SEDESOL 2018

Ollin Demian Langle Chimal

Título del proyecto: USO DE DATOS MASIVOS PARA LA EFICIENCIA DEL ESTADO Y LA INTEGRACIÓN REGIONAL

Clave: ATN/OC 15822-RG

Puesto: Científico de Datos Senior

Recolección y limpieza de información

Entregable número: 1

Acrónimo del proyecto:	Estimación de Ingreso
Nombre completo del proyecto:	USO DE DATOS MASIVOS PARA LA EFICIENCIA DEL ESTADO Y LA INTEGRACIÓN REGIONAL
Referencia:	ATN/OC 15822-RG
URL del Proyecto:	http://www.plataformapreventiva.gob.mx

Tipo de Entregable:	Reporte (R)
Fecha de Entrega Contractual:	Abril - 2018
Fecha de Entrega	XXX
Número de Páginas:	41
Keywords:	estimación ingreso ciencia datos ingesta
Autor:	Ollin Demian Langle Chimal, Laboratorio de Datos, SEDESOL
Peer review:	ZZZ - Institution

Resumen

La Secretaría de Desarrollo Social (SEDESOL) es una entidad del gobierno mexicano destinada al apoyo de la población para el mejoramiento de sus condiciones de vida.

Un problema importante para SEDESOL es la correcta distribución de sus recursos por lo que es importante contar con una metodología que permita la generación de una focalización correcta para así poder ayudar a aquellos que realmente están en condiciones vulnerables.

En este reporte se detalla el proceso de obtención e integración de distintas fuentes de información que puedes ser de utilidad en dicho proceso, ya sea de manera directa o como auxiliar de las primeras. Las partes fundamentales para la realización de esto son, el estudio y evaluación de información generada por la misma Secretaría como lo es el Cuestionario Único de Información Socioeconómica (CUIS) y el Sistema de Focalización de Desarrollo (SIFODE), además de datos auxiliares que permitan la georreferenciación como archivos poligonales de carreteras y caminos. Para esto es importante llevar a cabo un sistema semiautomatizado para almacenar de manera adecuada todas las fuentes de datos para su futuro uso así como la información de los orígenes y naturaleza de las mismas.

Índice general

Índice de figuras

Índice de tablas

Lista de Acrónimos

CUIS Cuestionario Único de Información Socioeconómica

SIFODE Sistema de Focalización de Desarrollo

SEDESOL Secretaría de Desarrollo Social

ENIGH Encuesta Nacional de Ingresos y Gastos de los Hogares

PEA Población Económicamente Activa

LGDS Ley General de Desarrollo Social

INEGI Instituto Nacional de Estadística y Geografía

LBM Línea de Bienestar Mínimo

OSSE Organismos del Sector Social de la Economía

CUAPS Cuestionario Único para el Análisis de Programas Sociales

PUB Padrón Único de Beneficiarios

AGEB Área Geoestadística Básica

CENFEMUL Catálogo de Entidades Federativas, Municipios y Localidades

AWS Amazon Web Services

EC2 Elasting Compute Cloud

S3 Simple Storage Services

RDS Relational Database Service

CSV Comma Separated Values

1. Introducción

1.1. Fuentes de Información

La Secretaría de Desarrollo Social (SEDESOL) es la entidad mexicana encargada de dar ayuda y alivio a las personas que se encuentran en estado de marginación y/o carencia. Cuenta anualmente con un estimado de 19 programas sociales a nivel federal de los que se desprenden aproximadamente 330 subprogramas dando apoyo a más de 80 millones de mexicanos, lo que representa una covertura superior al 60 % del total de la población mexicana.

El mecanismo principal de integración a los programas sociales tiene como parte fundamental la aplicación del Cuestionario Único de Información Socioeconómica (CUIS)[5] el cuál es la base mínima de información de los universos de beneficiarios de cada programa. Algunos programas incorporan más preguntas a dicho cuestionario para obtener una mayor información referente a las condiciones específicas que pretenden mejorar. El CUIS es aplicado por hogar, el cual se define como el conjunto de personas que comparten techo, manutención y comida; es decir, en un inmueble pueden convivir diversos hogares si estos no comparten gastos y/o no comen de la misma preparación de alimentos.

Debido a que la información asentada en dicho cuestionario es la utilizada para que un programa decida otorgar un beneficio a un hogar, el aliciente para contestarlo de manera falaz es muy elevado pues el encuestado puede declarar ingresos menores o vulnerabilidades falsas con la intención de ser considerado dentro de la definición de población objetivo de cierto programa. Es importante considerar que la mayor parte de los programas no realizan una verificación domiciliar para corroborar las respuestas, pues dicho cuestionario es llenado por medio de una aplicación digital o en un módulo de SEDESOL.

De esta manera, es importante contar con un modelo estadístico que nos permita definir qué tan verosímil es que un hogar en partícular se encuentre dentro de las definiciones de vulnerabilidades o carencias. Por lo que este proyecto pretende presentar una metodología para realizar una estimación del ingreso por medio de información que se considere con alta probabilidad de ser verídica como lo es la versión del CUIS del programa PROSPERA[2] que realiza verificaciones a toda su población beneficiada.

El Sistema de Focalización de Desarrollo (SIFODE) es el encargado de consolidar la información socioeconómica de los hogares a través de los cuestionarios antes descritos, siendo posible que estos hayan sido beneficiarios anteriores de algún programa o no. A partir de los resultados de los cuestionarios el mismo SIFODE es quien está encargado de evaluar los mismos por medio del modelo multidimensional de pobreza dado por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL)[3].

1.2. Carencias Sociales

SEDESOL considera que la pobreza es causada en base a 6 tipos de carencias sociales, clasificadas de la siguiente manera:

- Seguridad Social
- Salud
- Educación
- Alimentación
- Vivienda
- Ingreso

Dichas carencias están definidas por CONEVAL[1] que a partir de la Ley General de Desarrollo Social (LGDS) aprovada en 2004 donde se establece la promoción de condiciones que aseguren el disfrute

de los derechos sociales, así como el impulso de un desarrollo económico con sentido social que eleve el ingreso de la población y contribuya a reducir la desigualdad**??**. Para poder garantizar lo anterior se creó el mencionado CONEVAL como un instrumento de evaluación y seguimiento de las políticas de desarrollo social. El mismo aunque es un organismo público, a la vez cuenta con autonomía técnica y de gestión. Entre sus deberes de ejercicio se encuentra establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza.

