

Issues, current status and perspectives about DDI in the National Institute of Statistics and Geography of Mexico (INEGI)

European DDI User Conference 2014 (EDDI14)
December, 2014

Silvia.fraustro@inegi.org.mx



Purpose of the presentation

Sharing Mexican experiences about:

- Steps to implement DDI at INEGI
- Where we are at this moment
- Future of DDI in the National Statistical Information System of Mexico (SNIEG)
- DDI for Big Data?

National Institute of Statistics and Geography of Mexico, INEGI

INEGI is a Constitutionally **Autonomous Entity** with:

- **Statistical and geographical** responsibilities
- Around 20,000 employees
- **10 Regional** Offices and **34 State** Offices all around the country
- **5 General Directorates** which generate and integrate statistical data: economic, social, demographic, government, public safety and justice

INEGI is responsible for Mexico's **National System of Statistical and Geographical Information**

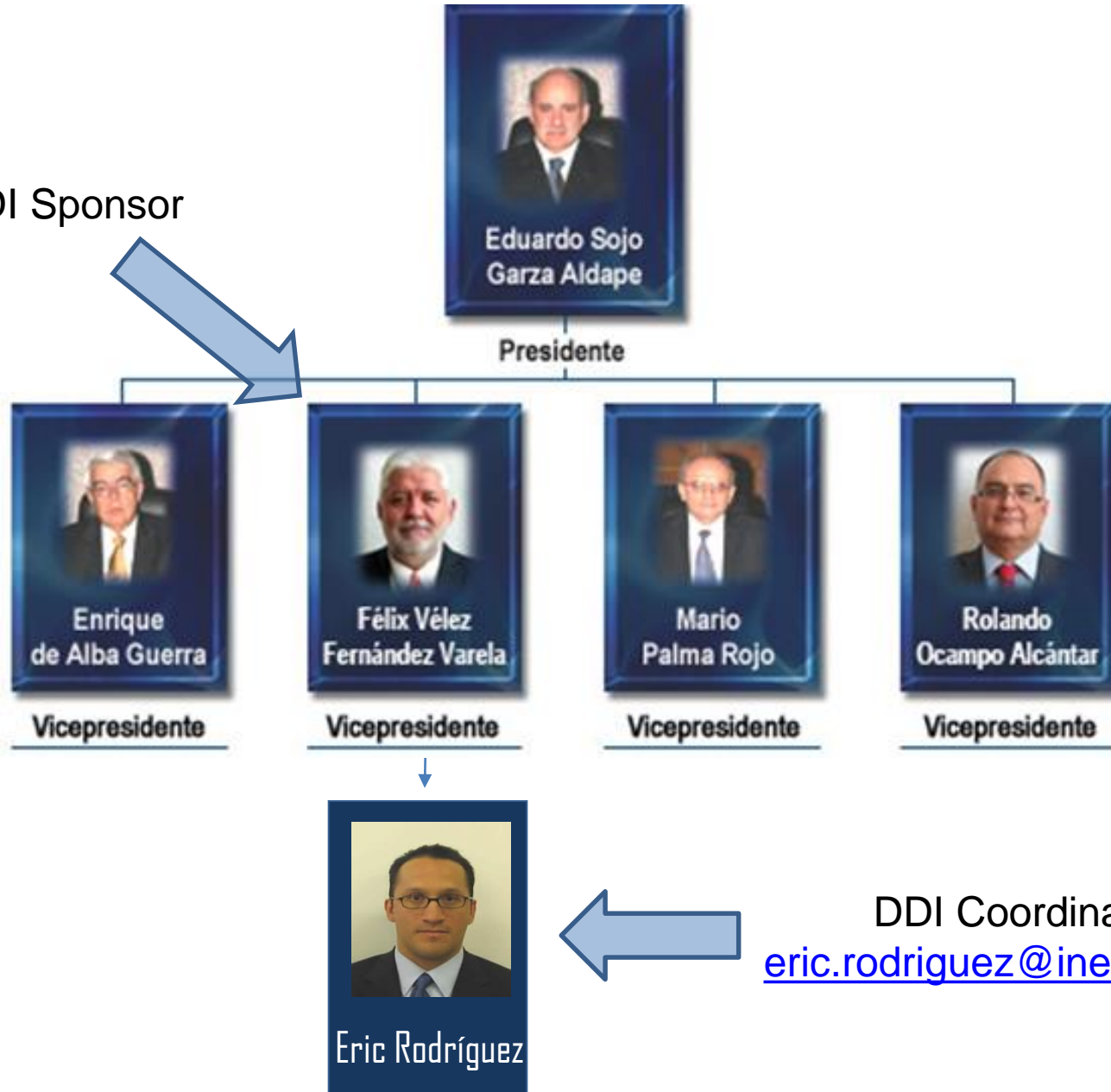


Steps to implement DDI at INEGI



Governing Board Commitment

DDI Sponsor



Steps to implement DDI at INEGI

Definitions

Support Material

Training

2008

2009

2010

2011

2012

2013

2014

2015

...

Projects Documentation

...

Dissemination of individual projects

...

National Data Archive

...



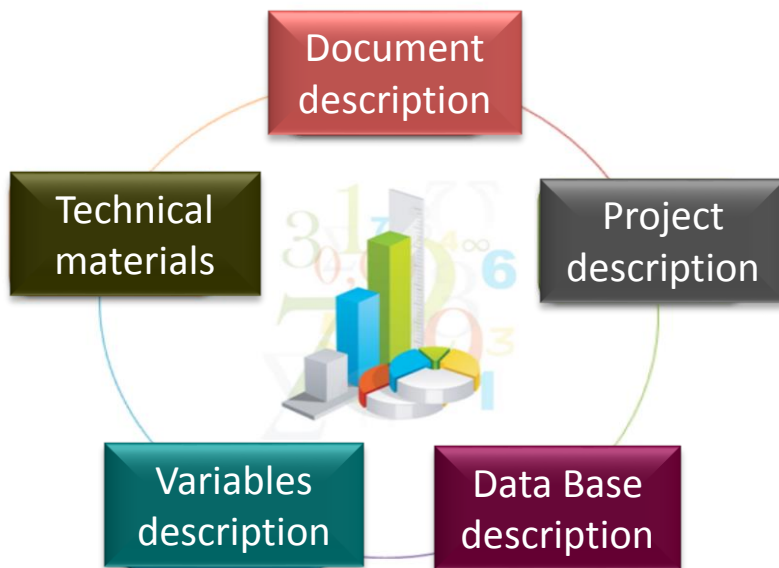
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Definition of Concepts 2008-2010

