

# IT Java Dev – Web Scraping

Technical Application Description

Application Overview and Functionality Documentation

Legal Considerations

Version 1.0

Application Development Team

## Table of Contents

Table of Contents .....	i
Summary of Changes .....	ii
Glossary .....	iii
Key Contacts .....	iv
1 Purpose .....	1
2 Web Scraping Issues Overview .....	1
3 CMB – California Medical Board / BreEZe .....	2
3.1 The LAC CMB Web Scraper .....	2
4 OIG – SAM .....	3
4.1 OIG .....	3
4.1.1 How the LAC OIG Web Scraper Works .....	4
4.2 SAM – System for Award Management Registration .....	7
5 AEVS – Medi-Cal Automated Eligibility Verification System .....	7
6 ADDENDUM: Manual Search Procedures .....	9
6.1 CMB - BreEZe Online Manual License Verification .....	9
6.1.1 Physician and Surgeon Licensing Types and Descriptions .....	9
6.1.2 CMB – BreEZe Manual Search Example .....	10
6.2 OIG, SAM Manual Searches .....	14
6.2.1 Manual Search for Excluded Practitioners .....	14
6.2.2 The LAC OIG Web Scraper and Exclusion Monitoring Application .....	17
6.2.3 SAM .....	19

## Summary of Changes

Version	Description	Created by	Date
0.1	First review draft with content from Justin Kim, Mikhail Liberchuck, Eleonora Groysman, Cindy Zhao; Placed manual description in Addendum	Carol Locus	06/09/2015
0.2	Combined SAM in same section as OIG	Carol Locus	06/10/2015
1.0	Final	Carol Locus	06/10/2015

## Glossary

---

<b>Term</b>	<b>Definition</b>
<b>AEVS</b>	Medi-Cal Automated Eligibility Verification System
<b>CMB</b>	California Medical Board
<b>LEIE</b>	List of Excluded Individuals and Entities
<b>OIG</b>	Office of Inspector General, under US Department of Health and Human Services
<b>SAM</b>	System for Award Management Registration

---

## Key Contacts

### IT Business Owner and Manager

Name	Email	Extension	Direct Line	Support Role
Robert [REDACTED]	[REDACTED]@lacare.org	[REDACTED]	( [REDACTED] )	Manager, Application Development and Support

### CMB

Name	Email	Extension	Direct Line	Support Role
Justin [REDACTED]	[REDACTED]@lacare.org	[REDACTED]	( [REDACTED] )	Senior J2EE Developer, Java

### OIG

Name	Email	Extension	Direct Line	Support Role
Mik [REDACTED]	[REDACTED]@lacare.org	[REDACTED]	( [REDACTED] )	Senior Java Developer

### SAM

Name	Email	Extension	Direct Line	Support Role
Ele [REDACTED]	[REDACTED]@lacare.org	[REDACTED]	( [REDACTED] )	Senior J2EE Developer, Java

### AEVS

Name	Email	Extension	Direct Line	Support Role
Cin [REDACTED]	[REDACTED]@lacare.org	[REDACTED]	( [REDACTED] )	Senior J2EE Developer, Java

## 1 Purpose


The purpose of this document is to provide technical information to IT and LAC management about key infrastructure and back end system information for IT/Java staff designing and implementing web scraping actions for CMB, OIG, SAM, and AEVS. The web scraping being performed automates public database searches and queries of provider and vendor data. This document offers information in order to facilitate support, maintenance, modifications and future updates regarding web scraping as described in this document, as necessary.


The purpose of these web scraping activities is to collect unprotected data in the public domain, offered by the either the state of California, or the federal government to protect itself from fraud. The federal data (from OIG, and later from SAM) is then compared to employees and vendors in the LAC network for the purpose of preventing known abusers from working with or providing services to the federal programs of Medicare and Medicaid.


## 2 Web Scraping Issues Overview


This paper describes the web scraping, applications and database descriptions in detail to address and clarify the possible legal issues regarding the procedures of web scraping, as related to finances and PHI. These issues are clarified in reference to how LAC IT reads the websites which are scraped, what is scraped and where the data collected is passed to.

Web scraping, as it is performed by LAC IT, is implemented in the same way that a browser works. In actuality, the web scraping, as described in this document, just automates the same keystrokes that would be done by a user sitting at his or her computer doing the query manually with a browser.

	<b>Note:</b> The illegal and criminal issues concerning web scraping are touched upon in two pieces of legislation: The Digital Millennium Copyright Act (DMCA) and the Computer Fraud and Abuse Act (CFAA). These acts were both conceived to give more protection to copyright owners. In a nutshell, these deem circumventing digital rights management that control access to copyrighted works as criminal, and give copyright owners greater protection of their digital content.
---	---

	<b>Note:</b> The web scraping that is implemented by the LAC IT Application Development department automates manual searches of data that is available online in the public domain. The data scraped is freely accessible with no accounts or passwords required to access the data.
---	--

	<b>Note:</b> The web scraping that is implemented by the LAC IT Application Development department automates manual searches without approaching any thresholds that would affect those websites bandwidth or capacity. Another words, the web scraping done by LAC IT is NOT producing any denial of service by over taxing these websites which would be illegal.
---	---

	<b>Note:</b> Hackers have not yet been able to piggy back any packets of information through a browser or web scraping that returns only text files (as is the case with the web scraping being done as described.) herein). A user would have to download an executable to infect the PC and network with a virus or other executable file.
---	--

### 3 CMB – California Medical Board / BreEZe

*The Mission of the Medical Board of California: The mission of the Medical Board of California is to protect health care consumers through the proper licensing and regulation of physicians and surgeons and certain allied health care professions and through the vigorous, objective enforcement of the Medical Practice Act, and to promote access to quality medical care through the Board's licensing and regulatory functions. Excerpted from: <http://www.mbc.ca.gov/> )*

The CMB refers to the California Department of Consumer Affairs to verify the status of practitioners.