La LGDS establece que la medición de la pobreza efectuada por CONEVAL debe ser efectuada cada 2 años a nivel estatal y cada 5 a nivel municipal, utilizando la información generada por el Instituto Nacional de Estadística y Geografía (INEGI)[4] considerando al menos los siguientes indicadores:

- Ingreso corriente per cápita
- Rezago educativo promedio en el hogar
- Acceso a los servicios de salud
- Acceso a la seguridad social
- Calidad y espacios de la vivienda
- Acceso a los servicios básicos en la vivienda
- Acceso a la alimentación
- Grado de cohesión social

La necesidad de usar los ocho indicadores implica la generación de una medición multidimensional de la pobreza. Históricamente se ha utilizado una defínción unidimensional en la que sólo se considera el ingreso de una persona y su capacidad para adquirir una serie de productos indispensables, definiendo así un umbral para el cuál un hogar se considere pobre. En una formulación multidimensional el número y el tipo de dimensiones a considerar están directamente asociados a la forma en que se conciben las condiciones de vida mínimas o aceptables para garantizar un nivel de vida digno para todos y cada uno de los miembros de una sociedad.

Figura 1.1: Esquema de carencias

Definidas estas carencias se establece que *una persona se encuentra en situación de pobreza multidimensional cuando no tiene garantizado el ejercicio de al menos uno de sus derechos para el desarrollo social, y si sus ingresos son insuficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades*.

Según la ONU, la pobreza es una experiencia específica, local y circunstancial[6], lo que implica que una medida de pobreza no debe ser del todo universal sino tomar en cuenta situaciones particulares de los hogares o individuos tales como si se encuentran en un ambiente urbano o rural.

Definición: **HOGAR** - Conjunto de personas que hacen vida en común dentro de una misma vivienda, unidos o no por parentesco, que comparten los gastos de manutención y preparan los alimentos en la misma cocina.

Existen definiciones alternas,

Definición 2: **HOGAR** - Hogar es la persona o grupo de personas que habitan en la misma vivienda, se rigen bajo una única administración doméstica, hacen compras conjuntas de los productos de consumo básico (despensa), cocinan en el mismo lugar y "comen todos de la misma olla".

1.2.1. Programas

Programa	**Objetivo**	**Población Objetivo**
Pensión para adultos mayores	Contribuir a dotar de esquemas de seguridad social que protejan el bienestar socioeconómico de la población en situación de carencia o pobreza, mediante el aseguramiento de un ingreso mínimo, así como la entrega de apoyos de protección social a personas de 65 años de edad en adelante que no reciban una pensión o jubilación de tipo contributivo superior a la línea de bienestar mínimo.	Personas de 65 años de edad en adelante, mexicanas y mexicanos por nacimiento o con un mínimo de 25 años de residencia en el país, que no reciban pensión mayor a $1092 pesos mensuales por concepto de jubilación o pensión de tipo contributivo.

Atención a jornaleros agrícolas	Contribuir a fortalecer el cumplimiento efectivo de los derechos sociales que potencien las capacidades de las personas en situación de pobreza, incidiendo positivamente en la alimentación, la salud y la educación mediante la reducción de las condiciones de precariedad que enfrenta la población jornalera agrícola y los integrantes de sus hogares.	Población jornalera agrícola integrada por mujeres y hombres de 16 años o más que laboran como jornaleras y jornaleros agrícolas, así como los integrantes de su hogar en situación de pobreza que tienen su residencia o lugar de trabajo en las regiones de atención jornalera, ya sea de forma permanente o temporal.

Estancias infantiles para apoyar a madres trabajadoras	Contribuir a dotar de esquemas de seguridad social que protejan el bienestar socioeconómico de la población en situación de carencia o pobreza mediante el mejoramiento de las condiciones de acceso y permanencia en el mercado laboral de las madres, padres solos y tutores que buscan empleo, trabajan o estudian y acceden a los servicios de cuidado y atención infantil.	**Modalidad de Apoyo a Madres Trabajadoras y Padres Solos** En esta modalidad la población objetivo son las madres, padres solos y tutores que trabajan, buscan empleo o estudian, cuyo ingreso per cápita estimado por hogar no rebasa la LB y declaran que no tienen acceso a servicios de cuidadoy atención infantil a través de instituciones públicas de seguridad social u otros medios, y que tienen bajo su cuidado al menos a una niña o niño de entre 1 y hasta 3 años 11 meses de edad (un día antes de cumplir los 4 años), o entre 1 y hasta 5 años 11meses de edad (un día antes de cumplir los 6 años), en casos de niñas o niños con alguna discapacidad.

		Modalidad de Impulso a los Servicios de Cuidado y Atención Infantil En esta modalidad la población objetivo son las personas físicas o personas morales, que deseen establecer y operar una Estancia Infantil, o que cuenten con espacios en los que se brinde o pretenda brindar el servicio de cuidado y atención infantil para la población objetivo del Programa en la modalidad de Apoyo a Madres Trabajadoras y Padres Solos, conforme a los criterios y requisitos establecidos en las presentes Reglas de Operación y sus Anexos.
3X1 para migrantes	Contribuir a fortalecer la participación social para impulsar el desarrollo comunitario mediante la inversión en Proyectos de Infraestructura Social, Servicios Comunitarios, Educativos y/o Proyectos Productivos cofinanciados por los tres órdenes de gobierno y organizaciones de mexicanas y mexicanos en el extranjero.	La población objetivo la constituyen las localidades seleccionadas por los clubes u organizaciones de migrantes para invertir en proyectos de Infraestructura Social Básica, Equipamiento o Servicios Comunitarios, Educativos, así como Productivos, durante el ejercicio fiscal correspondiente, considerando el presupuesto disponible y de conformidad con las presentes Reglas.

Empleo temporal	Contribuir a dotar de esquemas de seguridad social que protejan el bienestar socioeconómico de la población en situación de carencia o pobreza, mediante la mitigación del impacto económico y social de las personas de 16 años de edad o más que ven disminuidos sus ingresos o patrimonio ocasionado por situaciones económicas y sociales adversas, emergencias o desastres.	Mujeres y hombres de 16 años de edad en adelante que ven afectado su patrimonio o enfrentan una disminución temporal en su ingreso por baja demanda de mano de obra o por los efectos de situaciones sociales y económicas adversas, emergencias o desastres.
Seguro de vida para jefas de familia	Contribuir a dotar de esquemas de seguridad social que protejan el bienestar socioeconómico de la población en situación de carencia o pobreza, mediante la incorporación de jefas de familia en condición de pobreza, vulnerabilidad por carencias sociales o vulnerabilidad por ingresos a un seguro de vida.	Jefas de familia que se encuentran en situación de pobreza, en situación de vulnerabilidad por carencias sociales o en situación de vulnerabilidad por ingresos.