Getting acquainted with Mexican (local) statistics



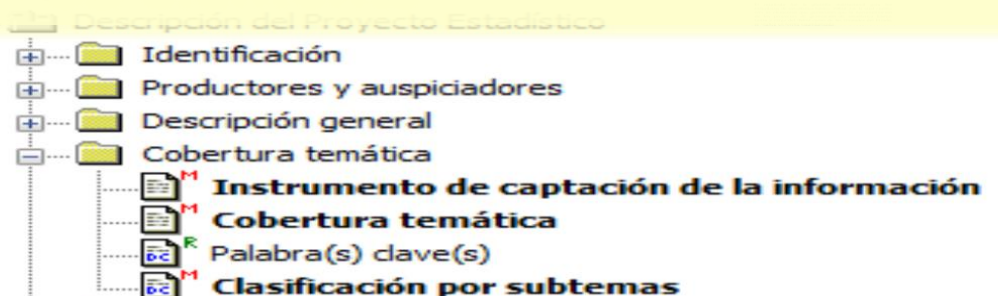
DDI Template Sections



Agreements (Directorates)



Metadata Template INEGI_2.0: 82 elements



Development of Support Material 2008-2011

- Inegi's microdata documentation plan
- Manuals for Microdata documentation
- List of inputs for microdata documentation
- Inegi's template for microdata documentation
- Quality review format

Training our Information People 2009-...

180 People trained



- Four workshops with World Bank
- Two production workshops, supported by ADP, OECD, the World Bank
- Nineteen internal workshops

Project Documentation 2009-...

90 Statistical Projects:

- 37 fully documented by 2012
- 27 more in 2013 and
- 26 in 2014

➤ 2015... Documentation will be a permanent task at INEGI: new censuses, surveys, administrative records and special projects.



Nesstar Editor 4.0.9

Dissemination of Individual Projects 2010-...

www.inegi.org.mx



INEGI INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Inicio | Contacto | INE

Estadística | **Geografía** | Investigación | Productos y servicios | Acerca del INEGI

Temas > Bancos de datos > Fuente / Proyecto > Aspectos normativos y metodológicos >

Censos agropecuarios
Censos de gobierno
Censos económicos
Censos y conteos de población y vivienda
Encuestas agropecuarias
Encuestas en establecimientos
Encuestas en hogares
Índices de precios
Otras encuestas
Paridades de Poder de Compra
PIB y Cuentas Nacionales de México
Registros administrativos

Presentación

Al acceder al apartado Fuente/Proyecto de la pestaña se encuentra información básica generada por los censos nacionales y oficinas de gobierno, las encuestas en hogares y establecimientos, el aprovechamiento de registros administrativos, también, incluye información derivada principalmente del Sistema de Cuentas Nacionales de México y de los índices nacionales de precios.

www.inegi.org.mx/est/contenidos/proyectos/estadistica/default.asp

X



INEGI INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Inicio | Contacto | INE Móvil | Seguirnos

Estadística | **Geografía** | Investigación | Productos y servicios | Acerca del INEGI

Inicio > Estadística > Encuestas en hogares >

Encuestas en Hogares

Encuesta Nacional de Ocupación y Empleo (ENOE)

Productos y servicios

- Consulta en línea
 - Serie con datos ajustados a proyecciones de población 2010
 - Nota sobre los resultados ajustados a las proyecciones de población 2010
 - Nota sobre el uso de nuevas clasificaciones (a partir del tercer trimestre de 2012)
 - Consulta interactiva de datos
 - Consulta interactiva de indicadores estratégicos (InfoLaboral)
 - Indicadores de informalidad laboral - descarga
 - Indicadores de ocupación y empleo - descarga
 - Microdatos - descarga
- Serie con datos originales (2005-2010)
- Series unificadas con criterios de la ENOE

Publicaciones

Aspectos metodológicos

- Presentación
- Catálogos y clasificaciones
- Documentos conceptuales y metodológicos
- Documentos sobre la base de datos
- Glosario
- Instrumentos de captación
- Manuales
- Marco muestral

Presentación

La Encuesta Nacional de Ocupación y Empleo (ENOE) es la censación y fuente de la Encuesta Nacional de Empleo Urbano (ENEU) y la Encuesta Nacional de Empleo (ENE), que por más de 20 años estuvieron proporcionando información de la población ocupada y desocupada. La sustitución se realizó con la finalidad de presentar una nueva encuesta que permitiera captar y conocer de mejor manera las características del mercado laboral mexicano.

A partir del 12 de agosto de 2013, los resultados de la ENOE se presentan en tres bloques: la serie de datos originales (2005 a 2010), generada con factores de expansión ajustados a las proyecciones demográficas oficiales elaboradas por el Consejo Nacional de Población (CONAPO) con base en los resultados definitivos del II Censo de Población y Vivienda 2005; la serie de datos ajustados a las proyecciones demográficas 2010, actualizadas en abril de 2013 por el CONAPO; y las Series unificadas con criterios de la ENOE, que asimilaron la información generada a partir de 2005, con la producida en años previos (ENEU y ENE), con base en el esquema conceptual de la ENOE.

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

INEGI's WebSite National Data Archive 2012-...

www.inegi.org.mx

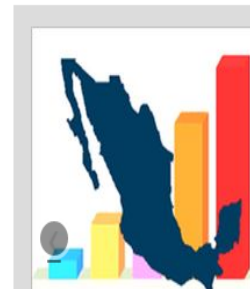


www3.inegi.org.mx/rnm/index.php/home

Red Nacional de Metadatos



[Inicio](#)



Información de Interés Nacional

La LSNEG establece que el Servicio Público de Información Estadística y Geográfica consiste en poner a disposición de los usuarios la totalidad de la Información de Interés Nacional (IIN). Actualmente, los metadatos DDI de 13 de los 17 proyectos de estadística básica del INEGI y que han sido decretados IIN, se pueden consultar a través de este sitio.