*DCA BreEZe Online Services: Welcome to the California Department of Consumer Affairs (DCA) BreEZe Online Services. BreEZe is DCA's new licensing and enforcement system and a one-stop shop for consumers, licensees and applicants! BreEZe enables consumers to verify a professional license and file a consumer complaint. Licensees and applicants can submit license applications, renew a license and change their address among other services.*

*(Excerpted from [http://www.mbc.ca.gov/Breeze/License\\_Verification.aspx](http://www.mbc.ca.gov/Breeze/License_Verification.aspx) )*

#### 3.1 The LAC CMB Web Scraper

The LAC CMB / BreEZe website at: <https://www.breeze.ca.gov/datamart/searchByLicNumber.do>

The web scraper automates the search for a provider to search by:

- License
- Last name
- First name

The results are collected and placed in:

- Table XXXXXXXXXX PROVIDER
- Procedure: XXXXXXXXXX chglacoroverider()

The web scraper for CMB updates and saves two things and automatically saves to the MPD:

1. License status
2. Expiration date

The web scrape template for the search parameters are:

[http://www2.dca.ca.gov/pls/wllpub/WLLQRYNA\\$LCEV2.QueryView?P\\_LICENSE\\_NUMBER={0}&P\\_LTE\\_ID={1}](http://www2.dca.ca.gov/pls/wllpub/WLLQRYNA$LCEV2.QueryView?P_LICENSE_NUMBER={0}&P_LTE_ID={1})

The license numbers searched for come from the LAC Provider Table.



**Note:** This web scraping process does not touch or mix any PHI, as patient information is in different databases.

## 4 OIG – SAM

### 4.1 OIG

The purpose of the Office of the Inspector General (OIG) website, (of Health and Human Services) is to provide lists of those excluded and not excluded by the federal government Department of Health and Human Services (HHS) from being providers, employees, or vendors for the Medicare and Medicaid federal programs.

As stated on their website at <http://oig.hhs.gov/about-oig/about-us/index.asp> :

*The OIG Mission: Office of Inspector General's (OIG) mission is to protect the integrity of the Department of Health & Human Services (HHS) programs as well as the health and welfare of program beneficiaries...Since its 1976 establishment, OIG has been at the forefront of the Nation's efforts to fight waste, fraud and abuse in Medicare, Medicaid and more than 100 other HHS programs...OIG develops and distributes resources to assist the health care industry in its efforts to comply with the Nation's fraud and abuse laws and to educate the public about fraudulent schemes so they can protect themselves and report suspicious activities. (Excerpted from: <http://oig.hhs.gov/about-oig/about-us/index.asp> )*

The OIG website presents search results containing practitioners that are not permitted by the federal government to provide services in the Medicare, Medicaid programs.

In the case of OIG it is the federal government offering up the information. However, it seems that per the Privacy Act of 1974, it would be up to the individual to prevent such disclosures:

*The right of privacy has evolved to protect the ability of individuals to determine what sort of information about themselves is collected, and how that information is used... In matters of privacy, the FTC's role is one of enforcing privacy promises made in the marketplace. Several additional laws form the foundation on which the FTC carries out this charge: the Privacy Act of 1974 ...The Privacy Act of 1974 (5 U.S.C. § 552a) protects personal information held by the federal government by preventing unauthorized disclosures of such information. Individuals also have the right to review such information, request corrections, and be informed of any disclosures. The Freedom of Information Act facilitates these processes. (Excerpted from: [https://www.law.cornell.edu/wex/personal\\_information](https://www.law.cornell.edu/wex/personal_information) )*



### 4.1.1 How the LAC OIG Web Scraper Works

The LEIE database contains approximately 60,000 records which are the approximate number of people and entities that the federal HHS OIG website monitors.

The search and comparison and consolidation of all federal OIG and LAC OIG related data is automated with these components:

1. The consolidation of LAC medical employees and vendors into one list, from:
  - a. █core (employees)
  - b. █mon (vendors)
2. The re-mapped LEIE federal database, to match LAC data organization in:
  - a. █\_OIG
3. The final step and result is to electronically compare the two with the end result being an auto-generated report of excluded and non-excluded providers.



**Note:** To manually verify a name and social security number on OIG, staff needs to have and them manually input the social security number.



**Note:** The automated process, however, is totally discrete, and behind the scenes. No personal information needs to be input manually. The verification process is automated on the back end inside the software and computer equipment.



**Note:** This web scraping process does not touch or mix any PHI, as patient information is in different databases.

# IT Java Dev – Web Scraping-IT Functional Specification

The LAC OIG Exclusion Monitoring screen displays lists of employees, and vendors. The search fields are on the first line, just below the application title.

