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

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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?

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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

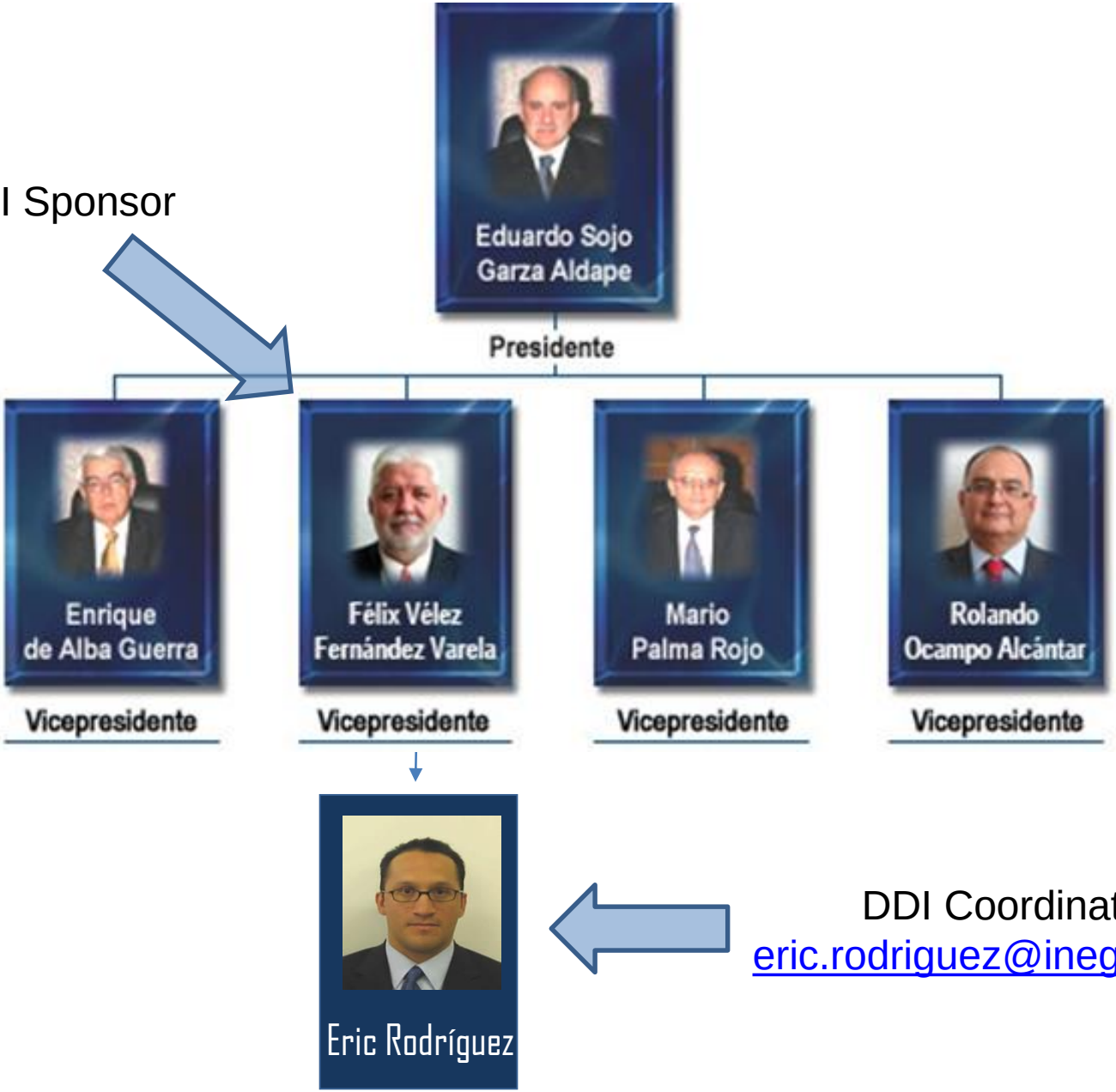
INEGI is responsible for Mexico's **National System of Statistical and Geographical Information**