Estadísticas

Miércoles 22 de Octubre de 2014
La biblioteca contiene

89 proyectos
24,431 variables

[Catálogo de proyectos](#)

Últimas incorporaciones

[México - Censo Nacional de Impartición de Justicia Estatal 2013](#)
Sep 15, 2014

[México - Censo Nacional de Gobierno, Seguridad Pública y Sistema Penitenciario Estatales 2013](#)
Sep 15, 2014

[México - Censo Nacional de Procuración de Justicia Estatal 2013](#)

Proyectos más consultados

[Módulo de Condiciones Socioeconómicas de la ENIGH 2008](#)
México

[Encuesta Nacional de Ingresos y Gastos de los Hogares 2002](#)
México

Steps to implement DDI at INEGI

Definitions

Support Material

Training

2008

2009

2010

2011

2012

2013

2014

2015

...

Projects Documentation

...

Dissemination of individual projects

...

National Data Archive

...



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

National Data Archive Capabilities

Red Nacional de Metadatos

[Inicio](#) > [Catálogo de proyectos](#)

Buscar por palabra clave

en descripción de la operación estadística

en documentación de la variable

ingreso

[Buscar](#) [Restablecer](#)

Catálogo de proyectos

Descripción

Lista de proyectos

Encontró 335 variables de 24941

ingreso x

Project and
Variable-level search

Variable-level
selection

<input type="checkbox"/>	CONACT	Condición de actividad México - Censo de Población y Vivienda 2010
<input type="checkbox"/>	RAN_INGR	Rango Ingreso México - Censo Nacional de Gobierno, Seguridad Pública y Sistema Penitenciario Estatales 2012
<input type="checkbox"/>	F311B2	LÍNEAS DE MERCANCÍAS COMERCIALIZADAS, PORCENTAJE 2 MÉXICO - Censos Económicos 2009
<input type="checkbox"/>	F411B3	TIPOS DE SERVICIOS PROPORCIONADOS, PORCENTAJE 3 MÉXICO - Censos Económicos 2009
<input type="checkbox"/>	M500A	INGRESO POR LA RENTA DE LOCALES, OFICINAS, MAQUINARIA Y MOBILIARIO MÉXICO - Censos Económicos 2009
<input type="checkbox"/>	F311A3	LÍNEAS DE MERCANCÍAS COMERCIALIZADAS, DESCRIPCIÓN 3 MÉXICO - Censos Económicos 2009

Comparar Variables [Refrescar](#) | [Limpiar](#) | [Descargar el archivo PDF](#)

CONACT	RAN_INGR	INGTRMEN
México - Censo de Población y Vivienda 2010	México - Censo Nacional de Gobierno, Seguridad Pública y Sistema Penitenciario Estatales 2012	México - Censo de Población y Vivienda 2010
Condición de actividad (CONACT) Archivo: Poblacion_Ampliado	Rango Ingreso (RAN_INGR) Archivo: JC_RECHU	Ingresos por trabajo mensualizados (INGTRMEN) Archivo: Poblacion_Ampliado
Información general Tipo: Discreta Casos válidos: 0 (0) Formato: characterInvalídos: 0 (0) Ancho: 2	Información general Tipo: Discreta Casos válidos: 52 Formato: numericInvalídos: 106 Ancho: 1 Mínimo: 1 Decimales: 0 Máximo: 7	Información general Tipo: Discreta Casos válidos: 0 (0) Formato: characterInvalídos: 0 (0) Ancho: 6
DEFINICIÓN Situación que distingue a la población de 12 y más años de edad, de acuerdo a si en la semana de referencia participó o no en la actividad económica, o si buscó vincularse a alguna. Se clasifica en: Población económicamente activa y Población no económicamente activa.	DEFINICIÓN Clasificación del personal destinado a la función de justicia cívica en el Distrito Federal de acuerdo con su rango de ingresos, al cierre del año 2011.	DEFINICIÓN Percepción monetaria mensual que la población ocupada obtiene o recibe del (los) trabajo(s) que desempeña en la semana de referencia. Se consideran los ingresos por concepto de ganancia, comisión, sueldo, salario, jornal, propina o cualquier otro devengado de su participación en alguna actividad económica.
UNIVERSO	UNIVERSO	

Variable analysis

Future of DDI in the National Statistical Information System of Mexico (SNIEG)

DDI in the National System of Statistical and Geographical Information of MX (SNIEG)

- INEGI has the legal mandate to coordinate the National System of Statistical and Geographical Information.
- Mexico is a federal country with 3 branches of government: Executive, Legislative and Judicial; **all of them are statistical data generators**, working as a system.
- So we must adopt best practices for documentation, data sharing and dissemination (DDI & SDMX, for example)

April 16, 2008

Yet to be completed:

Law for the National System of Statistical and Geographical Information

LEY DEL SISTEMA NACIONAL DE INFORMACIÓN ESTADÍSTICA Y GEOGRÁFICA

La Ley del SNIEG establece:
Art. 55.- El Instituto, en su calidad de unidad central
coordinadora del Sistema, tendrá las funciones siguientes:
I. Normar y coordinar el Sistema...
II. Normar y coordinar las Actividades (estadísticas y geográficas
que lleven a cabo las Unidades, tomando en cuenta los
estándares nacionales e internacionales...

Art. 88.- ...
... el Instituto deberá dar a conocer y conservar los metadatos o
especificaciones concretas de la aplicación de las metodologías
que se hubieren utilizado en la elaboración de la Información.
...



DIARIO OFICIAL DE LA FEDERACION

“Official Gazette”

Rules for developing
metadata of “basic statistical
projects” under the DDI
standard at SNIEG



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Technical specifications for
the application of the Rules