When the VENDORS tab is opened with a click, the vendors are displayed with various status codes, ACT, TERM, as illustrated here:

The screenshot shows the 'OIG EXCLUSION MONITORING - PILOT' application. At the top, there are navigation links for 'Home' and 'About'. Below this is the main title 'OIG EXCLUSION MONITORING'. A search section includes a dropdown for 'Search By Criteria' set to 'EqualTo(=)', a dropdown for 'Criteria' set to 'First Name', and a 'VALUE' input field. Below the search section are several tabs: 'EMPL.', 'VENDORS', 'CONSOLIDATION', 'OIG', 'VERIFICATION', and 'CLEAN'. The 'VENDORS' tab is active. A table displays a list of vendors with columns: 'Show Action', 'First Name', 'Last Name', 'SSN/SIN', 'Person ID', 'Status Code', 'City', 'State', and 'Zip'. Each row has a 'Select' button in the 'Show Action' column. At the bottom of the table, there is a blue bar with the text '12345678910...'. The table data is as follows:

Show Action	First Name	Last Name	SSN/SIN	Person ID	Status Code	City	State	Zip
Select	1055 Seventh, LLC		26-1484764	VESEVE01	ACT	Los Angeles	CA	90017
Select	1-800-Flowers.com Inc		11-3329949	VE1-800FLOW	ACT	New York	NY	10087-9901
Select	1-800-GOT-JUNK		95-480-9268	VE1-800-GOT	TERM	Buena Park	CA	90621
Select	1st Choice IT, Inc.		80-0549470	VEFIRS02	ACT	Encino	CA	91316
Select	2 Graphic Copy LA Inc		80-0588209	VE2GRA00	ACT	Los Angeles	CA	90017
Select	24 Hour Fitness		94-2899899	VE24FITNESS	ACT	Carlsbad	CA	92008
Select	3 Key Consulting, Inc.		26-4826073	VE3KEY	ACT	Woodland Hills	CA	91365-6571
Select	3M Company		41-0417775	VE3M00	ACT	St Paul	MN	55144-1000
Select	4 Imprint		39-1837105	VE4IMPRINT	ACT	Oskosh	WI	54901
Select	4 Service Inc.		57-1170617	VE4SER000	TERM	Encino	CA	91436
Select	401 JNH		26-3996876	VE40100	ACT	Los Angeles	CA	90071
Select	5R8, Inc.		11-3662065	VE5R8	TERM	Los Angeles	CA	90016
Select	70 Trees, LLC		20-1672429	VE70TR00	TERM	Evanston	IL	60201
Select	8020 Consulting LLC		46-1775726	VE8020C00	ACT	Woodland Hills	CA	91367
Select	9008 Group, Inc		95-3278704	VEACT1	ACT	Torrance	CA	90504

## IT Java Dev – Web Scraping-IT Functional Specification

The LAC OIG web scrape consists of doing the last verification with the web scrape inserting the employee social security number or the vendor tax ID number on the website to accomplish the final verification as illustrated below. The final verification report is returned as a completed cross checked list with the providers or vendors indicated with a No or a Yes. No means the employee or vendor is not excluded yes means yes, that employee or vendor is excluded from the federal programs.

This report page shows one Yes, for one that is excluded from the Medicare, and Medicaid networks:

### VERIFICATION REPORT



New Report

Created: 3/10/2015 2:57:00 PM

Person Id	Name Consolid.	OIG Id	Name_OIG	Result	Address Consolid.	Address_OIG
EM2400	Arturo Martinez	224059	ARTURO MARTINEZ	No		
EM2493	Janet Lee	221175	JANET LEE	No		
EM2498	Donna Stewart	240891	DONNA STEWART	No		
EM2516	James Brown	198248	JAMES BROWN	No		
EM2516	James Brown	198249	JAMES BROWN	No		
EM2516	James Brown	198250	JAMES BROWN	No		
EM2516	James Brown	198251	JAMES BROWN	No		
EM2516	James Brown	198252	JAMES BROWN	No		
EM2516	James Brown	198253	JAMES BROWN	No		
EM2516	James Brown	198254	JAMES BROWN	No		
EM2516	James Brown	198255	JAMES BROWN	No		
EM2516	James Brown	198256	JAMES BROWN	No		
EM2516	James Brown	198257	JAMES BROWN	No		
EM2562	Marlon Orellana	230069	MARLON ORELLANA	Yes, Match	1006 E. 32nd St, Apt C, Signal Hill, CA 90755	4236 NELSONBARK AVENUE, LAKEWOOD, CA 90712
EM2567	Melissa Johnson	217518	MELISSA JOHNSON	No		
EM2567	Melissa Johnson	217519	MELISSA JOHNSON	No		
EM2583	Mary Mitchell	226866	MARY MITCHELL	No		
EM2600	Karen Lawson	220960	KAREN LAWSON	No		
EM2658	Teresa Stevenson	240868	TERESA STEVENSON	No		
EM2675	Jaime Sanchez	236841	JAIME SANCHEZ	No		
EM2718	Rosa Hernandez	214357	ROSA HERNANDEZ	No		
EM2780	Lisa Reyes	234385	LISA REYES	No		

## 4.2 SAM – System for Award Management Registration

The Sam website About page give this information:

*The System for Award Management (SAM) is combining federal procurement systems and the Catalog of Federal Domestic Assistance into one new system. This consolidation is being done in phases. The first phase of SAM includes the functionality from the following systems:*

SAM.gov has replaced Central Contractor Registry, CCR.gov; Online Representations and Certifications Application ORCA.gov; and the Excluded Parties List System, EPLS.gov. Registration to Sam.gov is free.

The link to this federal service is: <https://www.sam.gov/portal/SAM/#1>

Currently web scraping this website is under research and development in the Java Department.

There is no web scraping being done on SAM at this time.

The goal is to fold this into the LAC OIG application to streamline and consolidate the verifying providers, medical employees, and vendors to verify their statuses and isolate those that need to be excluded from the network.