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

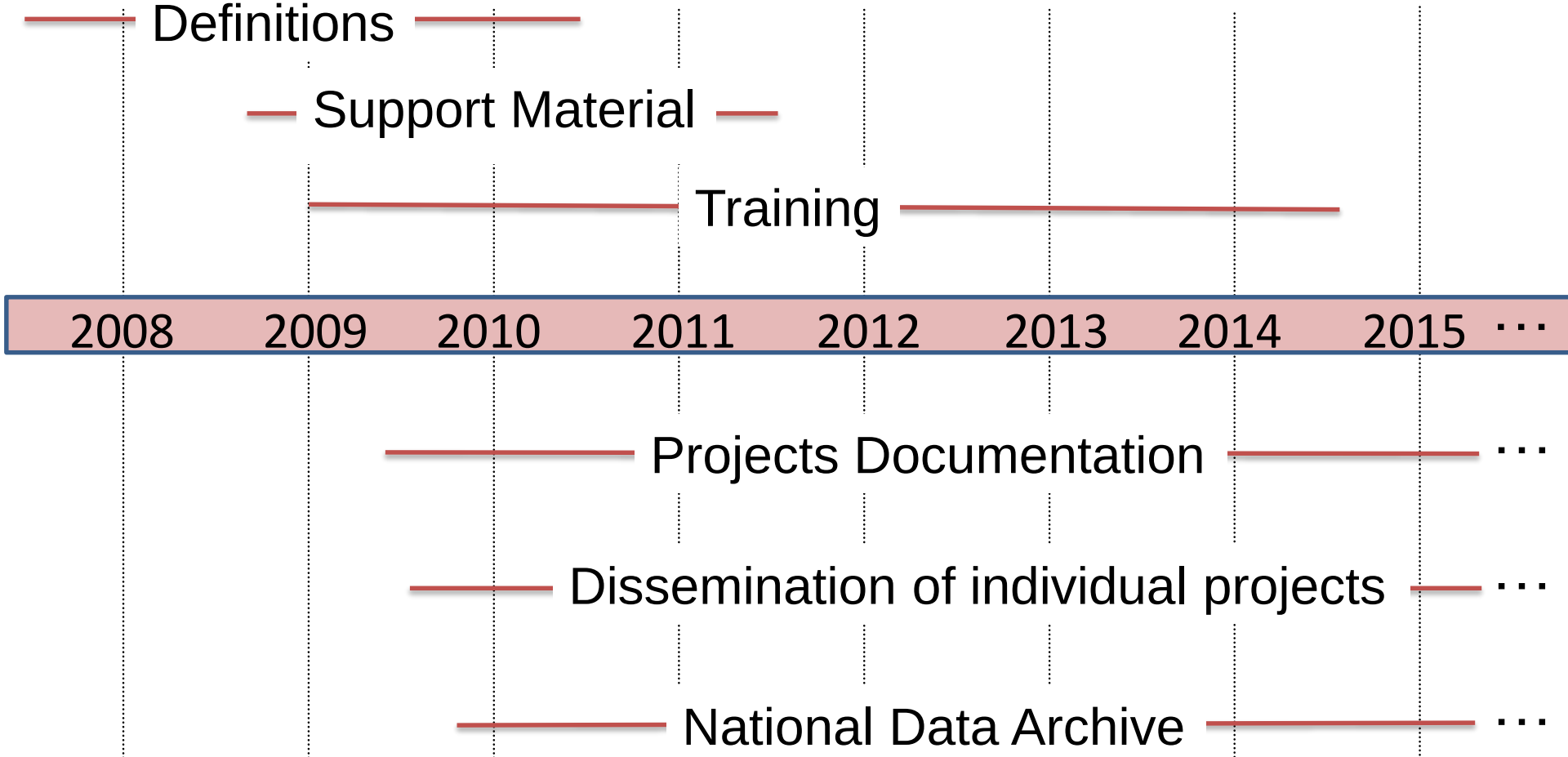
Steps to implement DDI at INEGI

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Governing Board Commitment



Steps to implement DDI at INEGI



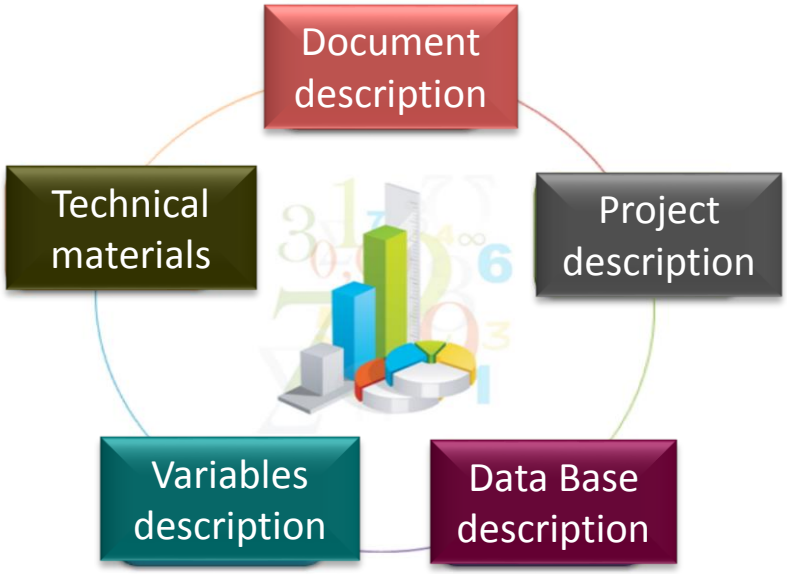
INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Definition of Concepts 2008-2010

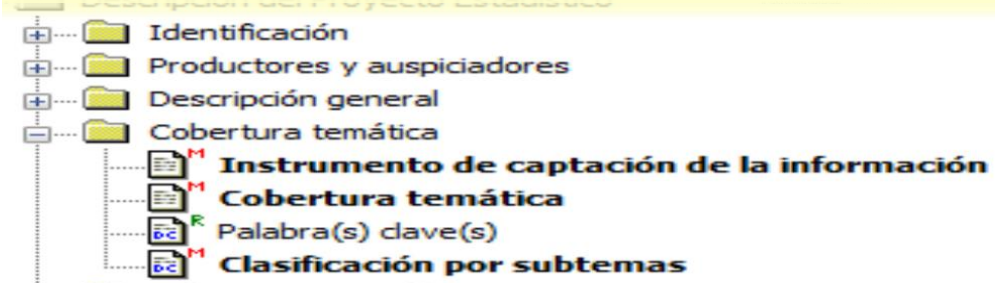
Getting acquainted with Mexican (local) statistics

Agreements (Directorates)

DDI Template Sections



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

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Project Documentation 2009-...

90 Statistical Projects:

Nesstar Editor 4.0.9

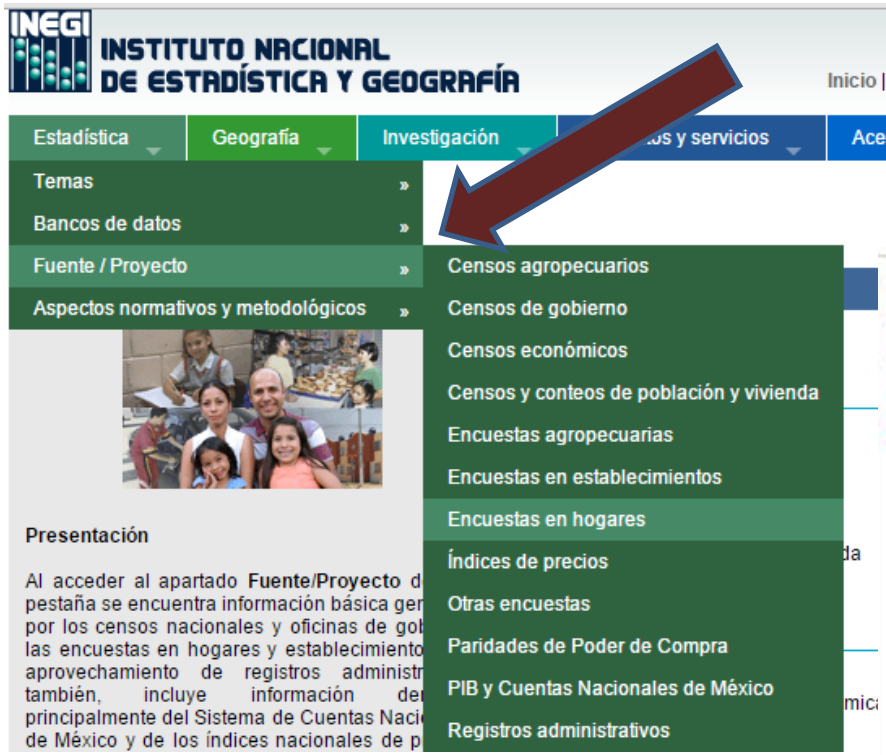
- 37 fully documented by 2012
- 27 more in 2013 and
- 26 in 2014