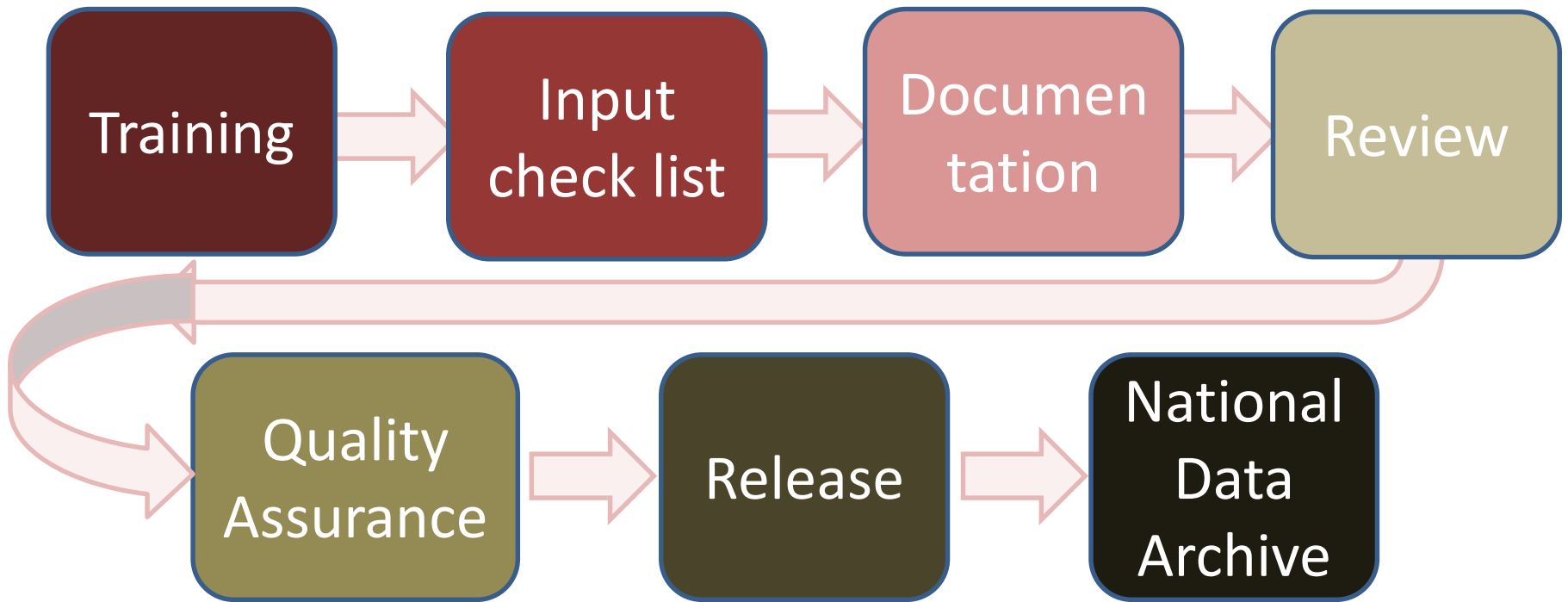
What the Draft Rules say

- INEGI has to define the metadata template, the process, formats and tools.
- **Mandatory for all “statistical data generators” to document their projects in DDI (data providers)**
- INEGI has to provide training
- “DDI Metadata Editor” (developed by World Bank)
- Dissemination through the National Data Archive
- **Providers of the System must enforce these Rules within 3 years**

Content of Technical specifications

- Definitions, the process, guidelines and manuals, which help to execute the norm
- The content of this specifications can be modified according to other DDI versions

Process for developing metadata under the DDI standard at SNIEG



Stages for the release of the Rule

- Review by the Governing Board
- Review by INEGI's Directorates
- Review by the 4 External Committees
- Public Consultation in the WebSite of SNIEG
- Approval by the Governing Board
- Publication in the Official Gazette
- Promotion of the Rule



Now

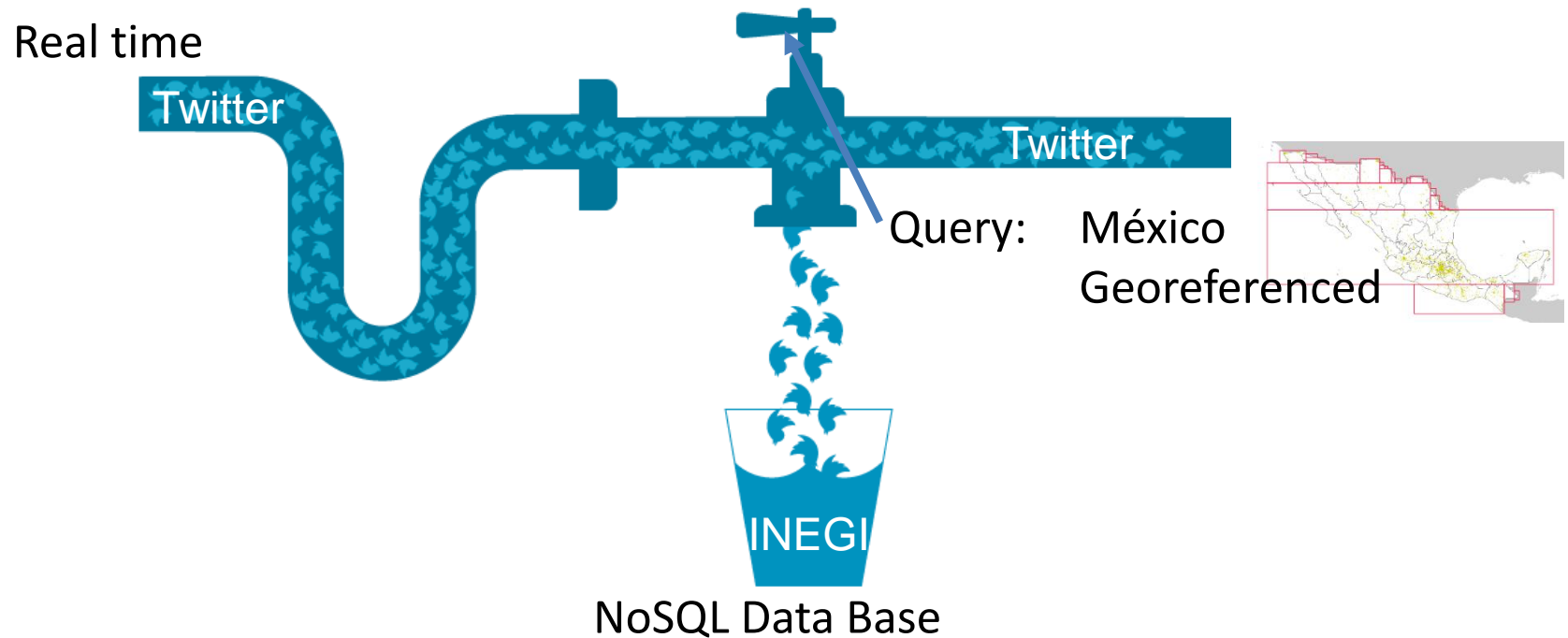
Nov 24, 2014

To do ...
First Quarter 2015

DDI for Big Data?