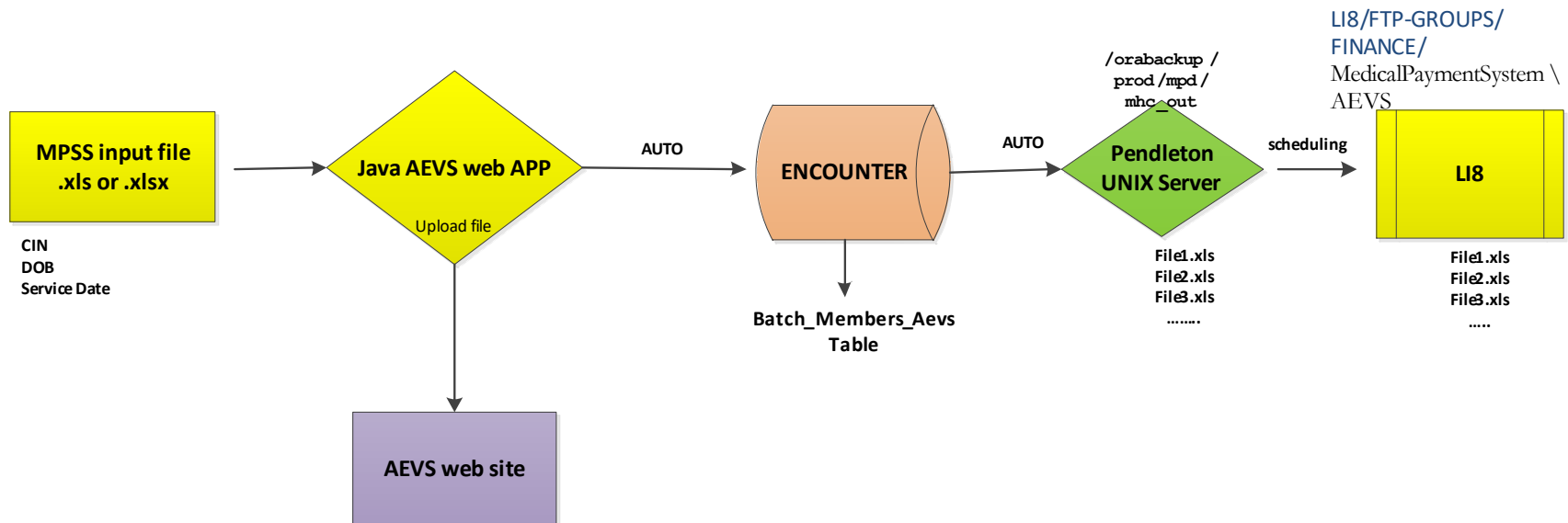
## 5 AEVS – Medi-Cal Automated Eligibility Verification System

The LAC procedure to verify Medical eligibility is a batch process as illustrated in this section.

The procedure begins with an input file from MPSS that is uploaded to the Java AEVS Web App that then searches (web scrapes) on the AEVS web site. The eligibility verification results load to the Encounter database which then auto generates a file that is uploaded to the Pendleton UNIX server. From there, the eligibility information is input into the LI8 finance medical payment system.

**AEVS Batch Process Work Flow for Medical Payment System and Services**

05/05/2015



## **6 ADDENDUM: Manual Search Procedures**

This addendum describes in detail how manual searches are done for CMB-BreEZe, IOG and SAM.

The web scraping described in the preceding sections automates these manual search procedures.

### **6.1 CMB - BreEZe Online Manual License Verification**

The CMB/BreEZe websites allow providers to be verified as being licensed, valid, in good standing and authorized to operate in California.

#### **6.1.1 Physician and Surgeon Licensing Types and Descriptions**

The BreEZe website offers manual searches with the following parameters:

- Choose a license type:
  - Physician and Surgeon
  - Search by Name
  - Search by License Number
  - Search by City
  - Search by County

## 6.1.2 CMB – BreEZe Manual Search Example

The link to search by a license number is:

<https://www.breeze.ca.gov/datamart/searchByLicNumber.do>

In this example the search is for 20454:


The screenshot shows the BreEZe search interface. At the top, there is a navigation bar with the CA.GOV logo, the Department of Consumer Affairs logo, and the BREZE logo. To the right of the logos are links for "About BreEZe", "FAQ's", and "Help Tutorials". Below the navigation bar is a "Skip navigation" link and "Logon | Contact Us" links. The main content area is titled "Search by License Number" and contains the following text: "Enter your search criteria and press 'Search' to find a license. Press 'Clear' to clear the search criteria. Press 'Back' to return to the previous screen." Below this text are four input fields: "Board:" with a dropdown menu, "License Type:" with a dropdown menu, "\* License Number (numeric portion only):" with a text input field containing "20454", and "Rows Per Page:" with a dropdown menu set to "5". To the right of these fields are three buttons: "Search", "Clear", and "Back". At the bottom of the page, there are links for "Back to Top", "Conditions of Use", "Privacy Policy", and "Accessibility", along with the copyright notice "Copyright © 2013 State of California".

The search returns a result which must be manually scanned and scrolled through and the name being searched for clicked on:

The screenshot shows the BREZE website interface. At the top, there is a navigation bar with the CA.GOV logo, the Department of Consumer Affairs logo, and the BREZE logo. To the right of the logo are links for "About BreZe", "FAQ's", and "Help/Tutorials". Below the navigation bar, there are links for "Skip navigation", "Logon", and "Contact Us". The main content area is titled "Search Results List" and contains instructions for navigating the search results. Below the instructions is a table with the following data:


Search Criteria	Search by License Number	
License Number:	20454	
Licensee Name	License Type	License Number
<a href="#">SHAW, LESLEY M (Previous Name)</a>	Psychologist	20454
<a href="#">SHEEHAN, MAUREEN C</a>	Public Health Nurse	20454
<a href="#">SIEGER, LANCE</a>	Physician and Surgeon G	20454
<a href="#">SIEGER, LANCE (Previous Name)</a>	Physician and Surgeon G	20454
<a href="#">TALISMAN, JENNIFER L</a>	Marriage & Family Therapist Intern	20454

Below the table, there are navigation buttons: "1 2 3 4 5 6", "First Page", "Previous Page", "Next Page", "Back", and "New Search". At the bottom of the page, there are links for "Back to Top", "Conditions of Use", "Privacy Policy", and "Accessibility", along with the copyright notice "Copyright © 2013 State of California".