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

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Dissemination of Individual Projects 2010-...

www.inegi.org.mx



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

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

www.inegi.org.mx

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

INEGI INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Inicio | Contac

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

Temas
Bancos de datos
Fuente / Proyecto
Aspectos normativos y metodológicos

Principios Fundamentales de las Estadís
Clasificaciones y catálogos
Cuestionarios
Documentos técnicos y normativos
Glosarios
Metodologías
Programas y proyectos estadísticos
Red Nacional de Metadatos

De interés

- Anuario estadístico y geográfico de los Estados Unidos Mexicanos
- Directorio Estadístico Nacional de Unidades Económicas (DENUE)
- Índices de

Inicio > Estadística >

Red Nacional de Metadatos

Inicio

Información de Interés Nacional

La LSNIEG 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

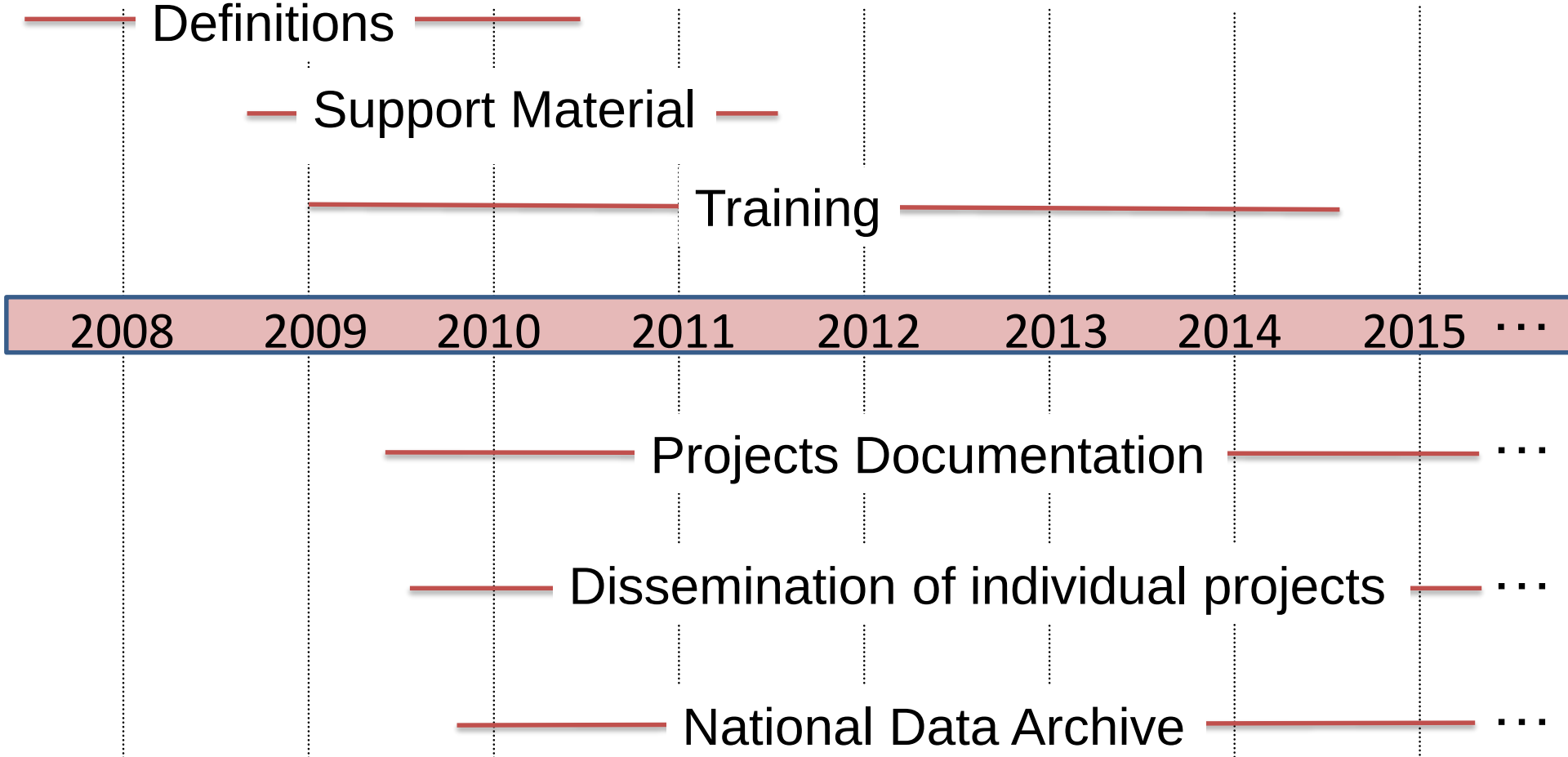
México - Censo Nacional de Procuración de Justicia Estatal 2013