Pilot Test: Use of social media data
for official statistics

Twitter as a data source

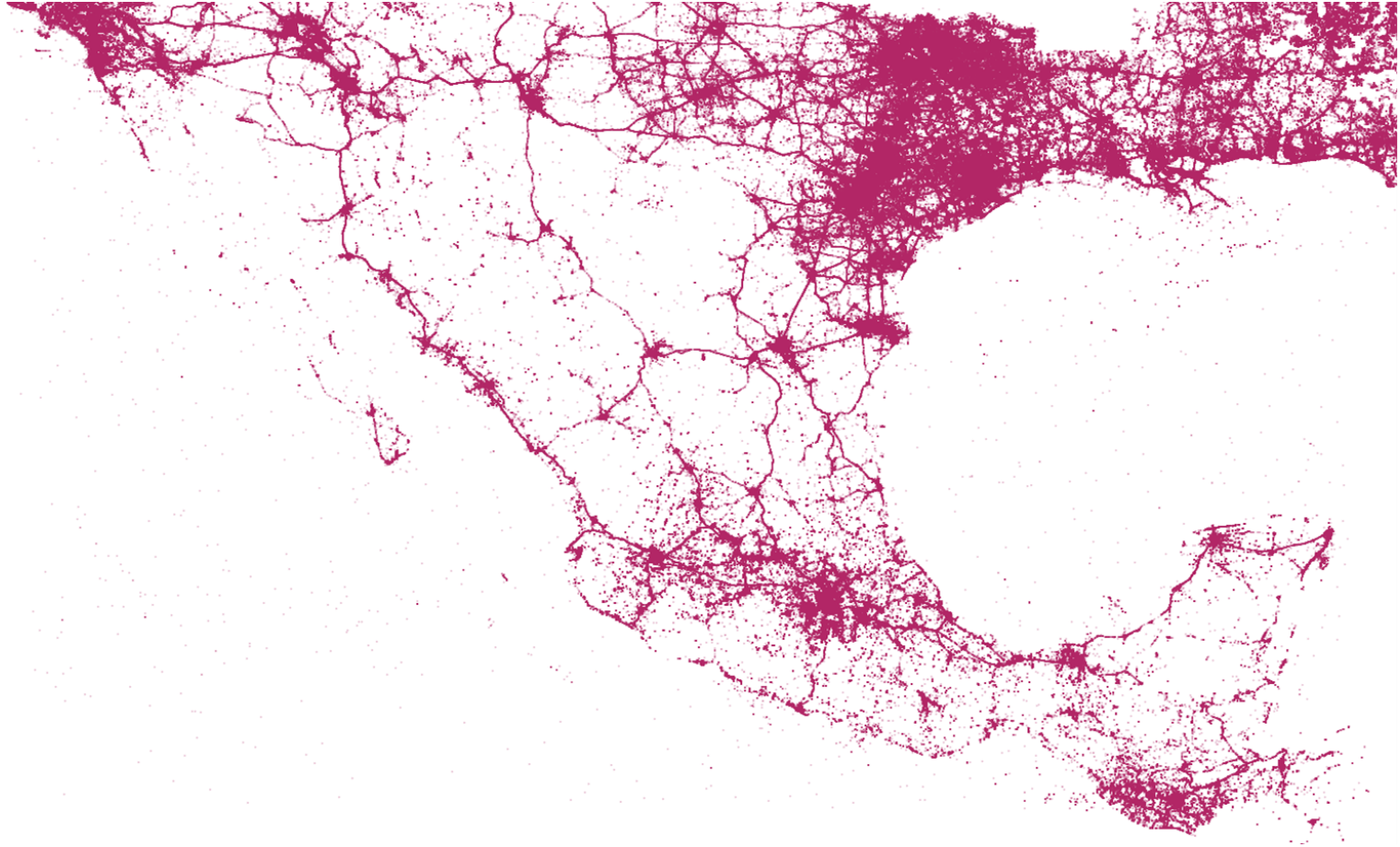


Why Twitter?



- Readily available
- Up to 1% of global tweets at no cost
- Around 12 M accounts in Mexico
- Geolocated tweets by 700 thousand accounts
- 100 M plus tweets downloaded since January 2014
- Even though its drawbacks: Not documented, not supported by “traditional” statistics methodologies

How do tweets look like...?

