

Servicio Especializado de Sistematización de
Registros y Sistemas de Víctimas de Trata de
Personas

PLAN DE TRABAJO

Dr. José Manuel Magallanes, Ph.D

17 de Noviembre de 2017

1. Presentación

El servicio contratado busca tener datos integrados de víctimas de *la trata de personas*. La integración se hará en base a los registros de distintos sistemas que contengan información de estas víctimas. Las víctimas pueden tener información histórica.

La finalidad del trabajo es que, a partir de los datos integrados, se pueda identificar los lugares donde tal actividad ilícita se lleve a cabo con mayor incidencia, en relación a los demás lugares identificables del país.

2. Insumos requeridos

A fin de llevar a cabo este trabajo, es necesario que la Dirección General de Seguridad Democrática del Ministerio del Interior, o alguna oficina alternativa que designe ésta, provea los datos provenientes de:

1. Sistema de Registro y Estadística del delito de Trata de Personas y afines (RETA), que contienen información de la investigación de delitos de trata a nivel policial.
2. Sistema de Información Estratégica sobre Trata de Personas (SISTRA), información de la investigación de delitos de trata de personas en el Ministerio Público.
3. Sistemas de denuncias e información policial (SIDPOL y ESINPOL), sobre denuncias formuladas ante comisarías y unidades especializadas.
4. Registros de denuncias y/o investigaciones seguidos a nivel del Ministerio de Justicia (MINJUS).
5. Registros de atención de víctimas de trata en Centros de Emergencia Mujer, del Ministerio de la Mujer y Poblaciones Vulnerables (MIMP).
6. Registros de atención de víctimas de trata de los Centros de atención Residencial del MIMP

Estos datos no son de libre acceso, por lo que sólo personal autorizado puede solicitarlo.

3. Limitaciones del Trabajo

De la sección anterior se desprende que el consultor no puede iniciar su trabajo si el cliente desde el MINITER no nos proporciona oportunamente los insumos.

4. Actividades de la Consultoría

Este trabajo consta de una serie de pasos. La secuencia a seguir va desde el almacenamiento hasta la detección de patrones[1]. Estas etapas se describen a continuación.

4.1. Almacenamiento de la Datos

Los datos entregados por el cliente del MININTER será colectados y almacenados en un repositorio. Según el tamaño de los archivos entregados, los datos serán almacenados en repositorios tales como *Box*, *Dropbox*, *Github* o *AmazonWS*. Los repositorios serán sólo temporales, es decir, se existirán mientras dure el trabajo. La manipulación de los Datos se hará usando diversos programas, tales como R y Python, sin descartar el uso de SPSS y/o STATA. Ni el uso de repositorios, ni el tipo de programas estadísticos a utilizar generarán gastos adicionales al cliente del MININTER, ni generarán excusa de ningún tipo para demorar los entregables del trabajo.

4.2. Preparación de los Datos

Los datos, antes de ser analizados, pasarán por una serie de pasos, cada uno de ellos con diversa duración. Así tenemos:

4.2.1. Identificación de Escalas

Los valores a entregar pueden estar en diversos tipos de escalas, usándose para este caso la clásica diferenciación: *nominal*, *ordinal*, *intervalo* y *razón* [2]. Sabiendo en qué tipo de escala está una variable en particular, sabremos qué técnicas debemos elegir para los cálculos y visualizaciones posteriores. Al final, se habrá identificado la escala de cada variable, pero aun no se formateará pues primero habrá que verificar si los datos están *limpios*.

4.2.2. Limpieza de valores

Una vez identificada la escala de las variables (columnas) se procederá a verificar que los valores de cada celda estén no contengan caracteres que alteren los cálculos. Este es uno de los procesos más tediosos. Al final, se tendrán valores verificados y confiables.

4.2.3. Formateo de valores

Los datos limpios deben ser justamente transformados para que los programas estadísticos los reconozcan en la escala respectiva. Al final, se tendrá data con el formato que los programas reconocerán.

4.2.4. Integración de Datos

En esta etapa se integrará los datos de todas las tablas. Sólo se podrá integrar las tablas que tengan un campo *clave* común que permita la integración o *merge*. Si alguna tabla no tuviera campos en común, será tratada de manera individual, y se evaluará como esos datos contribuyen al análisis posterior. Al final se tendrá una tabla única con todas las tablas integrables.

4.2.5. Formateo de Datos

La data integrada podría necesitar formato diferente al de una tabla de datos clásica (resultado anterior). Al final, se tendrá la data integrada en el formato adecuado, entre los cuáles pueden ser: panel, serie temporal, red, data de sobrevivencia, existe la posibilidad de algún formato no convencional (*NoSQL*).

4.3. Análisis de Datos

Esta etapa permitirá conocer qué se tiene. Aquí comienza la etapa de exploración de la data integrada formateada; lo cual permite hacer análisis posteriores mucho más complejos.

4.3.1. Análisis Exploratorio

La exploración de datos permite saber cómo se comporta cada variable. El comportamiento sugiere la distribución de probabilidad que esta produciendo tales datos, las técnicas multivariadas dependen de saber tales distribuciones para justamente saber cómo estimar los parámetros de la población. En caso la distribución no fuese fácilmente identificable, deberemos utilizar técnicas *no paramétricas*[3].

4.3.2. Análisis Multivariado

El análisis multivariado permitirá saber qué tanto podemos organizar los casos usando toda o la mayor parte de la información disponible, y qué tanto podemos reducir la alta dimensionalidad de las variables originales en tres o dos dimensiones. Al final, la data integrada estará lista para saber qué indicadores compuestos se podrán proponer y calcular.

4.4. Detección de Patrones

4.4.1. Creación de Indicadores

Luego del análisis multivariado, podemos proponer los indicadores compuestos adecuados. Los indicadores serán calculados cambiando la unidad de análisis; es decir, considerando que la data integrada tiene en cada fila a una *persona*, éstas serán agregadas en una unidad geográfica comprensiva. De esta manera, los indicadores serán a nivel geográfico.

4.4.2. Minería de la Incidencia

La minería consiste en encontrar patrones y casos atípicos entre los indicadores. Según el tipo con los que se cuente, podremos encontrar diversos patrones a nivel nacional. Al final, se podrá producir la información requerida por el cliente del MINITER, saber dónde hay mayor incidencia.

5. Cronograma

Considerando el 14 de noviembre como el inicio de este trabajo (fecha de recepción de este trabajo), el cronograma, basado en semanas de cinco días, es el siguiente:

Actividades y Sub Actividades		Semanas								
		I	II	III	IV	V	VI	VII	VIII	IX
PLAN DE TRABAJO										
	Primer Entregable									
ALMACENAMIENTO DE LOS DATOS										
PREPARACION DE LOS DATOS										
	Identificación de Escalas									
	Limpieza de Valores									
	Formateo de Valores									
	Integración de Datos									
	Formateo de Tablas									
ANÁLISIS DE LOS DATOS										
	Análisis Exploratorio									
	Análisis Multivariado									
	Segundo Entregable									
DETECCIÓN DE PATRONES										
	Creación de Indicadores									
	Minería de la Incidencia									
	Tercer Entregable									

Referencias

- [1] Jose Manuel Magallanes Reyes. *Introducing to data science for social and policy research: collecting to organizing data with R and Python*. Cambridge University Press, Cambridge, United Kingdom New York, USA Port Melbourne, Australia Delhi, India Singapore, 2017. OCLC: 998518603.
- [2] Stanley S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946.
- [3] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric statistical methods*. John Wiley & Sons, Inc, Hoboken, New Jersey, third edition edition, 2014.