Proyectos más consultados

Módulo de Condiciones Socioeconómicas de la ENIGH 2008
México

Encuesta Nacional de Ingresos y Gastos de los Hogares 2002

Steps to implement DDI at INEGI



INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

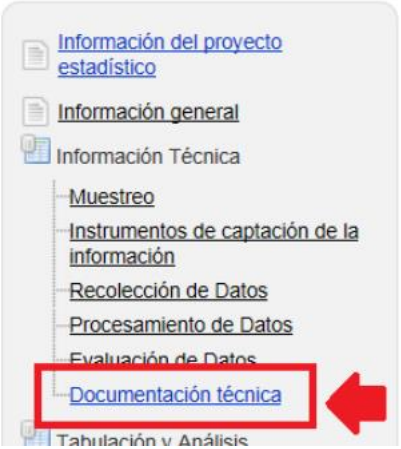
National Data Archive Capabilities

Project and Variable-level search

Variable-level selection

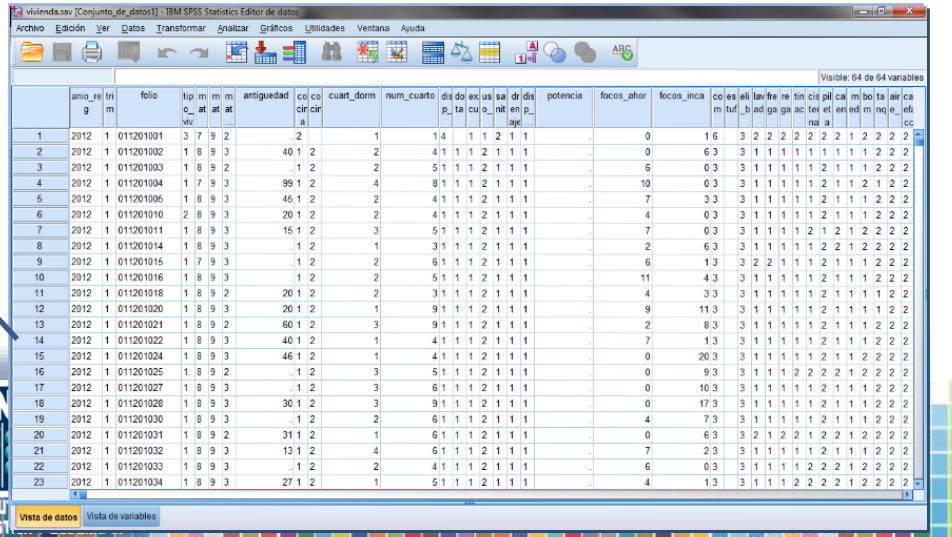
Variable analysis

Conceptual and Technical access



Microdata download

Database Views



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

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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)

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

April 16, 2008

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. Nomar 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.
...

SNIEG INEGI

Yet to be completed:

ESTADOS UNIDOS MEXICANOS

DIARIO OFICIAL DE LA FEDERACION

"Official Gazette"

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

INEGI

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Technical specifications for the application of the Rules

INEGI INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍ

What the Draft Rules say

DIARIO OFICIAL DE LA FEDERACIÓN

Rules

- 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**

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

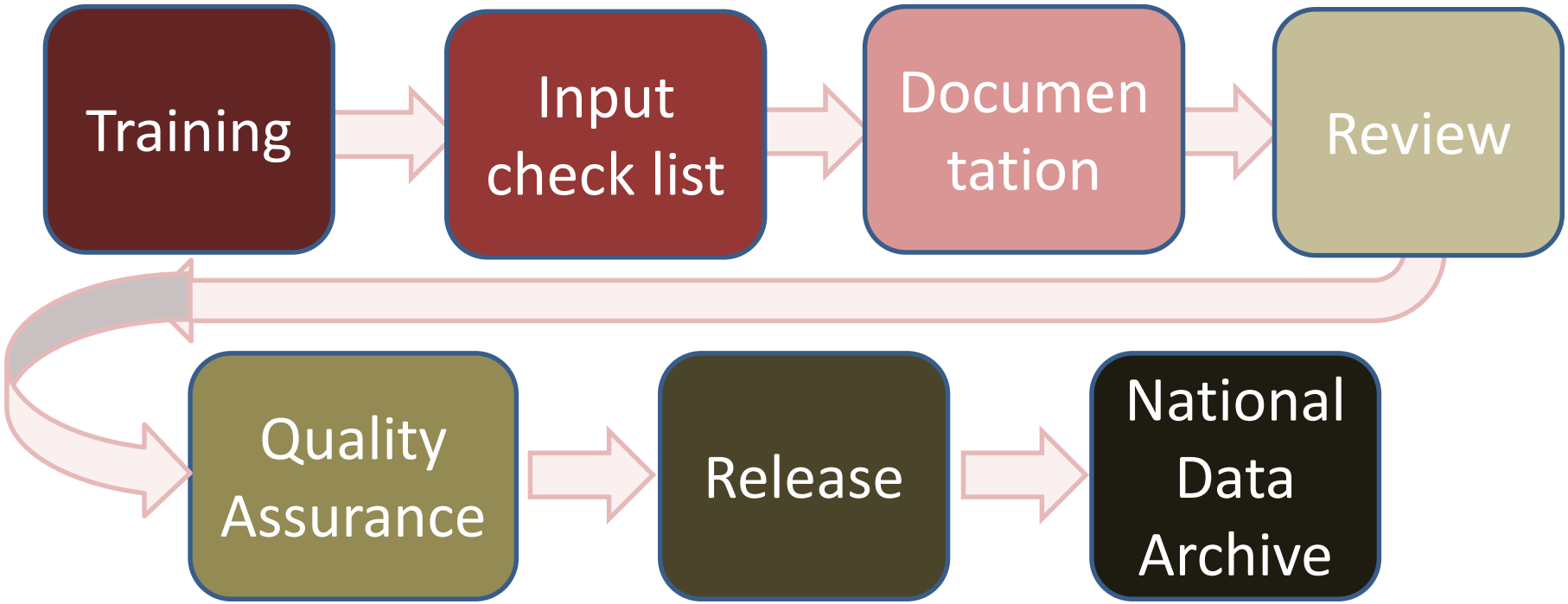
Content of Technical specifications

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



INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

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 …

Fist Quarter 2015

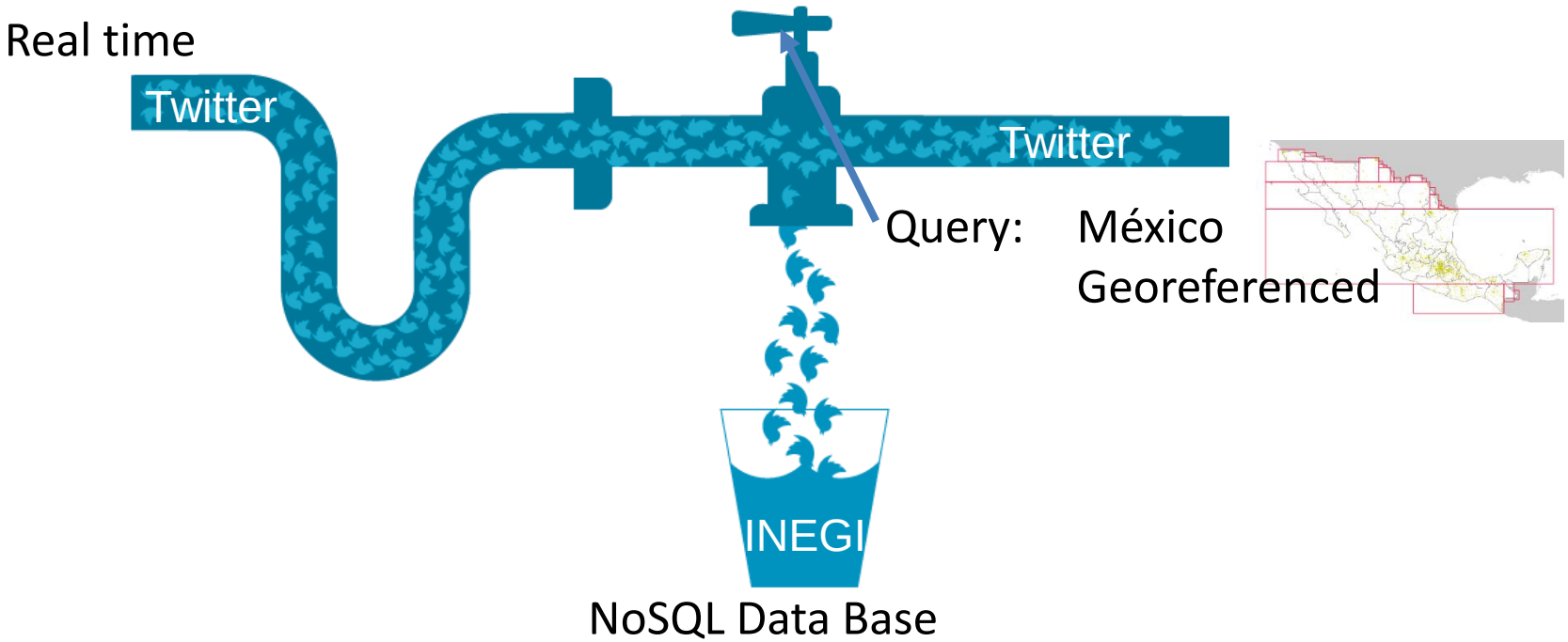
INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