Fomento a la economía social (opciones productivas)	Contribuir a mejorar el ingreso de personas en situación de pobreza-mediante el fortalecimiento de capacidades y medios de los Organismos del Sector Social de la Economía que adopten cualquiera de las formas previstas en el catálogo de OSSE, así como personas con ingresos por debajo de la línea de bienestar integradas en grupos sociales, que cuenten con iniciativas productivas.	Organismos del Sector Social de la Economía que adopten cualquiera de las formas previstas en el catálogo de OSSE, así como personas con ingresos por debajo de la línea de bienestar integradas en grupos sociales, que cuenten con iniciativas productivas.

Tabla 1.2: Programas y población objetivo

1.3. Información

Como se mencionó anteriormente, el SIFODE es la institución encargada de rocolectar e integrar los datos generado por el CUIS que sean de utilidad para la definición de pobreza multidimensional hecha por CONEVAL. Dicha información socioeconómica se clasifica en los siguientes apartados:

- Domicilio de la vivienda que habita regularmente el hogar
- Datos personales de los integrantes del hogar
- Variables correspondientes a la evaluación del model multidimensional, de acuerdo a la estimación del ingreso y seis carencias sociales
- Variables relacionadas a intervenciones de distintos programas sociales
- Contiene 3.6 millones de hogares y 11.9 millones de personas y se conforma por la información captada de 2011 a 2014 para hogares PEA, y los más recientes levantamientos

- Los logares PEA ascienden a 1.3 millones con 5.5 millones de personas

La entrevista del CUIS se realiza por hogar por lo que es importante que el encuestador identifique los distintos hogares que existen en una casa-habitación para la correcta detecctión de hogares en situación de carencias.

1.3.1. Condiciones

Condiciones por hogar:

- **Todo hogar debe de tener un JEFE**. Respecto a dicho jefe se establece la relación de parentesco de los demás integrantes.
- **No puede señalarse más de un JEFE DE HOGAR**.
- **Se debe considerar a TODOS los integrantes que viven normalmente en el hogar**. Incluyendo a no parientes, infantes o todo aquel con el que se compartan gastos.
- Se deben de incluir a las personas que usualmente viven en el hogar pero que por alguna situación se encuentran fuera; por vacaciones, trabajo o cuestiones de salud por ejemplo.
- Incluir a todas las personas que vivan temporalmente en ese hogar que no tienen algún otro lado para ir.
- **Obligatoriamente se debe considerar al jefe del hogar**. Aún cuando por alguna situación se encuentre residiendo en algún otro domicilio.
- **Personas a excluir**. Visitantes temporales, servientes o empleados cuando no comparten gastos.

2. Fuentes de datos

El objetivo del proyecto ATN/OC 15822-RG en general es el realizar una mejor focalización de los recursos para aquellos hogares que en verdad se encuentren en estado de carencia para así poder hacer una mayor diferencia, un objetivo particular también es el de realizar una estimación de ingreso con las variables con las que se cuentan, las cuales se obtienen a partir de las fuentes que se mencionaron en el capítulo de introducción.

2.1. SIFODE

SIFODE cuenta con una recopilación de datos de CUIS tomando en cuenta ciertas variables que son de interés para distintos ámbitos del propio sistema, el total de personas que están registradas en el mismo asciende a 39.9 millones.

2.1.1. SIFODE Calificación

Se le da el nombre de calificación a un set de datos que se utiliza para definir universos potenciales de participantes del SIFODE a partir de unas variables socioeconómicas prestablecidas. Como se ha mencionado anteriormente, la unidad básica es el hogar, por lo que existe un identificador único para cada uno de ellos que es llamado *LLAVE_HOGAR_H*, dicho identificador consta de 50 caracteres que identifican encuestas únicas y está integrado por el o los folios que identifican a una encuesta de manera única declarados por programa, a partir de este se genera otro identificador para cada una de las personas que componen el hogar, llamado *C_INTEGRANTE*, el cual consta de números consecutivos del 1 al número total de integrantes del hogar. Por último, existe una variable llamada *NEW_ID* el cual es un identificador único de personas en padrones correspondiente a SIFODE, este

se construye a partir del nombre, apellido paterno, apellido materno, fecha de nacimiento, entidad de nacimiento, sexo y CURP.

Las bases históricas de CUIS tienen *LLAVE_HOGAR_H* y *C_INTEGRANTE* y las bases de SIFODE tienen *LLAVE_HOGAR_H*, *C_INTEGRANTE* e *ID_UNICO_39_9*, que es el mismo que el *NEW_ID* por lo que es de esta manera en que se pueden relacionar CUIS, SIFODE y PUB.

Campo	Descripción
Identificadores únicos	
LLAVE_HOGAR_H	Identifica encuestas únicas integradas por el o los folios que identifican a una encuesta de manera única declaradas por el programa
C_INTEGRANTE	Número de renglón de integrante en la encuesta
ID_UNICO_39_9	Identificador Único de Personas en Padrones corresponiente a SIFODE 39.9
Indentificación de personas	
NUME_PER	Informante adecuado
TIE_CURP	La persona tiene CURP
NB_CURP	CURP
NB_NOMBRE	Nombre
NB_PRIMER_AP	Primer apellido
NB_SEGUNDO_AP	Segundo apellido
FCH_NACIMIENTO	Fecha de nacimiento
C_CD_EDO_NAC	Entidad de nacimiento
C_CD_SEXO	Sexo
EDAD	Años cumplidos al momento de la encuesta
EDAD_ACTUAL	Edad actual calculada a 2017
VAL_NB_RENAPO	Validación del CURPO con RENAPO
ANIO_NAC	Año de nacimiento

