Session organized by the Asociación Mexicana de Estadística: Strengthening synergies between academic and official statistics: success stories.

Instructions: Click on the link to access each author's presentation.

Organiser: Victor Alfredo Bustos y de la Tijera

Chair: Martha María Téllez Rojo

Discussant: Andrés Christen

Participants:

Victor Alfredo Bustos y de la Tijera: A view from both sides.

Pedro Luis do Nascimento Silva: Win-Win Partnerships– Public Statistics Meets Academia

Erin Lundy: Collaboration and Engagement with Academia: The Statistics Canada Experience

IAOS-ISI 2024
MEXICO CONFERENCE

Strengthening Synergies Between Academic and Official Statistics: A view from both sides.

Alfredo Bustos

INEGI

International Statistical Institute

IAOS
IMPROVING OFFICIAL STATISTICS

Index

CIMAT-INEGI M. Sc. Program In Statistics

CONACyT-INEGI Fund

Mexican Tweeterers' mood

Lessons learnt

Capacity building

The Research Centre in Math, known as CIMAT, is in charge of an evolving graduate program aimed at improving capabilities of INEGI staff.

Collaboration

Escrow aimed at funding research and development at INEGI's request. As a result, alternative methodologies have been introduced.

Other exchanges

INEGI's first experience with Data Science applications. American and Mexican institutions were involved.

Follow-up

Projects derived from previous experiences: Automatic collection from electronic systems, and roadmap to produce stats from social networks and surveys.