DDI for Big Data?

Pilot Test: Use of social media data for official statistics

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Twitter as a data source

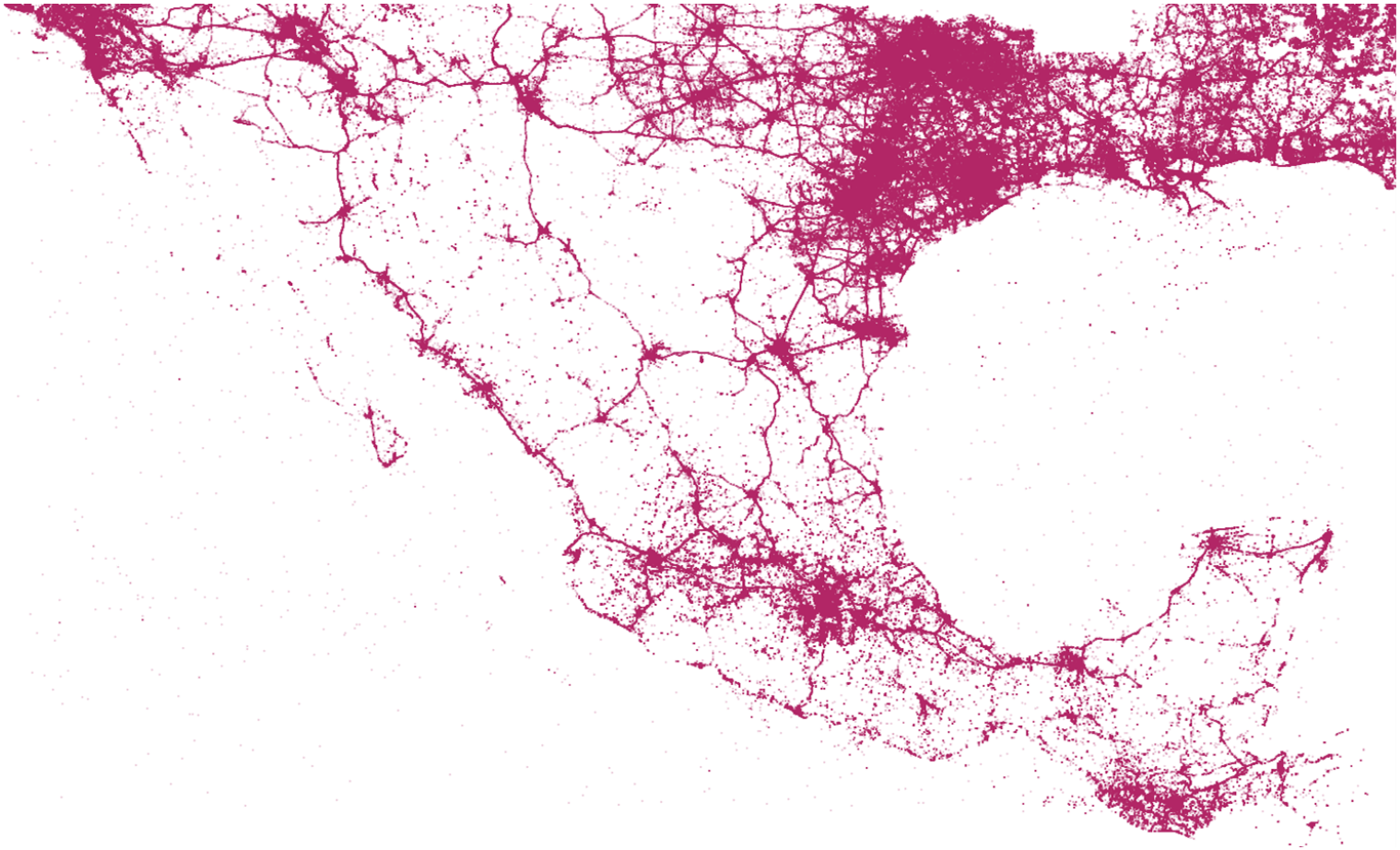


Why Twitter?

- 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

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

How do tweets look like…?



INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Topics of Study in the Pilot Test