Clasificación de pobreza y carencias

POBREZAP	Personas en pobreza
POB_EXTREMP	Personas en pobreza extrema
POB_EXTREM_ALIMP	Personas en pobreza extrema alimentaria
POB_MODP	Personas en pobreza moderada
VUL_CARENP	Personas vulerables por carencia
VUL_INGRESOP	Personas vulnerables por ingresos
NP_NVP	Personas no pobres y no vulnerables
CARENCIADAP	Personas con al menos una carencia
CARENCIAS3P	Personas con al menos tres carencias
IC_REZEDU_1	Indicador de carencia por rezago educativo
IC_REZEDU15M	Personas de 3 a 15 años con rezago educativo
IC_REZEDU_81	Personas de 16 años o más que nacieron antes de 1982 y que no tienen primaria completa
IC_REZEDU_82	Personas de 16 años o más que nacieron antes de 1982 y que no tienen secundaria completa
JUBILADO_IC	Cuenta con jubilación o pensión como prestación laboral
IC_ASALUD	Indicador de acceso a servicios de salud (SIFODE)
SEG_POP_1	Personas afiliadas al Seguro Popular
IMSS_1	Personas afiliadas al IMSS
ISSSTE_1	Personas afiliadas al ISSSTE
PEMEX_1	Personas afiliadas al PEMEX, Marina o Defensa

PRIVA_1	Personas afiliadas a una clínica u hospital privado
SERV_SAL_99	Personas que no cuentan con servicio de salud
SERV_SAL	Acceso a servicio de salud
IC_SS	Indicador de carencia por acceso a la seguridad social
SSPEI	Personas económicamente inactivas sin acceso a seguridad social
SSPEA	Personas económicamente activas sin acceso a seguridad social
SSINTOCUP	Personas ocupadas sin acceso a seguridad social
SSAM	Adultos mayores sin acceso a seguridad social
IC_CV	Indicador de carencia por calidad y espacios de la vivienda
IC_PISO	Indicador de carencia del material de piso
IC_TECH	Indicador de carencia del material de techo
IC_MURO	Indicador de carencia del material de muros
IC_HAC	Indicador de carencia por hacinamiento
IC_SBV	Indicador de carencia en el acceso a los servicios básicos de la vivienda
ISBV_AGUA	Indicador de carencia de acceso al agua
ISBV_DREN	Indicador de carencia de servicio de drenaje
ISBV_LUZ	Indicador de carencia de servicio de electricidad
ISBV_COMBUS	Indicador de carencia de servicio de combustible para cocinar
IC_ALI	Indicador de carencia por acceso a la alimentación

INSEG_ALIM	Personas que cuentan con inseguridad alimentaria
IA_AD	Inseguridad alimentaria en adultos
IA_MEN	Inseguridad alimentaria en menores
SUMCARENP	Número de carencias a nivel persona
POB_LBM	Personas con ingreso estimado inferior a la LBM
POB_LB	Personas con ingreso estimado inferior a la LB

Información socioeconómica de personas

C_LENGUA_IND	Lengua indígena
OTRO_DIAL	Especificar otra lengua indígena
HABL_ESP	También habla español
INDIGENA	Se considera indígena
LEER_ESCR	Sabe leer escribir
C_ULT_NIVEL	Nivel escolar (último aprobado)
C_ULT_GRA	Grado (años aprobados)
NIV_ED	Nivel educativo
ASIS_ESC	Actualmente asiste a la escuela
C_CON_TRA	Condición laboral
C_POS_OCUP	Ocupación principal
C_CON_RES	Condición de residencia
C_CD_PARENTESCO	Parentesco al jefe(a) del hogar
PADRE	Renglón de identificación del padre en el hogar
MADRE	Renglón de identificación de la madre en el hogar
C_CD_EDO_CIVIL	Estado civil

CONYUGE	Num. Renglón del cónyuge
INT0A15	Personas de 0 a 15 años de edad
INT16A64	Personas de 16 a 64 años de edad
INT65A98	Personas de 65 a 98 años de edad
MUJ15A49	Mujeres de 15 a 49 años de edad
MUJER	Mujer
HOMBRE	Hombre
ACTA_NAC	Tiene acta de nacimiento
TRAB_SUBOR	En su trabajo principal del mes pasado tuvo un jefe o supervisor
TRAB_INDEP	En su trabajo principal del mes pasado se dedicó a un negocio o actividad por su cuenta
TRAB_PRESTA_A	Prestación laboral: incapacidad (enfermedad, accidente, maternidad)
TRAB_PRESTA_B	Prestación laboral: SAR o AFORE
TRAB_PRESTA_C	Prestación laboral: crédito para vivienda
TRAB_PRESTA_D	Prestación laboral: guardería
TRAB_PRESTA_E	Prestación laboral: aguinaldo
TRAB_PRESTA_F	Prestación laboral: seguro de vida
TRAB_PRESTA_G	Prestación laboral: no tiene ninguna
TRAB_PRESTA_H	Prestación laboral: no sabe o no responde
C_SALUD_HOGA	Donde se atienden los integrantes del hogar A
C_SALUD_HOGB	Donde se atienden los integrantes del hogar B
ENF_ART	Enfermedad diagnosticada: artritis
ENF_CAN	Enfermedad diagnosticada: cáncer
ENF_CIR	Enfermedad diagnosticada: cirrosis

ENF_REN	Enfermedad diagnosticada: deficiencia renal
ENF_DIA	Enfermedad diagnosticada: diabetes
ENF_COR	Enfermedad diagnosticada: enfermedades del corazón
ENF_PUL	Enfermedad diagnosticada: efisema pulmonar
ENF_VIH	Enfermedad diagnosticada: VIH
ENF_DEF	Enfermedad diagnosticada: deficiencia nutricional (anemia/desnutrición)
ENF_HIP	Enfermedad diagnosticada: hipertensión
ENF_OBE	Enfermedad diagnosticada: obesidad
TIENE_DISCA	Discapacidad tipo: caminar, moverse, subir o bajar escaleras
DISCA_ORI	Discapacidad origen: caminar, moverse, subir o bajar escaleras
DISCA_GRA	Discapacidad grado: caminar, moverse, subir o bajar escaleras
TIENE_DISCB	Discapacidad tipo: ver, o solo ve sombras aún usando lentes
DISCB_ORI	Discapacidad origen: ver, o solo ve sombras aún usando lentes
DISCB_GRA	Discapacidad grado: ver, o solo ve sombras aún usando lentes
TIENE_DISCC	Discapacidad tipo: hablar, comunicarse o conversar
DISCC_ORI	Discapacidad origen: hablar, comunicarse o conversar