LOCSISI 2024 MEXICO CONFERENCE

CIMAT-INEGI Master programs in Statistics

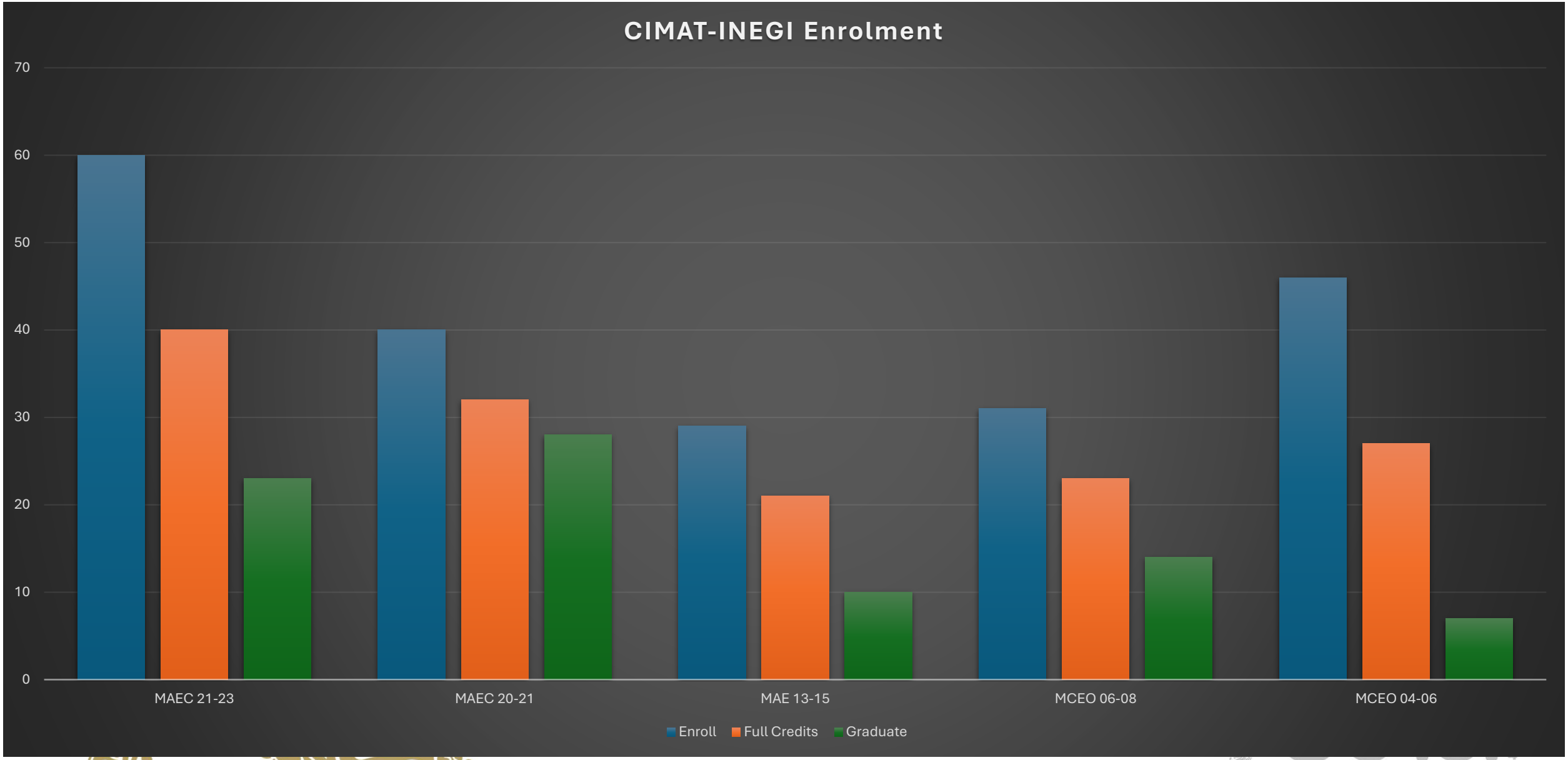
IOS-ISI 2024
MEXICO CONFERENCE

CIMAT-INEGI Master programs

Three stages:

- ✓ **Master in Official Statistics (2004-2006 and 2006-2008)**
- ✓ **Master in Statistical Analysis (2013-2015)**
- ✓ **Master in Statistical Analysis and Computing (2020-2021, 2021-2023, 2022-2024 and 2024-2026)**

Over 200 colleagues have gone through the program, at some point.
Only 79 have graduated. Some are already in decision-making positions.


CIMAT-INEGI Enrolment
70
60
50
40
30
20
10
0
MAEC 21-23
MAEC 20-21
MAE 13-15
MCEO 06-08
MCEO 04-06
Enroll
Full Credits
Graduate

CONACyT-INEGI Fund

IOS-ISI 2024
MEXICO CONFERENCE

Background

- The National Council for Science and Technology (CONACYT) created what came to be known as the CONACyT Funds.
- Their purpose was to coordinate efforts in generating knowledge, innovation, technological development, and formation of human resources.
- Researchers proposed innovative solutions to public administration problems and shortcomings.
- Generating new knowledge and/or raising new value added businesses.
- Through joint efforts, resources were allocated to fund research and development relevant to the Federal Government.
- The model they followed was an Escrow.
- Funds were allocated via competitive and efficient processes

IOS-ISI 2024
MEXICO CONFERENCE

CONACYT-INEGI Projects

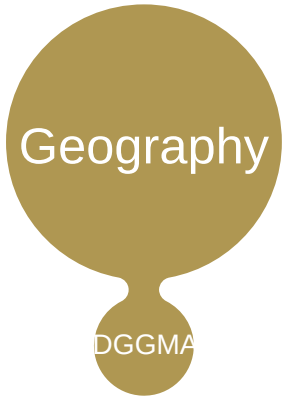
2011	2012	2013	2014	2015
Tenders 8 Completed 2	Tenders 8 Completed 7	Tenders 7 Completed 5	Tenders 7 Completed 7	Tenders 6 Completed 5

2016-1	2016-2	2017	2018-1	2018-2
Tenders 6 Completed 4	Tenders 6 Completed 6	Tenders 4 Completed 4	Tenders 2 Completed N/A	Tenders 1 Completed N/A

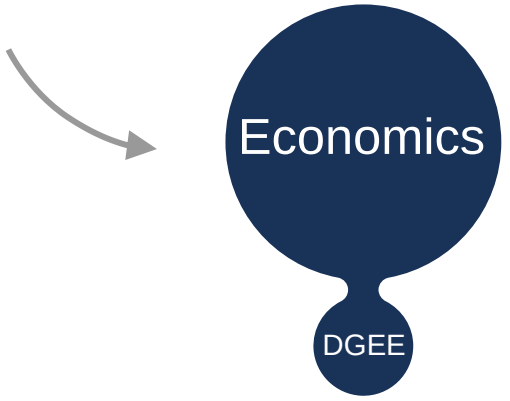
IOS-ISI 2024
MEXICO CONFERENCE

Examples

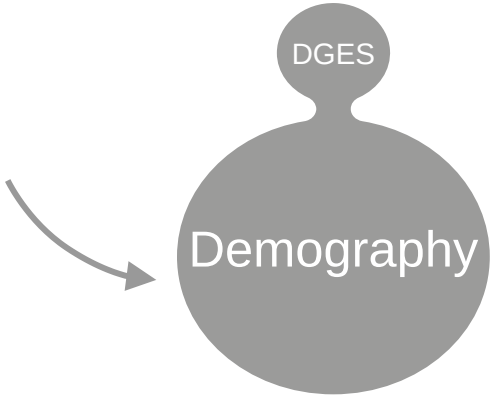
Satellite Hyperspectral and Lidar images for land use and vegetation types.