 **Note:** The search box does not take any letters; numbers only may be entered. In this search result there are many with the same license number. There are letter codes denoting the area of practice.



The provider searched for was clicked on that the public record is displayed:



Department of Consumer Affairs



[About BreZe](#) | [FAQ's](#) | [Help/Tutorials](#)

[Skip navigation](#)  
[Logon](#) | [Contact Us](#)

### License Details

The Department of Consumer Affairs encourages you to verify the license statuses of any licensees that may appear in a 'Related License' section below. You can verify these licensees by selecting 'New Search' and conducting a new search using the 'Search by Personal or Business Name' option. Please note that the 'Related License' section will only appear below if this license is related to another license. Not all licensees have a related license.

If the License Details below include 'Date of Graduation', the month and date of graduation may not be available. In this instance it will be displayed as '01/01/YYYY' where YYYY represents the year of graduation. Please note that not all license types disclose 'Date of Graduation' on the License Details screen.

Press "Previous Record" to display the previous license.  
 Press "Next Record" to display the next license.  
 Press "Search Results" to return to the Search Results list.  
 Press "New Search Criteria" to do another search of this type.  
 Press "New Search" to start a new search.

**License Number: 20454** *Current Date: 06/08/2015 10:26 AM*

Name:	SIEGER, LANCE	
License Type:	Physician and Surgeon G	
License Status:	License Renewed & Current <span style="color: blue;">🔗</span>	
Expiration Date:	08/31/2016	
School Name:	PA012 - UNIVERSITY OF PITTSBURGH SCHOOL OF MEDIC	
Date of Graduation:	01/01/1969	
Original Issuance Date:	05/28/1971	

**Addresses**

Address of Record (Required)	Address	1000 W CARSON ST BOX 468 TORRANCE , CA LOS ANGELES 90509 <a href="#">View on a map</a>
------------------------------	---------	--

**Survey Information**

The following information is **self-reported by the licensee and has not been verified by the Board.**

Are you retired?	No
Activities in Medicine	Administration - 1.9 Hours Patient Care - 1.9 Hours Research - 1.9 Hours Teaching - 1.9 Hours Telemedicine - None
Patient Care Practice Location	Zip: 90502 County: LOS ANGELES
Patient Care Secondary Practice Location	Zip: 90502 County: LOS ANGELES
Telemedicine Practice Location	Not identified
Telemedicine Secondary Practice Location	Not identified
Current Training Status	Not in Training
Areas of Practice	Hematology - Primary Nuclear Medicine - Secondary Oncology - Secondary Pediatrics - Secondary
Board Certifications	American Board of Nuclear Medicine - Nuclear Medicine American Board of Pediatrics - Pediatric Hematology-Oncology
Postgraduate Training Years	5 Years
Cultural Background	Declined to Disclose
Foreign Language Proficiency	Declined to Disclose
Gender	Male

The non-physician search webpage example is at:

[http://www2.dca.ca.gov/pls/wllpub/WLLQRYNA\\$LCEV2.QueryView?P\\_LICENSE\\_NUMBER=1663&P\\_LTE\\_ID=835](http://www2.dca.ca.gov/pls/wllpub/WLLQRYNA$LCEV2.QueryView?P_LICENSE_NUMBER=1663&P_LTE_ID=835)

The screenshot shows the CA.GOV website interface. At the top, there is a navigation bar with the CA.GOV logo, a search bar, and links for 'About Us' and 'Contact Us'. Below the navigation bar, there are tabs for 'BOARD ACTIVITY', 'APPLICANT/REGISTRANT', 'LICENSEE', 'CONSUMER', and 'FORMS/PUBLICATIONS'. The main content area is titled 'SPEECH-LANGUAGE PATHOLOGY AND AUDIOLOGY BOARD'. It features a table with license details for RIFKIND PATRICE ANGELIQUE, including license type (Audiologist), license number (1663), status (VALID Definition), expiration date (February 29, 2016), issue date (March 22, 1996), and city (VALENCIA). Below this table is a section for 'Related Licenses/Registrations/Permits' with a table listing other licenses held by the individual. At the bottom, there is a 'Public Disclosure' section stating that no information is available from this agency and a disclaimer.

Licensee Name:	RIFKIND PATRICE ANGELIQUE
License Type:	Audiologist
License Number:	1663
License Status:	VALID Definition
Expiration Date:	February 29, 2016
Issue Date:	March 22, 1996
License or Registration Class:	Dispensing Audiologist. May dispense hearing aids without holding a hearing aid dispensers license
City:	VALENCIA
County:	LOS ANGELES
Actions:	No

Number	Name	Type	Status	Actions
1341	AGUILAR CYNTHIA	Speech-Language Pathology and Audiology Aide	VALID	No
413559	RIFKIND PATRICE	Branch Office	CANCELLED	No
423559	RIFKIND PATRICE	Branch Office	DELINQUENT	No
8663	LATKOVICH KIMBERLY A	Hearing Aid Dispenser - Trainee	CANCELLED	No

**Public Disclosure**

No information available from this agency

Public Record Action(s)

This information is updated Monday through Friday - Last updated: JUN-08-2015

**Disclaimer**

All information provided by the Department of Consumer Affairs on this web page, and on its other web pages and internet sites, is made available to provide immediate access for the convenience of interested persons. While the Department believes the information to be reliable, human or mechanical error remains a possibility, as does delay in the posting or updating of information. Therefore, the Department makes no guarantee as to the accuracy, completeness, timeliness, currency, or correct sequencing of the information. Neither the Department, nor any of the sources of the information, shall be responsible for any errors or omissions, or for the use or results obtained from the use of this information. Other specific cautionary notices may be included on other web pages maintained by the Department. All access to and use of this web page and any other web page or internet site of the Department is governed by the Disclaimers and Conditions for Access and Use as set forth at California Department of Consumer Affairs' Disclaimer Information and Use Information.