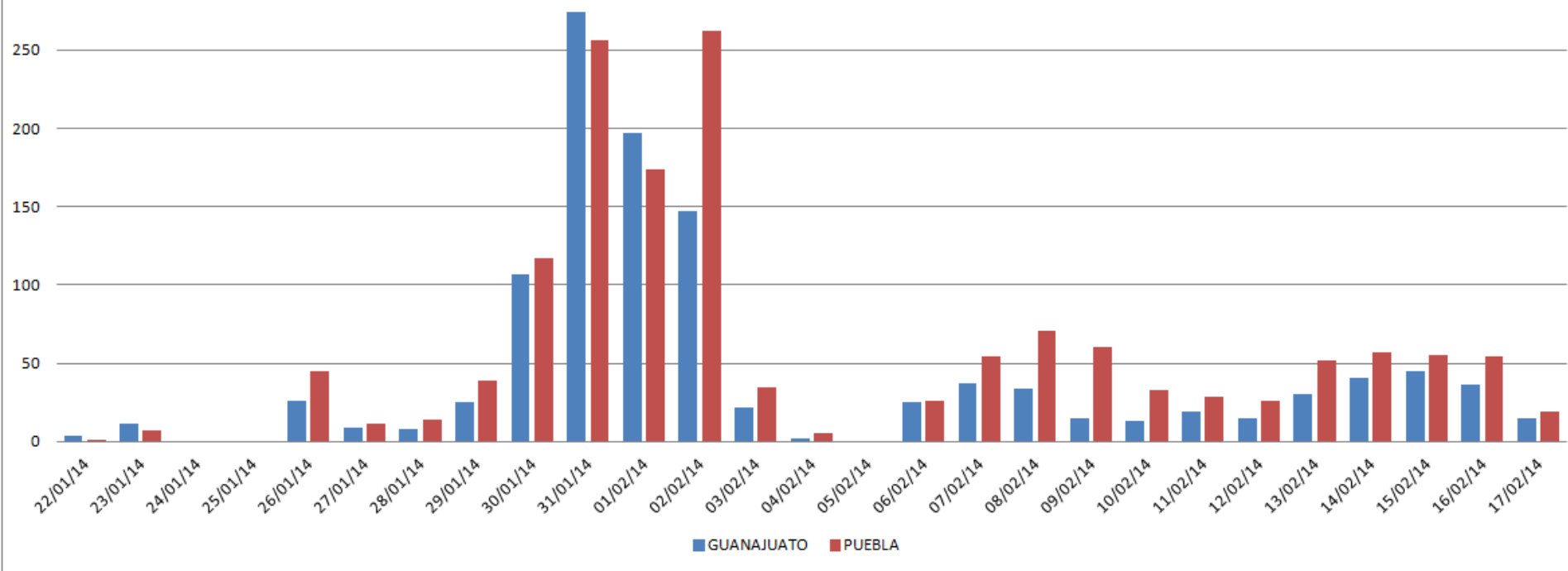


Topics of Study in the Pilot Test

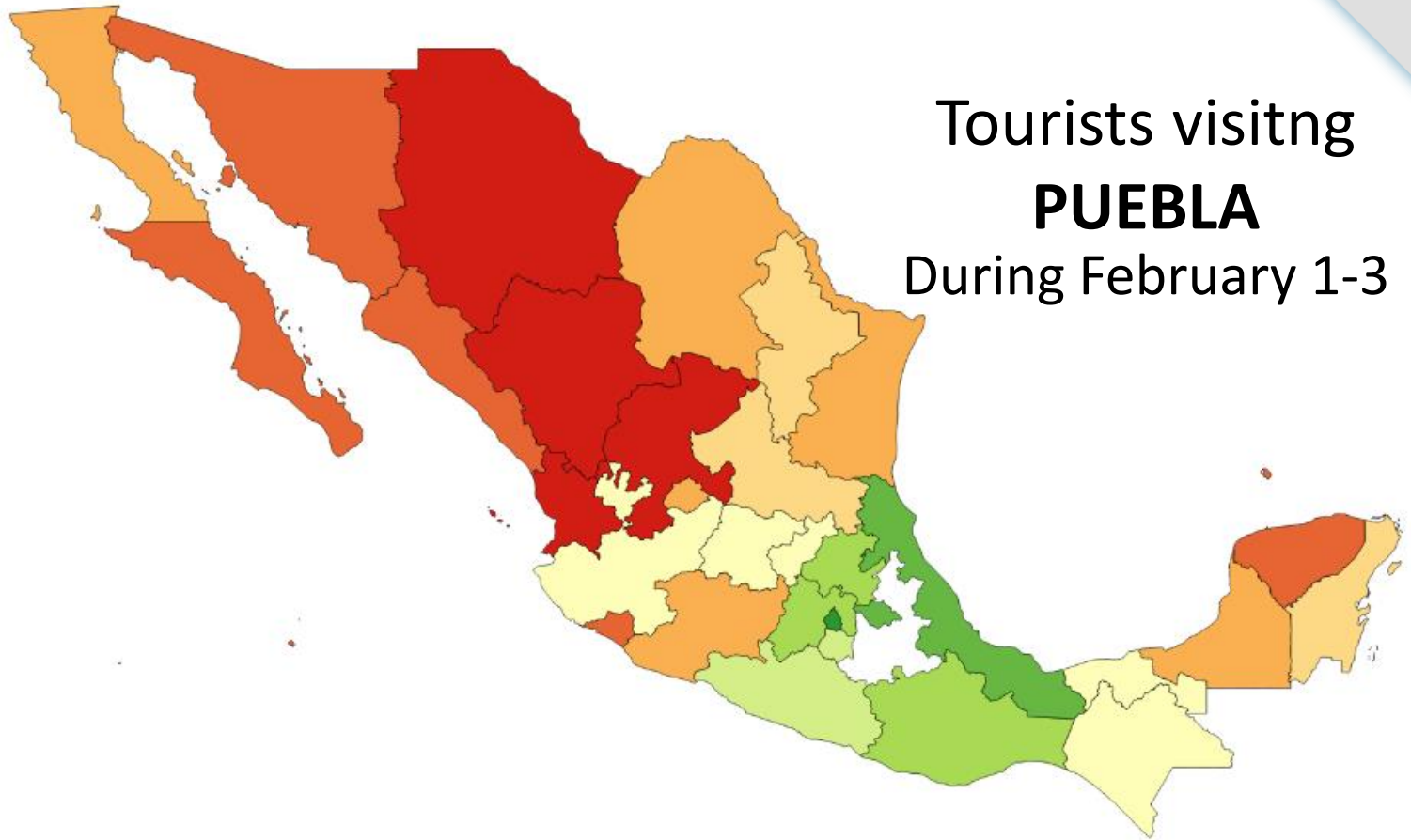
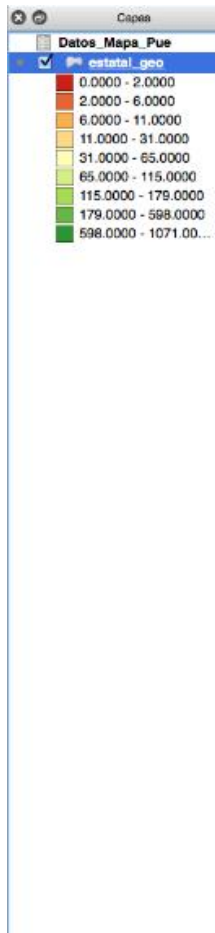
- **Twitter- Tourist Exercise**
- Mobility studies
- Twitter for Subjective Well-being

