, data cleaning, enriching of information, and collapsing of duplicate records for the seven states studied. Using weekly and monthly reports, the two employees would need to work 225 staff days (4 hours per day per employee, over a one-year time frame) to accommodate all fifty states. The reason for the half days is that there was a noticeable drop in quality if employees did this same work for multiple days in a row without a break. The salary plus benefits was calculated from the employees’ actual payroll and benefits statements. This cost could be removed if more of the processes were automated, though the time requirement would still be substantial.

9.2 Contractor Frame

The total cost of the Contractor frame is estimated to be \$461,933, based on the following information.

The contract cost was \$400,000. The contract took 5 months from initiation to completion. This process was streamlined since NASS had worked with the contractor before on web crawling/scraping projects. Multiple NASS staff (ranging from senior executives to GS-13s) were involved in authoring, negotiating requirements and in finalizing/funding the contract. An estimated ten NASS staff days were devoted to procuring and finalizing the contract.

Once finalized, a discovery meeting (four NASS staff for one day) with contractors was held, three informal meetings and multiple email correspondences over the four-month web crawling/scraping project. Two NASS staff compiled keywords and a database of probable websites to scrape. These processes were estimated to take twenty staff days.

The contractor provided NASS with an initial deliverable, a listing of possible hemp operations. Two NASS staff reviewed the initial deliverable, recorded items to improve upon and participated in multiple meetings with the contractor. This process required a total of five staff days.

After receiving the final deliverable, NASS staff reviewed the listing of operations, removed gas stations, grocery stores and hardware stores, marked and combined multiple duplicate operations information into one record. These processes took approximately forty staff days to complete.

Overall, the complete process of procuring a contract to having a finalized listing took 75 NASS staff days. Hence, the total cost of the contractor frame was \$400,000 + 75 NASS staff days. Using the same pay scale rate as used above for the internal frame, we obtain a grand total of $\$400,000 + 61,933 = \$461,933$.

9.3 USDA Frame

We estimated the USDA frame took 78 staff days over three months. This equates to approximately \$66,354. The time and costs to build the USDA frame is explained below.

NASS has 45 state statisticians (GS-14 pay level) covering the United States. These statisticians are the primary liaisons between NASS and each state's department of agriculture office. These state statisticians were also responsible for working with their state's department of agriculture's colleagues in compiling a listing of industrial hemp growers. This involved contacting (telephone and email (with follow-ups) each state department of agriculture office. An estimated total of 23 staff days (45 state statisticians working for 4 hours each) on this task.

After receiving these lists, some data clean-up and data enrichment by the state statisticians were conducted. This includes adding street addresses and phone numbers. Estimating one day per state statistician, equates to 45 staff days total.

An additional internal list building effort involved obtaining data from FSA. This process to obtain, compile and review the information took one GS-14 statistician one day.

NASS also obtained hemp grower listings from USDA's AMS and RMA . To compile and transform into a useable format took one GS-14 two staff days.

All lists, mentioned above, were merged and duplicate growers were removed. This took a GS-14 five staff days. A final review before sending to NASS's list frame unit to be added to the agricultural operations was conducted. This took two days by one GS-14.

The time equates to 78 staff days, which equate to \$66,354. This was based on using a conservative GS-14 Step 5 (Washington D.C.) pay rate of \$191,407 per year which includes benefits). One caveat is that the GS-14 pay rate used was based on Washington DC's pay rate which may be higher or lower for the 45 state statisticians depending on location. The Washington D.C. pay rate to provide a conservative number.

10. Capabilities and Limitations of Each Frame

The capabilities and limitations of each frame were examined.