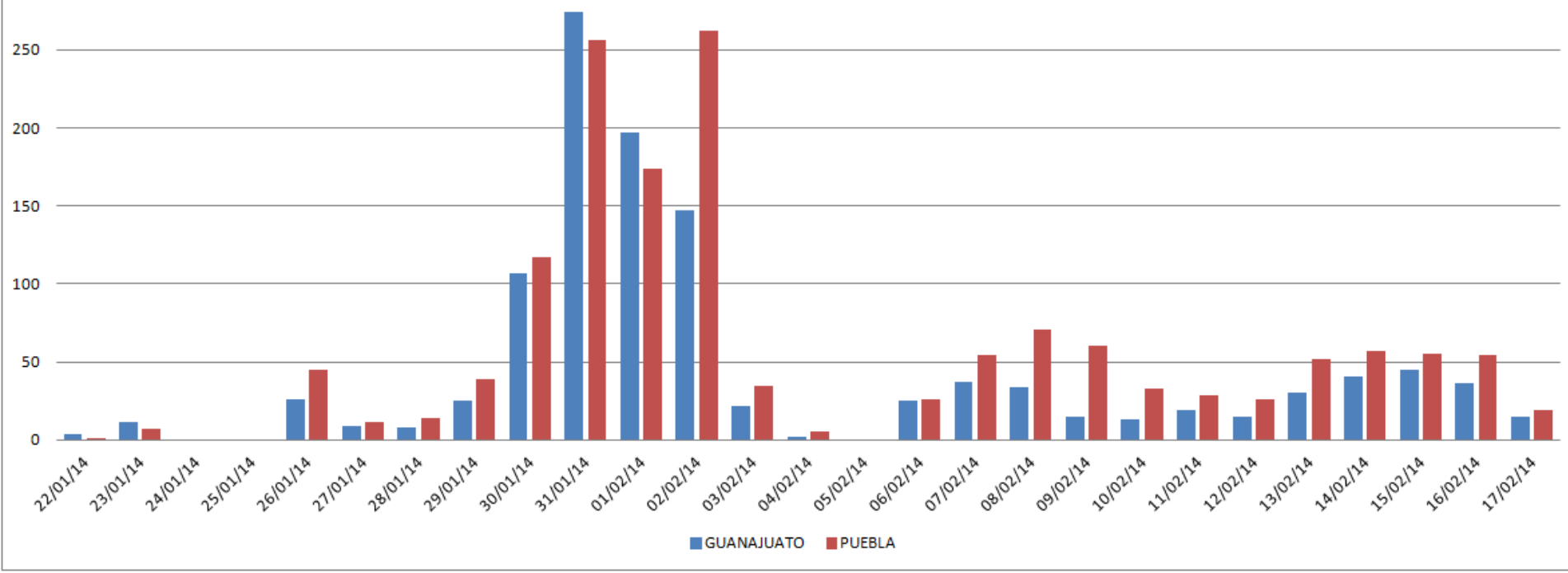
- **Twitter- Tourist Exercise**
- Mobility studies
- Twitter for Subjective Well-being