Twitter- Tourist Exercise

Tweets from Guanajuato or Puebla before, during and after a long weekend

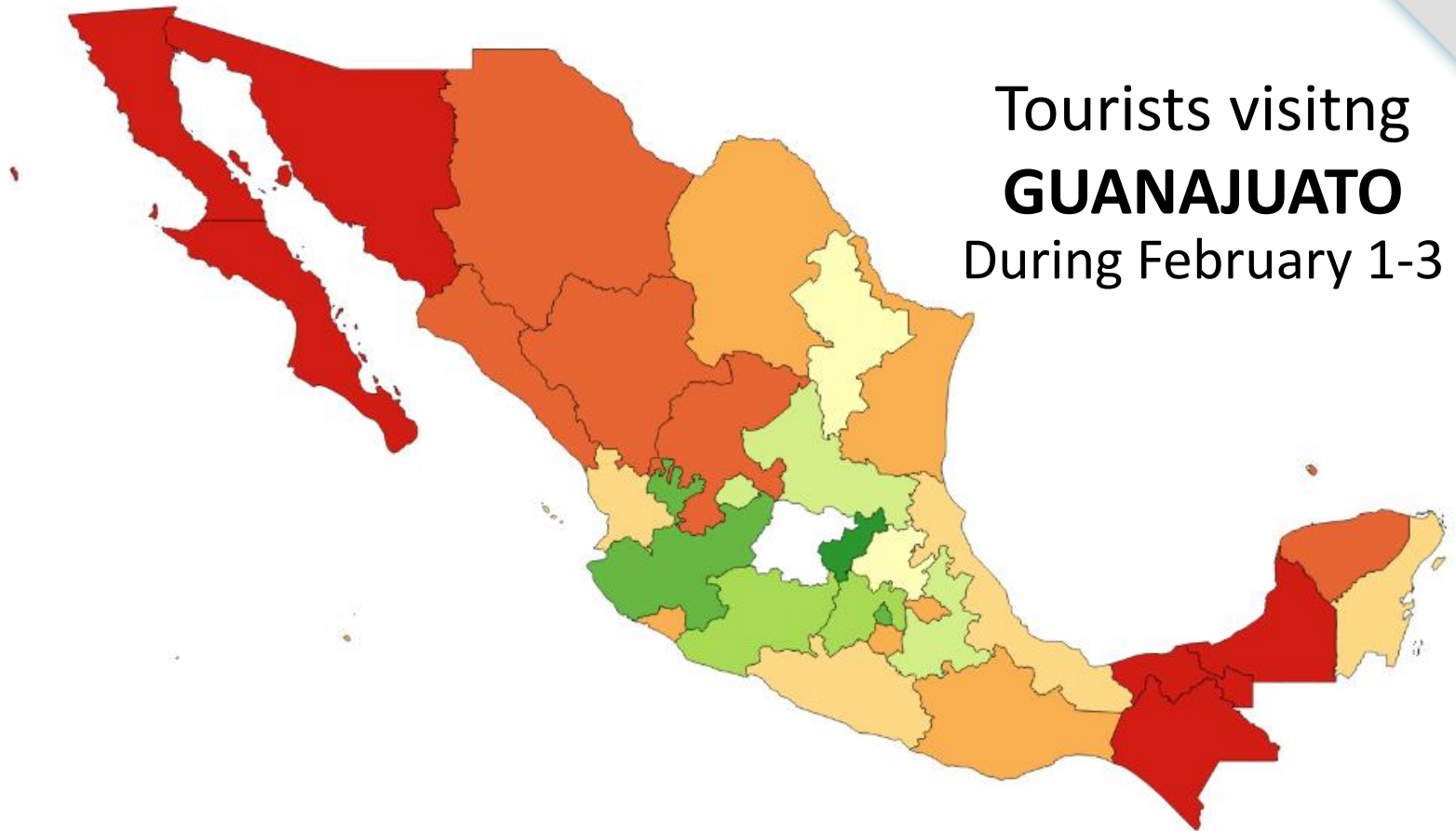
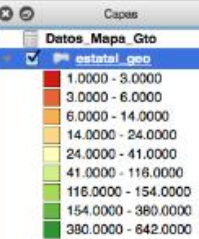


Twitter- Tourist Exercise



Find out **which state people twitted from and how long they stayed in that state.**

Twitter- Tourist Exercise



Find out **which state people twitted from and how long** they stayed in that state.

How to include metadata on the results obtained from sources of Big Data?

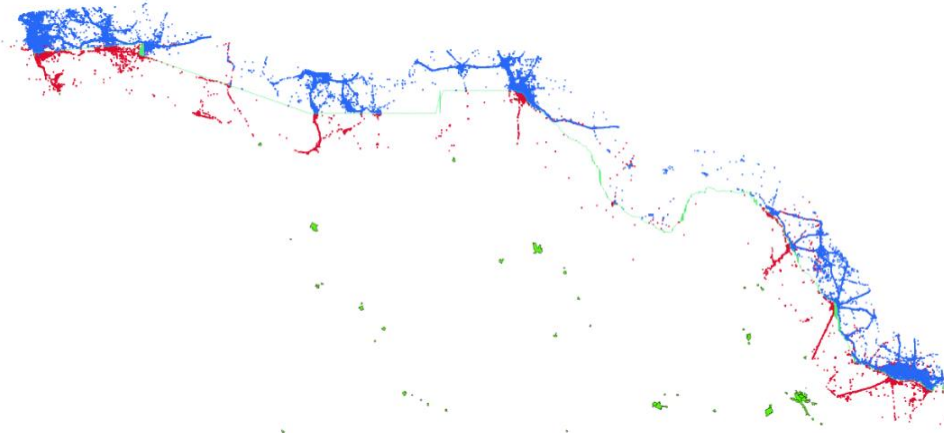
- Is Big Data a new Source for Official Statistics?
- If so.. It would be interesting to investigate if the capabilities of DDI can allow us the documentation of this new source
- The same happens with the resulting Statistics: How can we assure that they are properly understood, at least according to what the researcher pretended to express.
- We need to document the Project including metadata related to definitions, methodologies, places, origin of the sources, etc..

Topics of Study in the Pilot Test

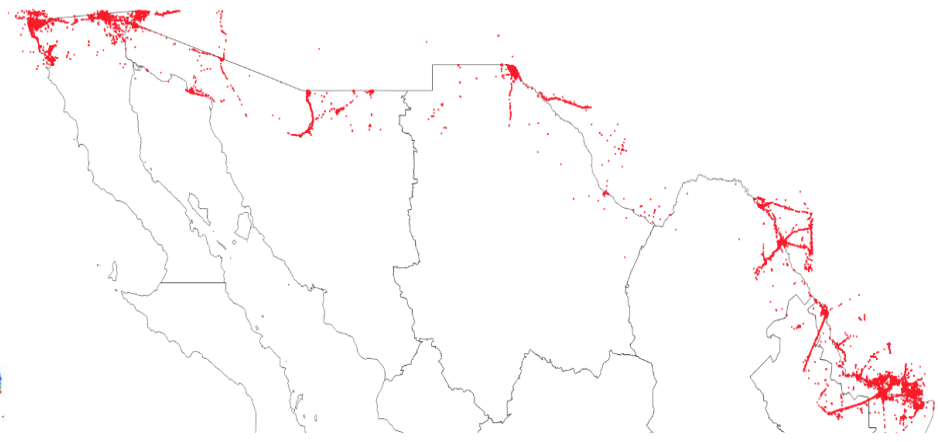
- Twitter- Tourist Exercise
- **Mobility studies**
- Twitter for Subjective Well-being

Mobility studies

Research for development of analytical method to measure trans-border mobility through Geo-referenced tweets.