10.1 Internal Frame

- Capabilities
 - Expertise in determining keywords and in manually web crawling for websites to be scraped
 - Web scrape and data cleanse 100,000 operations per year
 - Provide expertise in subject matter
 - Obtain complete data (contact information, business type)
 - Work on additional NASS projects at the same time
- Limitations
 - Requires additional NASS staff to scrape over 100,000 operations
 - Requires additional resources to scrape complex websites that require multiple virtual machines with multiple IP addresses

10.2 Contractor Frame

There are several articles including (Schooley 2023) and (Indeed 2022) on when to employ a contractor. There are also articles specifically on when to hire a contractor for web crawling/scraping, (Jones 2023) and (Rudzinski 2022). Three similar themes of when to hire a contractor are: 1) Not enough in-house staff available 2) Requires a specialized expert skillset/knowledge not available in-house and 3) One time event/process.

- Capabilities
 - Ability to scrape over 100,000 records in one year via manual and automated processes
 - Expertise to accommodate certain website's data security and scraping complexities

- Example: Requiring multiple virtual machines with multiple IP addresses to repeatedly scrape information
- Limitations
 - Most expensive of the three options
 - Requires actual financial dollars to procure a contract
 - NASS staff time required for daily meetings with the Contractor
 - NASS staff required to web-crawl/scrape the entity of interest first to uncover any issues to be discussed with Contractor
 - NASS staff time to determine and supply Contractor with initial inputs (keywords, websites for scraping) to save the Contractor time and thus NASS finances in the signed contractual agreement
 - NASS staff required to review initial deliverables for any content irregularities
 - NASS staff time required to clean and enrich the information

10.3 USDA Frame

- Capabilities
 - Easy to conduct
 - Requires contacting fellow USDA agencies for information
 - Least expensive of the three options
- Limitations
 - Additional NASS staff required for data clean-up
 - Lacks full contact information

11. Discussion & Lessons Learned

NASS continues to search for ways to cost-effectively improve various business processes. Web crawling/scraping is one way for NASS to build a survey frame and a way to improve its coverage of agricultural operations across the United States.

If NASS decides to contract web crawling/scraping, suggest continuing to author the contract to ensure the ability to make minor modifications without incurring additional costs. Also, internal staff need to have daily (and sometimes) multiple daily virtual meetings with the contractor to answer questions and provide technical support.

If NASS decides to incorporate web crawling/scraping practices into its operational frame building program, then the agency should consider authoring a standards document on web crawling /scraping. This document would standardize the key steps for successfully building survey frames using web crawling/scraping. The document would include the following;

- 1) how to determine key words
- 2) best practices for web crawling and web scraping
- 3) framework for building a list of websites to scrape
- 4) framework for constructing a database of operations
- 5) how to find missing addresses and phone numbers
- 6) data clean-up
 - i. how to find, merge and denote duplicate records

- ii. how to find and remove scraped operations not of interest
- 7) how to conduct a pilot web crawling/scraping in-house to determine any challenges and how best to resolve those before a web crawling/scraping project occurs in a production environment

Additionally, NASS will have to decide on how best to update these new frames by either employing a contractor or via in-house expertise.

12. Recommendations

- Continue to use USDA sources (if available) for the initial development of a new survey frame
 - Employ internal staff to enrich (addresses and phone numbers) and add additional operations to the USDA frame through web crawling and web scraping
- Evaluate whether employing a contractor is necessary for web crawling/scraping
 - Recommend employing a contractor for large web scraping over 100,000 operations to be scraped in less than six months
 - Recommend employing a contractor when there are multiple websites containing over 2,000 operations per website that cannot be easily scraped in-house
 - If a contract is required:
 - Recommend authoring the contract to ensure the ability to make minor modifications without incurring additional costs
 - Internal staff need to hold daily (and sometimes multiple) virtual meetings with the contractor to answer questions and provide technical support
 - Recommend web crawling/scraping in-house first to determine any challenges and how best to resolve those challenges, before entering into a contract
- If NASS decides to incorporate web crawling/scraping practices into its operational frame-building program, then the agency should author a standards document on web crawling/scraping. This document would standardize the key steps for successful frame building using web crawling/scraping and would include the following:
 - 1) how to determine key words
 - 2) best practices for web crawling and web scraping
 - 3) framework for building a list of websites to scrape
 - 4) framework for constructing a database of operations
 - 5) how to find missing addresses and phone numbers
 - 6) data clean-up
 - i. how to find, merge, and denote duplicate records
 - ii. how to find and remove scraped operations not of interest

- 7) how to conduct a pilot web crawling/scraping in-house to determine any challenges and how best to resolve those before a web crawling/scraping project occurs in a production environment

13. Conclusion

Overall, USDA and Internal frames provided more probable industrial hemp grower operations than the Contractor frame.