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Twitter- Tourist Exercise

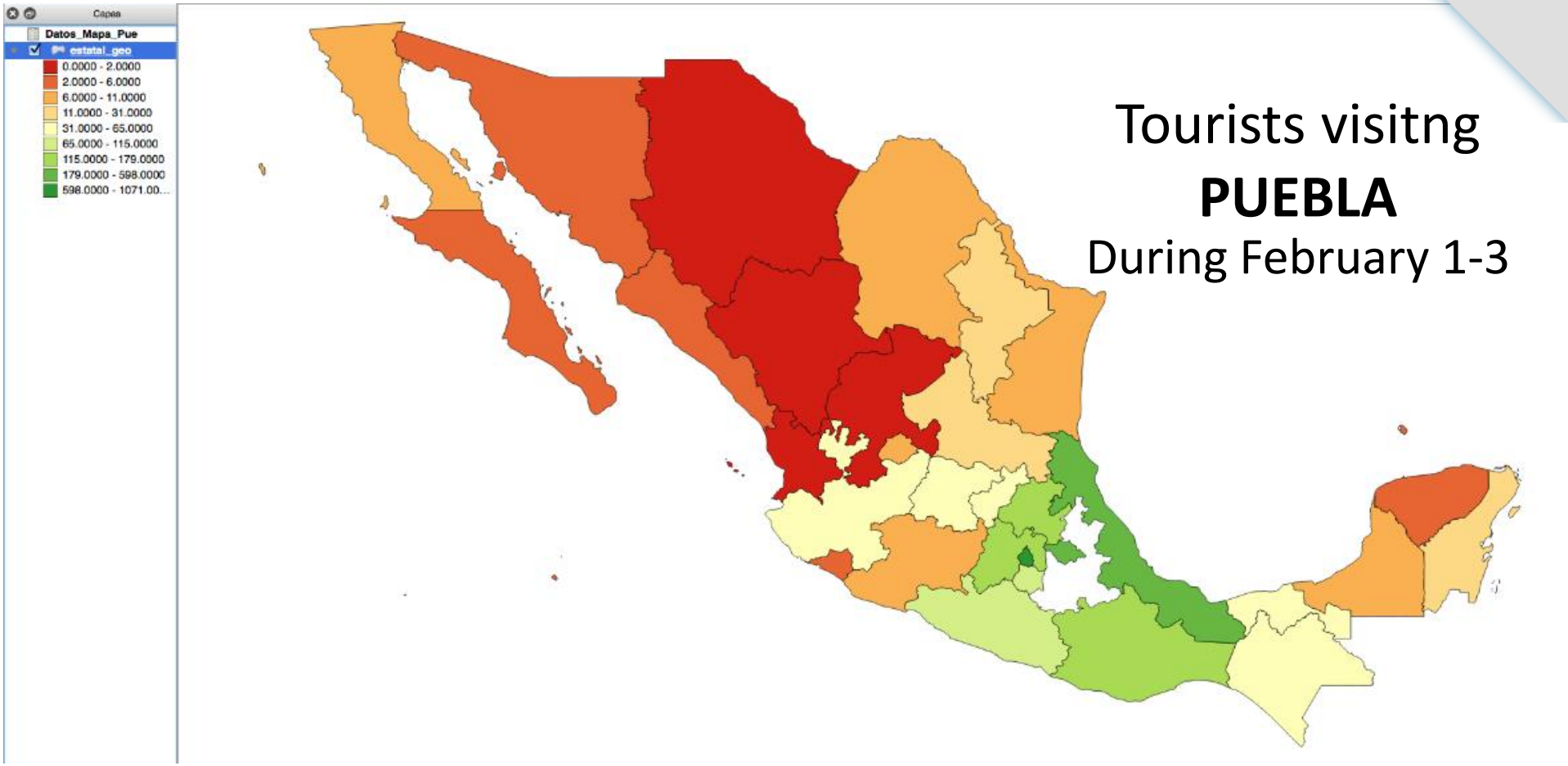
SECTUR
SECRETARÍA DE TURISMO

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



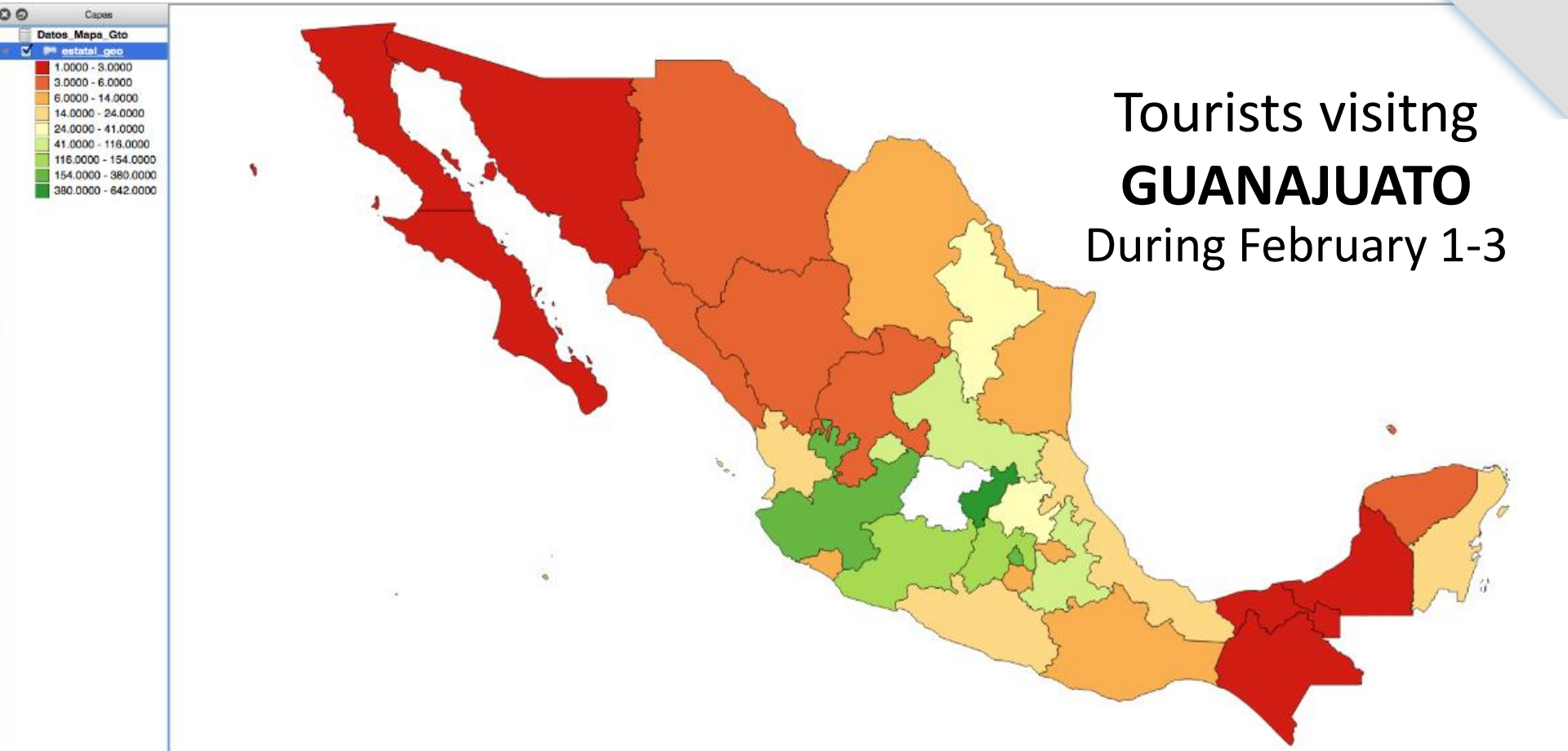
INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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 capabilites of DDI can aloud 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..

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

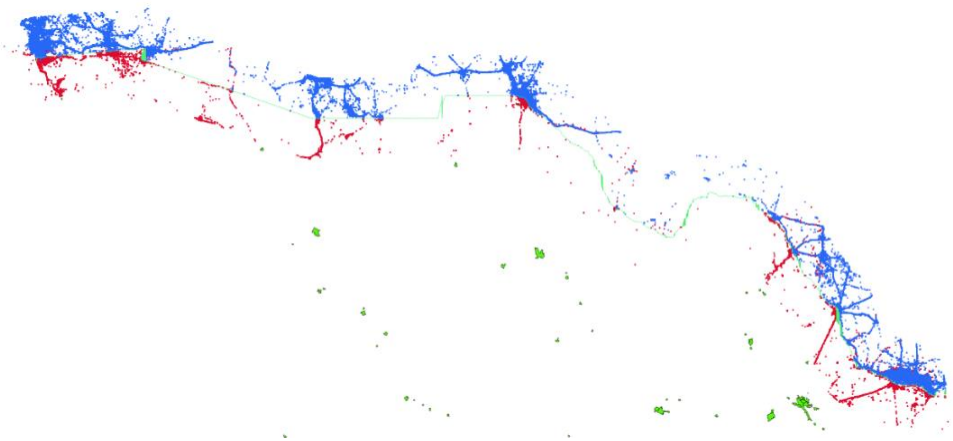
Topics of Study in the Pilot Test

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

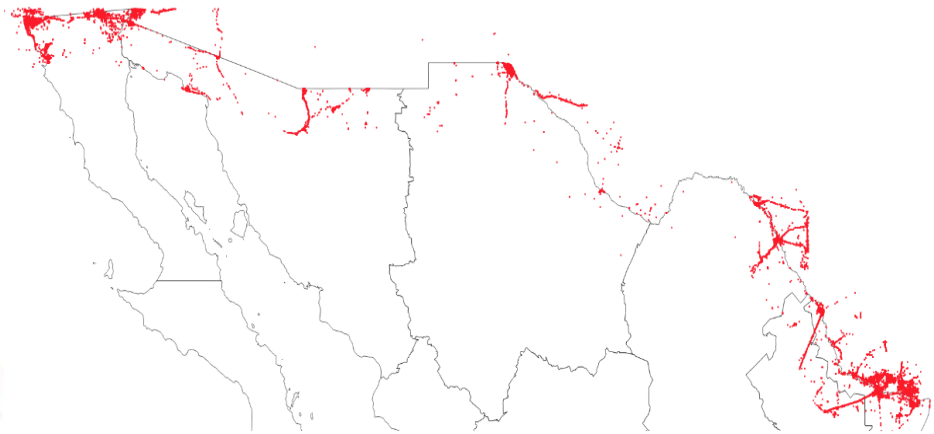
INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Mobility studies