## 6.2 OIG, SAM Manual Searches

The purpose of OIG and SAM is to verify that providers, medical employees, and vendors are not excluded from federal programs and employment.

### 6.2.1 Manual Search for Excluded Practitioners

In the past, the Compliance Department would search for practitioners manually to check each one, one at a time. Data about the practitioner must be entered into various data entry fields. Once the data was entered and a search result was returned, it had to be scanned by staff, and the one to be verified had to be selected. Then the social security or Tax ID number that the practitioner provided had to be manually entered.

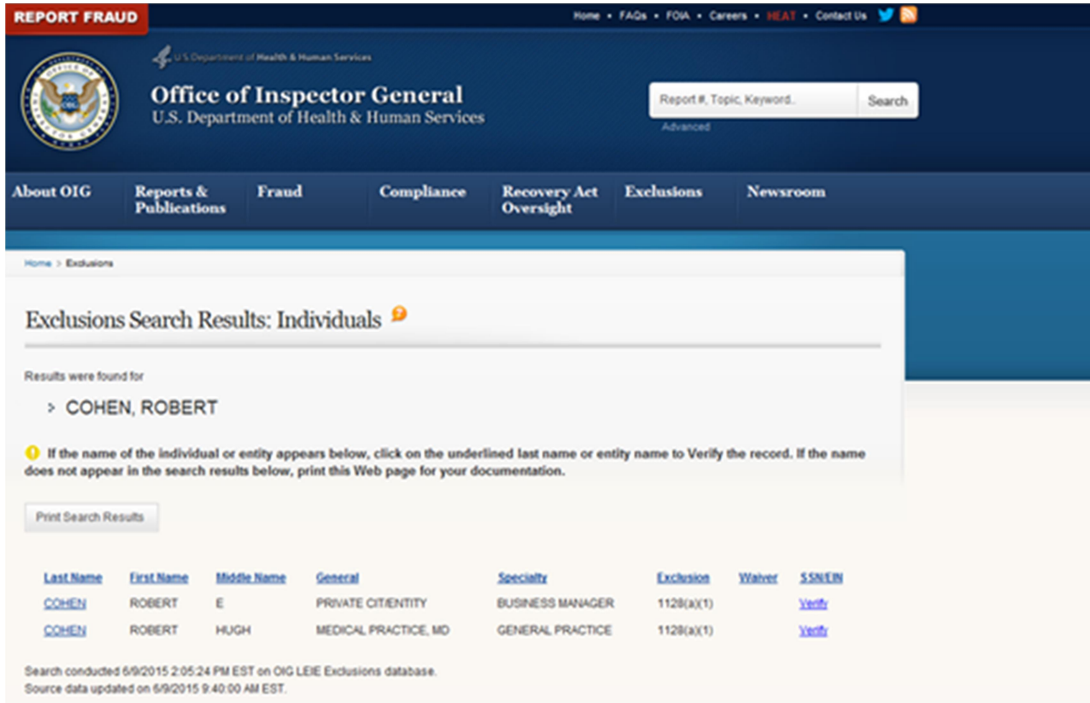
The screen for manual searches on the federal OIG website:



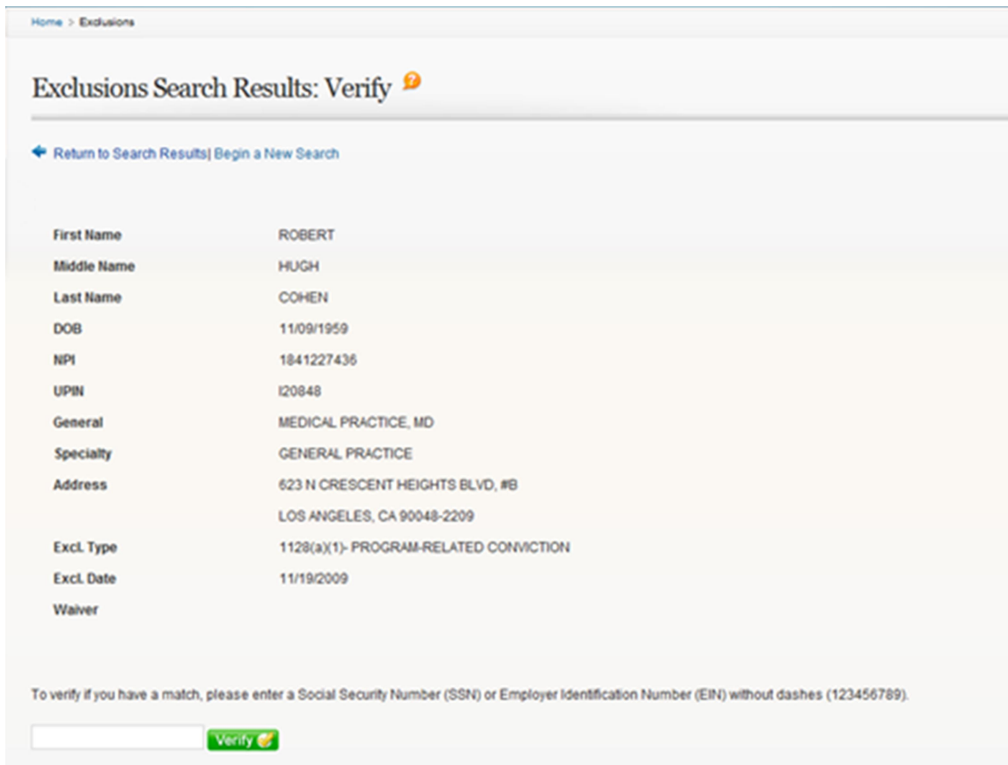
The screenshot shows the OIG website's search interface for the Exclusions Database. At the top, there is a navigation bar with a 'REPORT FRAUD' button and links for Home, FAQs, FOIA, Careers, HEAT, and CO. Below this is the OIG logo and the text 'U.S. Department of Health & Human Services Office of Inspector General U.S. Department of Health & Human Services'. A search bar is located on the right side of the header, labeled 'Report #, Topic, Keyword...' with an 'Advanced' link below it. The main navigation menu includes 'About OIG', 'Reports & Publications', 'Fraud', 'Compliance', 'Recovery Act Oversight', 'Exclusions', and 'Newsroom'. The current page is 'Home > Exclusions'. The main content area is titled 'Search the Exclusions Database' with a help icon. Below this is a section for 'Search For An Individual' with a help icon. There are three radio button options: 'Search For Multiple Individuals' (selected), 'Search For A Single Entity', and 'Search For Multiple Entities'. The search form has two input fields: 'Last Name' with the value 'COHEN' and '(and/or) First Name' with the value 'ROBERT'. At the bottom of the form are 'Search' and 'Clear' buttons.