DISCC_GRA	Discapacidad grado: hablar, comunicarse o conversar
TIENE_DISCD	Discapacidad tipo: oír, aún usando aparato auditivo
DISCD_ORI	Discapacidad origen: oir, aún usando aparato auditivo
DISCD_GRA	Discapacidad grado: oir, aún usando aparato auditivo
TIENE_DISCE	Discapacidad tipo: vestirse, bañarse o comer, desplazarse u otras de cuidado personal
DISCE_ORI	Discapacidad origen: vestirse, bañarse o comer, desplazarse u otras de cuidado personal
DISCE_GRA	Discapacidad grado: vestirse, bañarse o comer, desplazarse u otras de cuidado personal
TIENE_DISCF	Discapacidad tipo: poner atención, aprender cosas, o concentrarse
DISCF_ORI	Discapacidad origen: poner atención, aprender cosas, o concentrarse
DISCF_GRA	Discapacidad grado: poner atención, aprender cosas, o concentrarse
OTR_ING_A	Otros ingresos: es maestro de escuela de gobierno (federal, estatal, municipal)
OTR_ING_B	Otros ingresos: es dueño de una tienda
OTR_ING_C	Otros ingresos: es dueño de algún negocio
OTR_ING_D	Otros ingresos: es arrendatiario de algún transporte

OTR_ING_E	Otros ingresos: es doctor(a) o enfermero(a) de gobierno (federal, estatal, municipal)
OTR_ING_F	Otros ingresos: es servidor público de gobierno (federal, estatal, municipal)
OTR_ING_G	Otros ingresos: ninguna de las enteriores
INAPAM	Tiene tarjeta del instituto nacional de las personas adultas mayores (INAPAM)
AM_A	Recibe dinero por: programa pensión para adultos mayores
AM_B	Recibe dinero por: componente de apoyo para adultos mayores del Programa Prospera
AM_C	Recibe dinero por: otros programas para adultos mayores (estatal o municipal)
AM_D	Recibe dinero por: ningún programa para adultos mayores
AM_E	Recibe dinero por: no sabe/no responde

Información socioeconómica de hogares

CON_REMESA	El hogar recibe dinero proveniente de otros países
C_ESCUSADO	Tipo de baño o escusado de la vivienda
USO_EXC	El baño o escusado es para uso exclusivo de los habitantes de su hogar
FOGON_CHIM	Aparato que usa para cocinar
TS_VHS_DVD_BR	Indicadora de tenencia de VHS, DBD o BLU RAY
TS_REFRI	Indicadora de tenencia de refrigerador

TS_VEHI	Indicadora de tenencia de vehículo (carro, camioneta o camión)
TS_MICRO	Indicadora de tenencia de horno (microondas o eléctrico)
TS_TELEFON	Indicadora de tenencia de teléfono (fijo)
TS_COMPU	Indicadora de tenencia de computadora
TS_EST_GAS	Indicadora de tenencia de estufa / parrilla de gas
TS_INTERNET	Indicadora de tenencia de internet
TS_CELULAR	Indicadora de tenencia de celular
TS_TELEVISION	Indicadora de tenencia de televisión (microondas o eléctrico)
C_SIT_VIV	Relación con la vivienda en que habita (propia, hipotecada, rentada, etc.)
ESCRITURA1	Algún integrante del hogar tiene a su nombre las escrituras
ESCRITURA2	Otro integrante del hogar tiene a su nombre las escrituras
TIE_AGRI	Alguna persona del hogar posee o utilizó en los últimos 12 meses tierras para la agricultura o aprovechamiento forestal
PROP_TIERRA1	Las tierras pertenecen a algún integrante del hogar (propias)
PROP_TIERRA2	Las tierras pertenecen a más de un integrante del hogar (propias)
C_MAIZ	Tierras: cultiva maiz
C_FRIJ	Tierras: cultiva frijol
C_CERE	Tierras: cultiva cereales

C_FRUT	Tierras: cultiva frutos
C_CANA	Tierras: cultiva caña de azúcar
C_JITO	Tierras: cultiva jitomate
C_CHIL	Tierras: cultiva chile
C_LIMN	Tierras: cultiva limón
C_PAPA	Tierras: cultiva papa
C_CAFE	Tierras: cultiva café
C_CATE	Tierras: cultiva aguacate
C_FORR	Tierras: cultiva forrajes
C_NING	Tierras: no se cultiva
CUL_RIEGO	Para cultivar utiliza: sistemas de riego
CUL_MAQUINA	Para cultivar utiliza: maquinaria (tractor y/o otros)
CUL_ANIM	Para cultivar utiliza: ayuda de animales
CUL_FERORG	Para cultivar utiliza: composta / fertilizantes orgánicos
CUL_FERQUIM	Para cultivar utiliza: fertilizantes químicos
CUL_PLAGUI	Para cultivar utiliza: plaguicidas
CUL_RIEGO	Para cultivar utiliza: sistemas de riego
USO_HID_TRA	En el hogar se emplea la hidroponía o la agricultura de traspatio (huertos) para el cultivo de productos
PROYECTO	Algún integrante del hogar le gustaría realizar un proyecto productivo o de servicio
P_AGRI	Proyecto productivo: agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza

P_MANU	Proyecto productivo: manufactura (elaboración de productos)
P_COME	Proyecto productivo: comercio (compra-venta de bienes)
P_TRAN	Proyecto productivo: transporte (mercancías o personas)
P_PROF	Proyecto productivo: servicios profesionales, científicos y/o técnicos (oficios)
P_EDUC	Proyecto productivo: servicios educativos (capacitación)
P_SALD	Proyecto productivo: servicios de salud y de asistencia social (enfermería, cuidado de personas)
P_RECR	Proyecto productivo: servicios de esparcimiento, culturales y deportivos, y otros servicios recreativos
P_ALOJ	Proyecto productivo: servicios de alojamiento temporal y de preparación de alimentos y bebidas
P_COMU	Proyecto productivo: servicios de telecomunicaciones (café internet, casetas telefónicas)
P_OTRO	Proyecto productivo: otro
TIPO_PROY_ESP	Especificar otro proyecto

Tabla 2.2: Catálogo de variables socioeconómicas SIFODE 39.9

2.1.2. SIFODE Domicilio

Otra sección importante dentro de la información recolectada por el SIFODE es la correspondiente a la geolocalización de los hogares, la cual se encuentra en forma de domicilio y algunas referencias.