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



Mexican (red) and US(blue) tweets

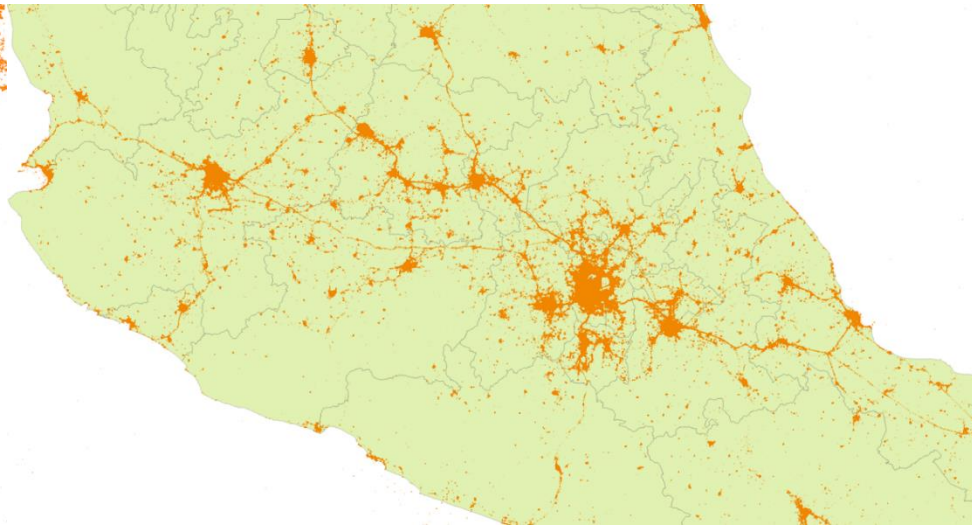


Mexican tweets

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Mobility studies

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



INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Topics of Study in the Pilot Test

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

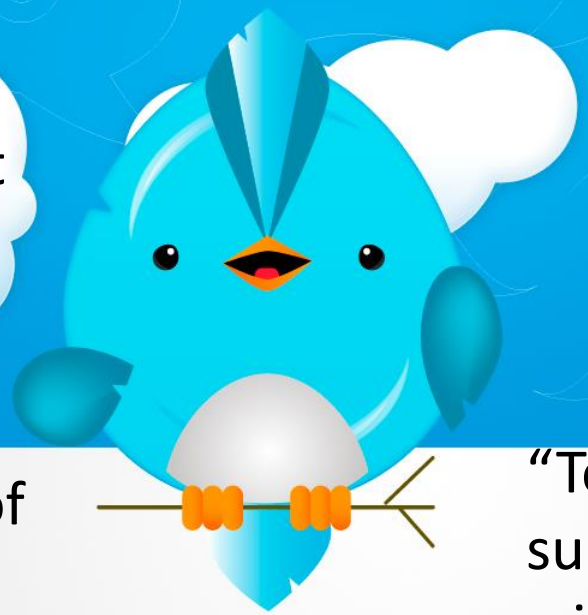
INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Twitter for Subjective Well-being

INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

http://cienciadedatos.inegi.org.mx/pioanalisis

App to create our training set of tweets



Humans grade a sample of tweets as possitive, neutral or negative, and classify them in several subjects.

PIOANALISIS

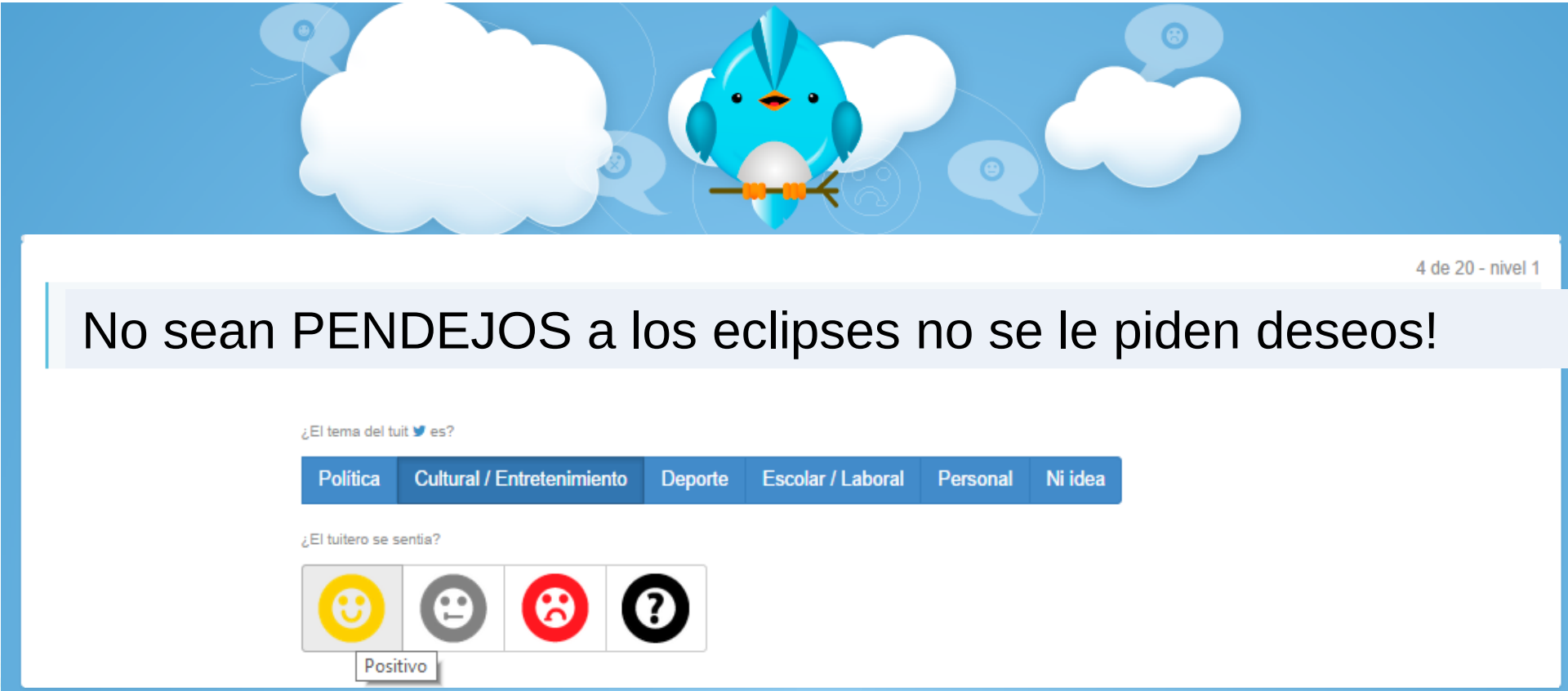
Bienestar Subjetivo y las redes sociales como fuente de datos

"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?



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

INEGI
INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

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