# IT Java Dev – Web Scraping-IT Functional Specification

The manual search result; the provider must be manually checked and selected:

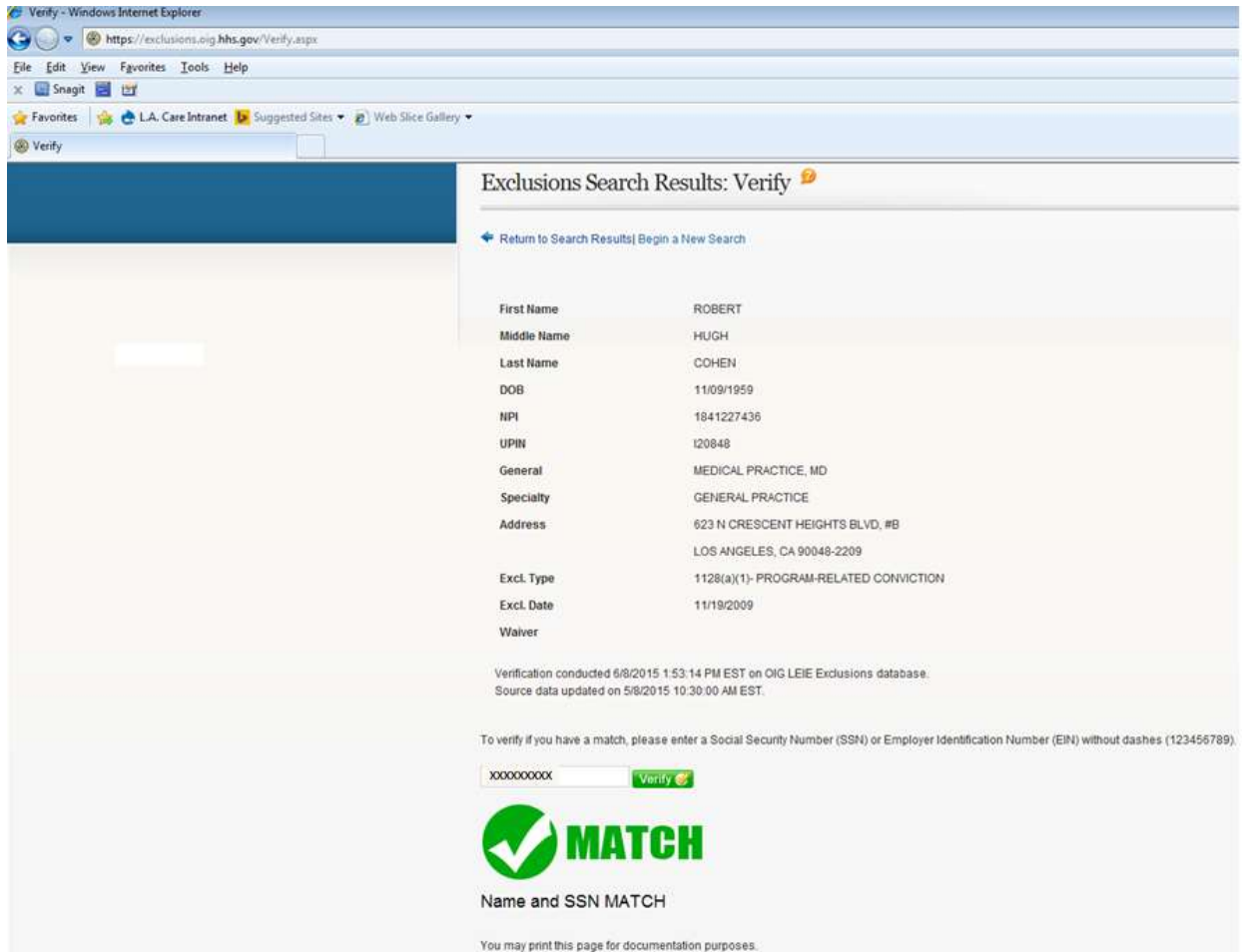


The manually selected search result, must be verified by inputting the social security number (the SSNs are not otherwise displayed from the website, only for the final verification to see if the provider is exclude from the federal medical networks:



## IT Java Dev – Web Scraping-IT Functional Specification

This screen capture illustrates the last step of the process. The Social Security Number (SSN) was input manually into the Name and SSN Match field. The Exclusion was matched, so it is verified and confirmed. In this example, the exclusion is due to a PROGRAM RELATED CONVICTION:




The screenshot shows a web browser window titled 'Verify - Windows Internet Explorer' with the URL 'https://exclusions.oig.hhs.gov/Verify.aspx'. The page content includes a search results table and a verification confirmation.