In the future, recommend that NASS use USDA's lists as a starter frame and employ internal staff to enrich (addresses and phone numbers) and add additional operations to the USDA frame through web crawling/scraping. This would provide the best frame requiring the least amount of resources (staff and financials).

14. References

- AskUSDA, (2023). "What is Industrial Hemp", U.S. Department of Agriculture. Link is <https://ask.usda.gov/s/article/What-is-Industrial-Hemp#:~:text=Industrial%20hemp%20is%20from%20the,controlled%20under%20drug%20enforcement%20laws>
- Barcaroli, G., D. Fusco, P. Giordano, M. Greco, V. Moretti, P. Righi, and M. Scarno (2016). "ISTAT Farm Register: Data Collection by Using Web Scraping for Agritourism Farms", In Proceedings of the ICAS VII Seventh International Conference on Agricultural Statistics, 1017-1086. DOI: <https://doi.org/10.1481/icasVII.2016.f29d>
- Bosch, O. ten, D. Windmeijer, A. Van Delden, and G. Van den Heuvel (2018). "Web Scraping Meets Survey Design: Combining Forces", In Proceedings of the BIGSURV18 Conference, available at: https://www.europeansurveyresearch.org/bigsurv18/uploads/73/61/20180820_BigSurv_WebscrapingMeetsSurveyDesign.pdf
- Drotelff, L., (2020). "2020 Outlook: Licensed US hemp acreage falls 9% from 2019, but grower numbers increase 27%", Hemp Industry Daily. Link is: <https://hempindustrydaily.com/2020-outlook-licensed-u-s-hemp-acreage-falls-9-from-2019-but-grower-numbers-increase-27/>
- Fulton, B. R., and King, D. P. (2022). "Using Google Maps to Generate Organizational Sampling Frames", SocArXiv Pre-print. DOI: <https://doi.org/10.31235/osf.io/qtu8n>
- Hyman, M., L. Sartore, and L. J. Young (2021). "Capture-Recapture estimation of characteristics of U.S. Local Food Farms Using a Web-Scraped List Frame", Journal of Survey Statistics and Methodology, 00, 1-26. DOI: <https://doi.org/10.1093/jssam/smab008>
- Indeed (2022), The Pros and Cons of Hiring an Employee vs. a Contractor, Link: <https://www.indeed.com/career-advice/career-development/hiring-employee-vs->

contractor

- Jones, A. (2023). “How Much does Web Scraping Cost?”, Octoparse. Link: <https://www.octoparse.com/blog/how-does-web-scraping-cost>
- Rhodes, B. B., A. E. Kim, and B. R. Loomis (2016). “Vaping the Web: Crowdsourcing and Web Scraping for Establishment Survey Frame Generation”, In Proceedings of the 2015 Federal Committee on Statistical Methodology Research Conference, available at: https://nces.ed.gov/fcsm/pdf/H3_Rhodes_2015FCSM.pdf.
- Rudzinski, R. (2022). “When Should You Hire a Dev Agency vs a Contractor vs an Internal Developer”, Tragic Media. Link: <https://tragic.media/blog/startup/when-should-you-hire-dev-agency-vs-contractor-vs-internal-developer>
- Schooley, S. (2023), When to Hire a Full-Time Employee vs. Contractor, Link: <https://www.business.com/articles/contractors-vs-employees-benefits-and-drawbacks/>
- Vote Hemp, (2018), “2018 U.S. Hemp Crop Report, Vote Hemp”, Link: <https://www.votehemp.com/wp-content/uploads/2019/06/Vote-Hemp-Crop-Report-2018-v2.pdf>
- Young, L. J., M. Hyman, and B. R. Rater (2018). “Exploring a Big Data Approach to Building a List Frame for Urban Agriculture: A Pilot Study in the City of Baltimore”, Journal of Official Statistics, 34(2), 323-340. DOI: <http://dx.doi.org/10.2478/JOS-2018-0015>