Campo	Descripción
Identificadores únicos	
LLAVE_HOGAR_H	Identifica encuestas únicas integradas por el o los folios que identifican a una encuesta de manera única declaradas por el programa
C_INTEGRANTE	Número de renglón de integrante en la encuesta
ID_UNICO_39_9	Identificador Único de Personas en Padrones corresponiente a SIFODE 39.9
Ubicación de personas y hogares	
S_CVE_ENTIDAD_FEDERATIVA	Clave de entidad
S_NOM_ENTIDAD_FEDERATIVA	Entidad federativa
S_CVE_MUNICIPIO	Clave de municipio
S_NOM_MUNICIPIO	Municipio o delegación
S_CVE_LOCALIDAD	Clave de localidad
S_NOM_LOCALIDAD	Localidad
S_TIPOLOC	Tipo de localidad
S_LONGITUD	Longitud
S_LATITUD	Latitud
S_CVE_AGEB	Clave de la cartografía censal al cierre del censo de población y vivienda 2010
S_CVE_MANZANA	Clave Manzana
S_C_TIPO_VIAL	Tipo de vialidad
S_NOMBRE_VIAL	Nombre de vialidad

S_CP	Código postal
S_C_TIPO_ASENTAMIENTO	Tipo de asentamiento humano
S_NOMBRE_ENTRE_VIAL_1	Nombre de vialidad entre la calle 1
S_C_TIPO_ENTRE_VIAL_1	Tipo de vialidad 1
S_C_TIPO_VIAL_POS	Nombre de vialidad entre la calle 2
S_NOMBRE_ENTRE_VIAL_2	Nombre de vialidad entre la calle 2
S_C_TIPO_ENTRE_VIAL_2	Codificación tipo de vialidad
S_NOMBRE_VIAL_POS	Nombre de la vialidad la calle de atrás
S_DESC_UBIC	Descripción de la ubicación
S_NUM_EXT	Número exterior
S_LETRA_EXT	Letra exterior
NUMEXTNUM2	Número exterior anterior
S_NUM_INT	Número interior
S_LETRA_INT	Letra interior
S_ORIGEN_DOMICILIO	Origen del domicilio
S_CARRETERA	Nombre compuesto de la carretera, conforme a la norma técinca sobre domicilios geográficos del INEGI
S_CAMINO	Nombre compuesto del Camino, conforme a la Norma técnica sobre domicilios geográficos del INEGI

Tabla 2.4: Información domiciliaria de SIFODE 39.9

Campo	Descripción
Identificadores únicos	
LLAVE_HOGAR_H	Identifica encuestas únicas integradas por el o los folios que identifican a una encuesta de manera única declaradas por el programa
C_INTEGRANTE	Número de renglón de integrante en la encuesta
ID_UNICO_39_9	Identificador Único de Personas en Padrones corresponiente a SIFODE 39.9
Indicadores	
EJERCICIO	Año de captura de la encuesta
CONSISTENCIA	Índice de consistencia de la información
VIGENCIA	Índice de la vigencia de la información
USABILIDAD	Índice de la usabilidad de la información
INT_CARE	Índice de la intensidad de la carencia

Tabla 2.6: Información de los indicadores de SIFODE 39.9

2.1.3. Red Carretera

A partir del año 2014 el INEGI ha liberado archivos shape[1] con la información de los caminos registrados dentro de la República Mexicana.

Campo	Descripción
Red Carretera	
wkt	Well Known Text, es una codificación o sintaxis en formato ASCII estandarizada diseñada para describir objetos espaciales expresados de forma vectorial
id_red	Un número secuencial que se incrementa con cada ocurrencia del objeto
tipo_vial	Clasificación que se le da al objeto espacial, en función de lo determinado por la autoridad local
nombre	Nombre de la vialidad
codigo	Si cuenta con algún código que represente dicho camino como una carretera
cond_pav	Condiciones en las que se encuentra el pavimento de la vialidad
recubri	Tipo de recubrimiento si es que la tiene
carriles	Número de carriles
estatus	Si se encuentra habilitado o en construcción
condicion	Si se encuentra en operación o no
nivel	Si es puente, camino o túnel

[1] Un shapefile es un formato sencillo y no topológico que se utiliza para almacenar la ubicación geométrica y la información de atributos de las entidades geográficas.

peaje	Pago correspondiente a los derechos de tránsito o c irculación por determinados lugares, como algunas autopistas, puentes, túneles, aduanas, etc.
administra	Clasificación que se le da al objeto espacial en fu nción de la responsabilidad de mantenimiento
jurisdi	Si es federal o no
circula	Los sentidos en los que se recorren los caminos
escala_vis	Clasificación que se le da a la carretera o vialidad para fines de representación de acuerdo al rango de escalas definido para cada valor
velocidad	Velocidad máxima sobre la vialidad
union_ini	Intersección inicial con otra vialidad
union_fin	Intersección final con otra vialidad
longitud	Longitud en metros de la vialidad
ancho	Ancho en metros de la vialidad
calirepr	Calificador de posición

Tabla 2.8: Variables dentro de los archivos shape de la Red Carretera

3. Infraestructura

3.1. Infraestructura física

Para la realización de la ingesta de estos datos se decidió utilizar la infraestructura en la bube de Amazon Web Services (AWS). Llamaremos infraestructura física a la que consta del equivalente al hardware en los servicios de nube. Los servicios de AWS que se utilizan son:

- **EC2**. Elasting Compute Cloud. Servidores con capacidad de cómputo multipropósito.
- **S3**. Simple Storage Service. Servicio de almacenamiento de datos.
- **RDS**. Relational Database service. Servicio de base de datos relacionales con PostgreSQL.

Con estas tecnologías podemos realizar una ingesta de datos tal que pase de sus formatos originales a una base de datos propia y así poder consumir fácilmente información de la misma por medio de queries.

3.1.1. EC2

Es la solución en la nube de AWS para servidores, es decir, poder de cómputo. El uso de este servicio es mediante una instancia alojada en los servidores de Amazon. Se utiliza un sistema operativo basado en Linux porque otorga la flexibilidad necesaria para llevar a cabo las tareas planteadas de manera semiautomatizada.

3.1.2. S3

La necesidad de almacenar datos de gran escala de forma eficiente y segura se puede cubrir usando el servicio de almacenamiento S3. El cual es altamente confiable y de sencilla integración con los

demás servicios de AWS. El costo de s3 frente a un volumen de disco duro convencional es tan sólo una fracción del mismo por MB además de ofrecer mayor seguridad e integridad de los datos.