Exclusions Search Results: Verify	
First Name	ROBERT
Middle Name	HUGH
Last Name	COHEN
DOB	11/09/1959
NPI	1841227436
UPIN	I20848
General	MEDICAL PRACTICE, MD
Specialty	GENERAL PRACTICE
Address	623 N CRESCENT HEIGHTS BLVD, #B LOS ANGELES, CA 90048-2209
Excl. Type	1128(a)(1)- PROGRAM-RELATED CONVICTION
Excl. Date	11/19/2009
Waiver	

Verification conducted 6/8/2015 1:53:14 PM EST on OIG LEIE Exclusions database.  
Source data updated on 5/8/2015 10:30:00 AM EST.


To verify if you have a match, please enter a Social Security Number (SSN) or Employer Identification Number (EIN) without dashes (123456789).


XXXXXXXXXX


 **MATCH**

Name and SSN MATCH

You may print this page for documentation purposes.

 **Note:** To manually verify a name and social security number on OIG, staff needs to have them to manually input the social security number.

 **Note:** The automated process, however, is totally discrete, and behind the scenes. No personal information needs to be input manually. The verification process is automated on the back end inside the software and computer equipment.

 **Note:** This CMB web scraping process does not touch or mix any PHI, as patient information is in different databases.

## 6.2.2 The LAC OIG Web Scraper and Exclusion Monitoring Application

The purpose of the LAC OIG application is to compile consolidated lists of vendors and LAC medical employees to compare and find those excluded by the federal government Department of Health and Human Services (HHS) from being providers, employees, or vendors for the Medicare and Medicaid federal programs.

The product of LAC OIG application is the generating of monthly reports that illustrate the cross checks of LAC network providers against the federal OIG public website information.

There are two ways to verify a provider, manually as already described, or electronically automated.

The federal OIG generates monthly updates to their database that is available for free, public, unrestricted download, at: <http://oig.hhs.gov/>



Once the federal LEIE database is downloaded, it is remapped and input into the LAC OIG application to run the excluded reports after consolidating the vendors and employees to then compare to LEIE.

The end result is a report listing those practitioners that are excluded and not excluded (yes, no).

## IT Java Dev – Web Scraping-IT Functional Specification

The LAC OIG application front end is comprised of six tabs that sort and list the data collected:

1. EMPL = Employees at LAC
2. VENDORS = Vendors that provide services to LAC
3. CONSOLIDATION = LAC employees, vendors consolidated with data collected from OIG website
4. OIG = OIG information for that month
5. VERIFICATION = the end result of 'No' not exclude, or 'Yes' excluded
6. CLEAN = this tab resets the application for the new month of data comparisons

### 6.2.3 SAM

Manual searches on the Sam website may be done at:

[https://www.sam.gov/portal/SAM/?portal:componentId=24f38525-a064-4a77-b09f-3fa5986c2b48&interactionstate=JBPNS\\_r00ABXc0ABBFanNmQnJpZGdlVmlld0lkAAAAAQATL2pzZi9uYXZpZ2F0aW9uLmpzcAAHX19FT0ZfXw\\*\\*&portal:type=action#1](https://www.sam.gov/portal/SAM/?portal:componentId=24f38525-a064-4a77-b09f-3fa5986c2b48&interactionstate=JBPNS_r00ABXc0ABBFanNmQnJpZGdlVmlld0lkAAAAAQATL2pzZi9uYXZpZ2F0aW9uLmpzcAAHX19FT0ZfXw**&portal:type=action#1)

View assistance for Search Records

**SAM**  
SYSTEM FOR AWARD MANAGEMENT

USER NAME  **Quick Search enter your specific search term here.** ✕  
[Forgot Username](#) **Examples of search terms include the entity's name, etc.**

Enter Search Terms

HOME SEARCH RECORDS DATA ACCESS GENERAL INFO HELP

#### Search Records

Looking for entity registration records or entity exclusion records in SAM?  
\* Use **Quick Search** if you know an entity's Business Name, DUNS number or CAGE code.  
\* Use **Advanced Search** to structure your search using multiple categories and criteria.

Are you a Federal government employee?  
\* Create a SAM user account with your government e-mail address and log into SAM before searching to see FOUO information and registrants who chose to opt out of the public search.

Conducting small business-focused market research?  
\* In addition to what is contained in SAM, small businesses may supplement information about themselves in the SBA's [Dynamic Small Business Search](#).

Trying to find a contractor participating in the Disaster Response Registry?  
\* Use the new Disaster Response Registry Search to search for contractors willing to provide debris removal, distribution of supplies, reconstruction, and other disaster or emergency relief services in the event of a national disaster.

**QUICK SEARCH:**

(Example of search term includes the entity's name, etc.)

DUNS Number Search:

CAGE Code Search:

**ADVANCED SEARCH:**  
Use specific criteria in multiple categories to structure your search.