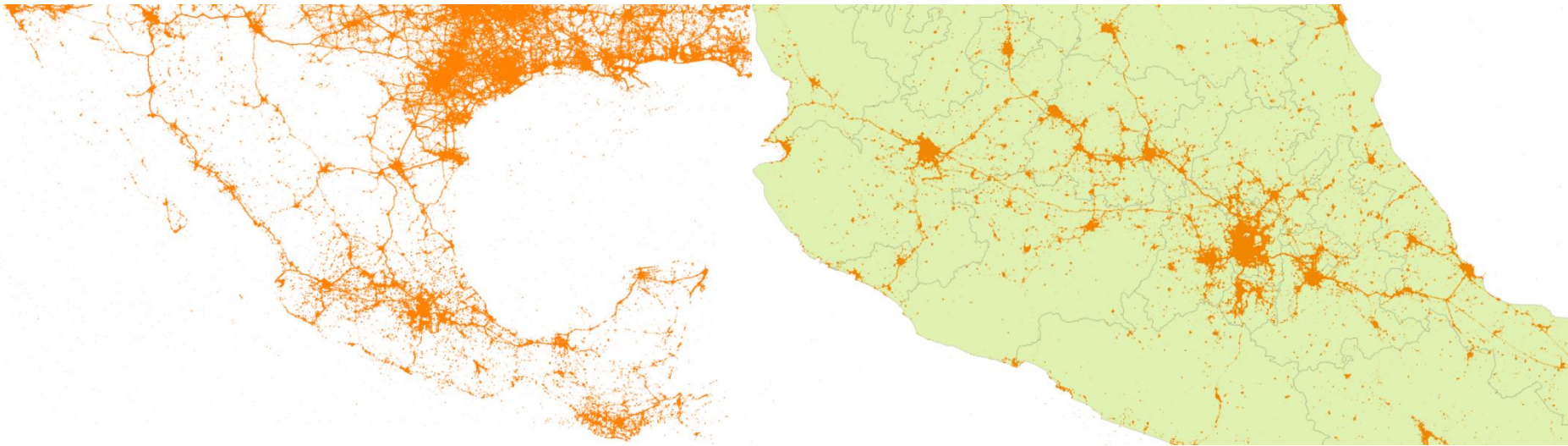
Mexican (red) and US(blue) tweets



Mexican tweets

Mobility studies

Using the National Road Network to find out whether domestic mobility analysis can be conducted, current work. (Plot of 70 M tweets)



Topics of Study in the Pilot Test

- Twitter- Tourist Exercise
- Mobility studies
- **Twitter for Subjective Well-being**

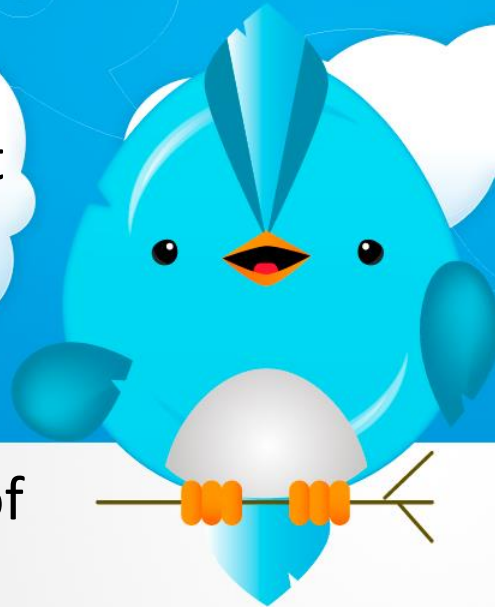
Twitter for Subjective Well-being



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

<http://cienciadedatos.inegi.org.mx/pioanalysis>

App to create
our training set
of tweets



Humans grade a sample of tweets as positive, neutral or negative, and classify them in several subjects. **Bienestar Subjetivo y las redes sociales como fuente de datos**

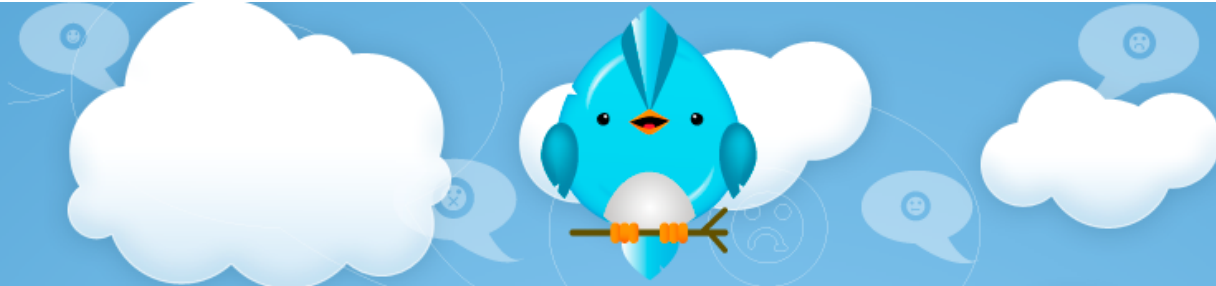
PIOANALISIS

“Tec Milenio University” supports the exercise with about 3000 students who are classifying the training set

Supervised learning method

Example of PIOANALISIS

How do you think that the person who tweeted was feeling in that moment?



4 de 20 - nivel 1

No sean PENDEJOS a los eclipses no se le piden deseos!

¿El tema del tuit es?

Política

Cultural / Entretenimiento

Deporte

Escolar / Laboral

Personal

Ni idea

¿El tuitero se sentía?



Positivo

DDI to document the way how the training set was obtained

- Could it be seen as a non probabilistic survey?
- The way humans graded the sample of tweets to get the training set for the machine learning method could also be enriched if we documented it with DDI.

Twitter for official statistics

Expected results?

If the pilot results are consistent with those of our traditional statistics (like “Self-reported Welfare Survey (BIARE)”), then:

- It would be possible to produce low-cost frequent measurements between traditional projects waves, and to publish them as experimental statistics.
- Document each project. Would it be possible with DDI?

WE DON'T KNOW YET



Conociendo México

01 800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

Thank you



@inegi_informa



INEGI Informa



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