3.1.3. RDS

El servicio de base de datos relacionales puede estar en parte manejado por la estructura de AWS permitiendo usar la capacidad de cómputo de una EC2 con la configuración del software de preferencia de manera sencilla. La selección de una base de datos es muy importante tomando en cuenta el tipo de datos que se van a almacenar en ella además de la forma en la que se desea acceder a los mismo. De esta manera, al pensar la estructura de datos de forma columnar y teniendo en mente la necesidad de uso de archivos de geometrías como los shape de la red carretera además del costo de licencia se optó por el uso de la base llamada PostgreSQL. Las ventajas de usar la tecnnología de PostgreSQL además de su robustez se basan principalmente en la capacidad de guardar objetos de tipo geométrico, abriendo la posibilidad de hacer cálculos geoespaciados gracias al paquete PostGIS que es una extensión del mismo PostgreSQL para dichos elementos.

3.2. Infraestructura de software

La infraestructura está basada en un sistema tipo UNIX basado en Linux porque además de ser abierto, también es compatible con todas las tecnologías que se van a utilizar.

3.2.1. Scripts

Una Shell de Unix es un intérprete de comandos, el cual es una interfaz textual de la computadora con el usuario que permite ejecutar comandos definidos por otros pedazos de software o programas, así mediante órdenes o instrucciones se puede comunicar con el kernel de la computadora. Para el manejo de datos por medio de Shell se utilizarán los siguientes programas:

- **AWK**. Es particularmente útil para el manejo de datos en listas indexadas y expresiones regulares.

- **SED**. Acrónimo de Stream Editor, sirve para leer y modificar flujos de datos pues lo hace línea por línea.
- **CURL**. Software de tranferencia de archivos.
- **UNZIP**. Descompresión de archivos zip.
- **CUT**. Extracción de segmentos de líneas.
- **OGR2OGR**. Conversión entre distintos tipos de archivos, entre ellos de shape a CSV.
- **CSVKIT**. Funcionalidades para trabajar con archivos CSV.

A su vez también se utilizan programas escritos en el lenguaje Python utilizando las siguientes librerías

- **PANDAS**. Añaden la capacidad de manejo de datos en forma de data frames con funcionalidades implementadas.
- **CSV**. Paquete para trabajar con archivos CSV.
- **SMART_OPEN**. Permite hacer streaming a archivos en S3.
- **BOTO**. Maneja la conexión con AWS.

Debido a que las fuentes de datos además de la naturaleza de los mismos es diferente, es común encontrarse con extensiones de datos distintas, esto es porque las mismas representan el tipo de datos que contienen dichos archivos y eso ayuda a identificar la forma de utilizarlos.

3.2.1.1. Pipeline

La forma de manejar la automatización de los datos se trabaja mediante el software llamado *Luigi* que es un orquestador manejado por dependencias y trabajando por batches. Las secciones del pipeline son:

- Source script. Descarga o hace stream del archivo fuente además de ordenarlo en un CSV separado por el caracter "—". Guarda temporalmente el archivo en local. Esta ingesta puede tener una temporalidad definida desde días hasta años.

- Diccionario. Se genera un diccionario de las variables del archivo el cual debe ser llenado a mano, una vez completado se guarda en S3.

- Base de datos. Se guardan tanto los archivos como sus diccionarios en la base de datos PostgreSQL.

- Linaje. Se genera un linaje de datos en una base de datos Neo4j que es orientada a grafos, así se generan relaciones con las tablas y sus columnas así como con su temporalidad.

4. Metadatos

4.0.1. Metadatos

Los metadatos sirven para entender los datos, lo cual es tan importante como los datos mismos por lo que es importante contar con diccionarios y catálogos que den información sobre la naturaleza de ellos.

4.0.1.1. Diccionarios

Los diccionarios tienen la siguiente estructura

Campo	Descripción
Identificadores únicos	
ID	Nombre de la columna
Nombre	Lo que representa el ID
Fuente	Fuente original de los datos
URL	Dirección url de la fuente si es que existe
Tipo	Tipo de la fuente de datos
Subtipo	Mayor granularidad del tipo de la fuente
Periodo	Periodo de descarga
Actualización_sedesol	La fecha en que se hizo una actualización de la base de datos
Metadata	Explicación de la variable
data_date	Fecha intrínseca al source script de descarga

Tabla 4.2: Diccionario de datos

4.0.1.2. Catálogo

Un catálogo de datos o DCAT es un documento con la información necesaria para entender el origen y la naturaleza de los datos en cuestión. Consta de un formato predefinido de manera homologada.

Bibliografía

[1] Coneval.

[2] Sitio de programa de inclusión social prospera.

[3] Consejo Nacional de Evaluación de la Política de Desarrollo Social. *Metodología para la medición multidimensional de la pobreza en México*. 2009.

[4] Instituto Nacional de Estad. Instituto nacional de estadstica y geografa (inegi).

[5] Dirección General de Geoestadística y Padrones de Beneficiarios. Instructivo Entrevistador CUIS, 2013.

[6] SEDESOL. El combate a la pobreza a través de la reducción de las carencias sociales.