15. Related Articles & Publications

- Arora, S. K., S. Kelley, and S. Madhavan (2021). “Building a Sample Frame of SMEs Using Patent, Search Engine, and Website Data”, Journal of Official Statistics, 37(1), 1-30. DOI: <http://dx.doi.org/10.2478/JOS-2021-0001>.
- Johnson, P., and D. Williams (2010). “Comparing ABS vs. Landline INTERNAL Sampling Frames on the Phone Mode”, Survey Practice, 3(3), 1-10. DOI: <https://doi.org/10.29115/SP-2010-0012>.
- Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn, and A. H. Mokdad (2006). “Address-Based versus Random-Digit-Dial Surveys: Comparison of Key Health and Risk Indicators”, American Journal of Epidemiology, 164, 1019-1025. DOI: <https://doi.org/10.1093/aje/kwj310>.
- Lo, A., S. Srikukenthiran, M. Chen, K. N. Habib, and E. J. Miller (2020). “Impact of Multiple Sample Frames on Data Quality of Household Travel Surveys: The Case of the 2016 Transportation Tomorrow Survey”, Transportation Planning and Technology, 43(6), 553-570. DOI: <https://doi.org/10.1080/03081060.2020.1780707>.

- Young, L. J., and M. Jacobsen (2021). “Sample Design and Estimation When Using a Web-Scraped List Frame and Capture-Recapture Methods”, *Journal of Agricultural, Biological and Environmental Statistics*, 27, 261-279. DOI: <https://doi.org/10.1007/s13253-021-00476-w>.

Appendix A

Hemp Industry Daily’s “2020 Outlook: Licensed US hemp acreage falls 9% from 2019, but grower numbers increase 27%”, authored by Laura Drotleff, June 19, 2020, link is: <https://hempindustrydaily.com/2020-outlook-licensed-u-s-hemp-acreage-falls-9-from-2019-but-grower-numbers-increase-27/>

Excerpt from the article:

Table 3: Top states

States that have led the nation in hemp production since 2014 have seen a marked decrease in licensed acreage. Of the top 10 hemp states by acreage licensed in 2018, only two – New York and North Carolina – have increased licensed acres and registered growers. Consider:

State	Outdoor (acres)		Indoor (square feet)		Growers	
	2019	2020	2019	2020	2019	2020
Colorado	80,000	61,854	9 million	15.4 million	2,300	2,017
Kentucky	58,000	32,000	6 million	4.6 million	1,047	960
Montana	40,000	11,685	–	150,000	250	95
Nevada	9,145	3,678	1.3 million	734,903	154	96
New York	5,000	29,777	–	9,042,279	278	667
North Carolina	11,572	16,434	4.5 million	7,276,394	933	1,503
North Dakota	2,175	Currently unknown	–	–	38	
Oregon	51,313	29,604	7.7 million	7,309,873	1,449	1,129
Tennessee	37,416	51,000 (thru 7/1/20)	2,643	–	2,900	3,830
Wisconsin	16,100 (combined)	11,626 (combined)	–	–	1,240	696

Appendix B

2018 U.S. Hemp Crop Report, Vote Hemp, 712 H Street NE, Washington, DC 20002 , link: <https://www.votehemp.com/wp-content/uploads/2019/06/Vote-Hemp-Crop-Report-2018-v2.pdf>

The 2018 U.S. Hemp Crop Report from Vote Hemp found that 23 states grew (or started to grow) a total of 78,175 acres of hemp in 2018. Among those states, the five leaders were: Montana: 22,000 acres, Colorado: 21,578 acres, Oregon: 7,808 acres, Kentucky: 6,700 acres and Tennessee: 3,338 acres.

Figure 1: 2018 U.S. Hemp Crop Report