A. Apéndice

Listing A.1: Archivo de ingesta de Red Carretera

```
#!/usr/bin/env bash

################
# Geom
# Red carretera INEGI
################

echo "Red de Carreteras INEGI"

year=$1
local_path=$2
local_ingest_file=$3

# Downloads
if [ $year = '2016' ]
then
    url="http://internet.contenidos.inegi.org.mx/contenidos/Productos/\
    prod_serv/contenidos/espanol/bvinegi/productos/geografia/caminos/2016/702825219000_s.zip"
elif [ $year = '2015' ]
then
    url="http://internet.contenidos.inegi.org.mx/contenidos/Productos/\
    prod_serv/contenidos/espanol/bvinegi/productos/geografia/caminos/2015/702825209575_s.zip"
elif [ $year = '2017' ]
then
    url="http://internet.contenidos.inegi.org.mx/contenidos/Productos/\
    prod_serv/contenidos/espanol/bvinegi/productos/geografia/caminos/2017/889463171836_s.zip"
elif [ $year = '2014' ]
then
    url="http://internet.contenidos.inegi.org.mx/contenidos/Productos/\
    prod_serv/contenidos/espanol/bvinegi/productos/geografia/caminos/702825278724_s.zip"
else
    echo 'url not defined for the selected year'
```

```
    exit 1
fi

curl $url > $local_path/red_carretera.zip

echo 'unzip'
unzip -o -d $local_path/temporal $local_path/red_carretera.zip

if [ $year = '2014' ]
then
    mv $local_path/temporal/producto/informa* $local_path/temp
    cd $local_path/temp
    rename -v s/Red_Vial/red_vial/ Red_Vial.*
    cd -
    ogr2ogr -t_srs EPSG:4326 -f CSV /vsistdout/ $local_path/temp/red_vial.shp -lco \
    GEOMETRY=AS_WKT | csvformat -D "|" | cut --complement -d'|' -f8 > $local_ingest_file
elif [ $year = '2016' ]
then
    mkdir $local_path/temp
    unzip -o -d $local_path/temp/ $local_path/temporal/conjunto_de_datos/\
    red_nacional_de_caminos_2016.zip
    ogr2ogr -t_srs EPSG:4326 -f CSV /vsistdout/ $local_path/temp/red_vial.shp \
    -lco GEOMETRY=AS_WKT | csvformat -D "|" | awk 'BEGIN{FS="|"}{if (NF<8) {printf "%s",$0} else print}' \
    | sed 's/\"N\/D|\"/N\/D/g' | sed 's/^M//g' > $local_ingest_file
elif [ $year = '2015' ]
then
    mv $local_path/temporal/red_nacional_de_caminos_2015/conjunto_de_datos $local_path/temp
    ogr2ogr -t_srs EPSG:4326 -f CSV /vsistdout/ $local_path/temp/red_vial.shp -lco GEOMETRY=AS_WKT | \
    csvformat -D "|" | sed -i -e 's/^d -1 M\n//g' > $local_ingest_file
elif [ $year = '2017' ]
then
    mv $local_path/temporal/conjunto_de_datos $local_path/temp
    ogr2ogr -t_srs EPSG:4326 -f CSV /vsistdout/ $local_path/temp/red_vial.shp -lco GEOMETRY=AS_WKT | \
    csvformat -D "|" > $local_ingest_file
fi

rm -r $local_path/temp $local_path/temporal $local_path/red_carretera.zip
```

Listing A.2: Archivo de ingesta de Sifode Calificación

```
#!/usr/bin/env bash

###############
# SIFODE
# Ingesta de datos
###############

echo "SIFODE"

# Downloads

aws s3 cp s3://sifode-raw $1/CALIFICACION_39_9.rar #--recursive

unrar p $1/temp/CALIFICACION_39_9.rar | csvformat -d "^" -D "|" | awk 'NR > 8 {
print }' | sed 's/
0\%//;s/\s//;s/\xEF\xBB\xBF///;s/\s\s\s\s//;s/\s[[:digit:]]%//;s/[[:digit:]]%//;s/\s[0-9]//'
| head -n -2 >> $2
```

Listing A.3: Archivo de ingesta de Sifode Domicilio

```
#!/usr/bin/env bash

###############
# SIFODE
# Ingesta de datos
###############

echo "SIFODE DOMICILIO"

# Downloads

aws s3 cp s3://sifode-raw/ $1/temp/ --recursive

unrar -inul p $1/temp/CALIFICACION_39_9.rar | csvformat -d "^" -D "|" >> $2
    sed 's/
0\%//;s/\s//;s/\xEF\xBB\xBF///;s/\s\s\s\s//;s/\s[[:digit:]]%//;s/[[:digit:]]%//;s/\s[0-9]//'
| head -n -2 >> $2

#unrar p $1/temp/CALIFICACION_39_9.rar | csvformat -d "^" -D "|" | awk 'NR > 8
{ print }' | sed 's/  0\%//' | sed 's/\s//' | sed 's/\xEF\xBB\xBF//' | sed//' |
    sed 's/\s\s\s\s//' | sed 's/\s[[:digit:]]%//' | sed 's/[[:digit:]]%//' |
    sed 's/\s[0-9]//' | head -n -2 >> $2
```

Listing A.4: Archivo de ingesta de Sifode

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

import pandas as pd
import csv
import argparse
import smart_open
import tarfile
import boto3
import botocore
import os

def ingest_sifode(data_date='', data_dir='', local_ingest_file=''):
    """
    This function downloads the compressed SIFODE file for each year from S3 as
    well as the headers for all the programs between 2013 and 2017 and builds a
    CSV containing the hole variables ready for concatenation.
    """

    s3 = boto3.resource('s3')

    with smart_open.smart_open('s3://sifode-raw/variables_totales.csv','r') as f:
        reader = csv.reader(f)
        headers = list(reader)

    headers = [item for sublist in headers for item in sublist]

    with smart_open.smart_open('s3://sifode-raw/variables_'+data_date+'.csv', 'r') as f:
        reader = csv.reader(f)
        headers_anio = list(reader)

    headers_anio = [item for sublist in headers_anio for item in sublist]

    s3.Bucket('sifode-raw').download_file('in_'+data_date+'_39_9.tar.gz',data_dir+\
            '/in_'+data_date+'_39_9.tar.gz')

    tar = tarfile.open(data_dir+'/in_'+data_date+'_39_9.tar.gz', "r:gz")
    tar.extractall(path=data_dir)

    df = pd.DataFrame(columns=headers,dtype=object)
```

```
    df2 = pd.read_csv(data_dir+'/in_'+data_date+'_39_9.txt', sep='^',\
            chunksize=1000,dtype=object)

    for chunk in df2:
        chunk.columns=headers_anio
        pd.concat([df,chunk],axis=0).to_csv(local_ingest_file,chunksize=1000,\
                sep='|',mode='a',index=False)

    os.remove(data_dir+'/in_'+data_date+'_39_9.txt')
    os.remove(data_dir+'/in_'+data_date+'_39_9.tar.gz')


if __name__ == '__main__':

    parser = argparse.ArgumentParser(description="Ingest SIFODE data")
    parser.add_argument('--data_date', type=str, default='',
                        help='First year to download, as string format yyyy')
    parser.add_argument('--data_dir', type=str, default='/data/sifode/',
                        help='Local path of ingest data')
    parser.add_argument('--local_ingest_file', type=str, default='',
                        help='Name of output file')

    args = parser.parse_args()
    data_date = args.data_date
    data_dir = args.data_dir
    local_ingest_file = args.local_ingest_file

ingest_sifode(data_date=data_date, data_dir=data_dir, local_ingest_file=local_ingest_file)
```

USO DE DATOS MASIVOS PARA LA EFICIENCIA DEL ESTADO Y LA INTEGRACIÓN REGIONAL

2 de mayo de 2018

FJR-1-ATN/OC 15822-RG

SEDESOL 2018