Models to enhance GNP sectoral disaggregation.



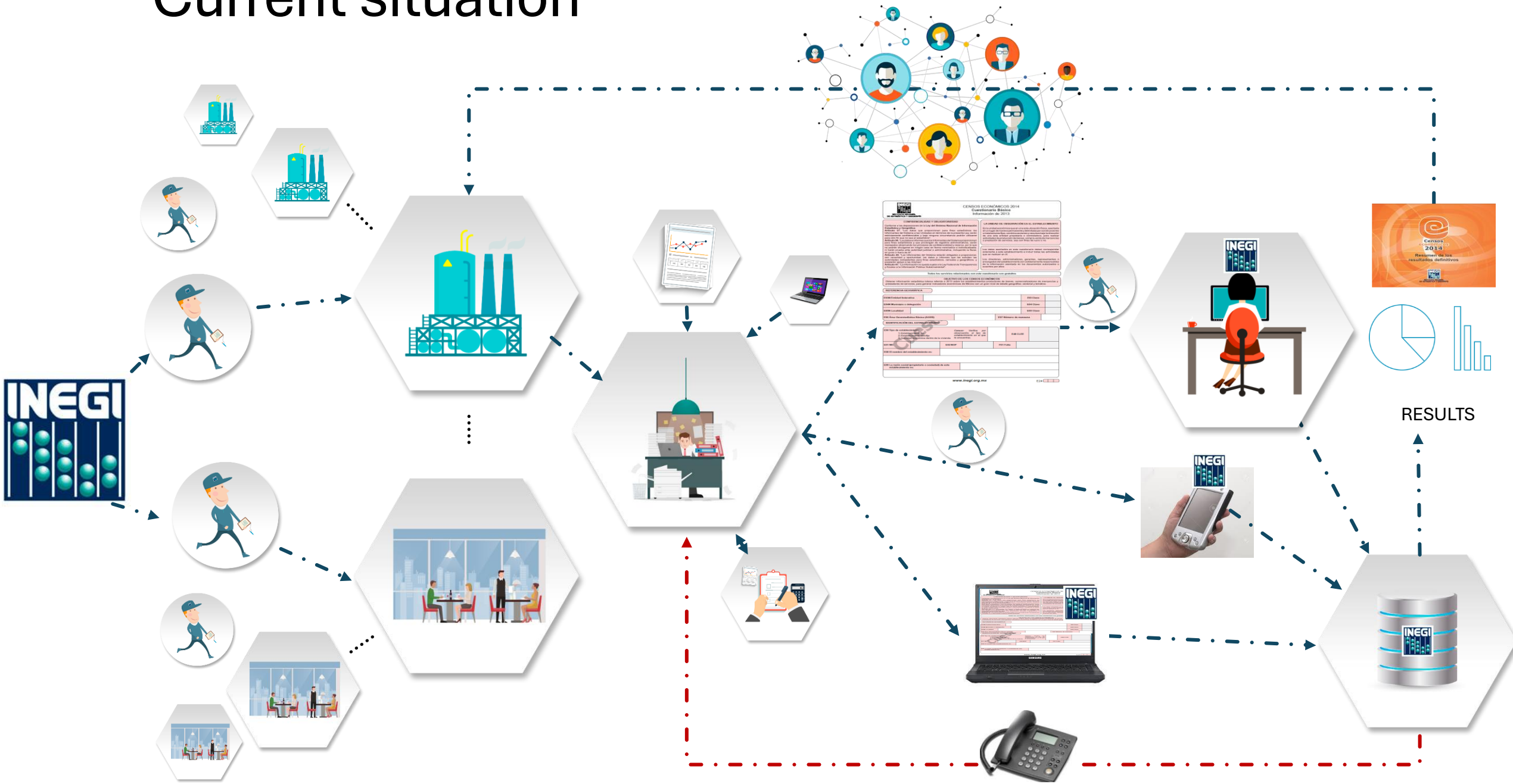
SAE to study socio-demographic characteristics of municipalities.



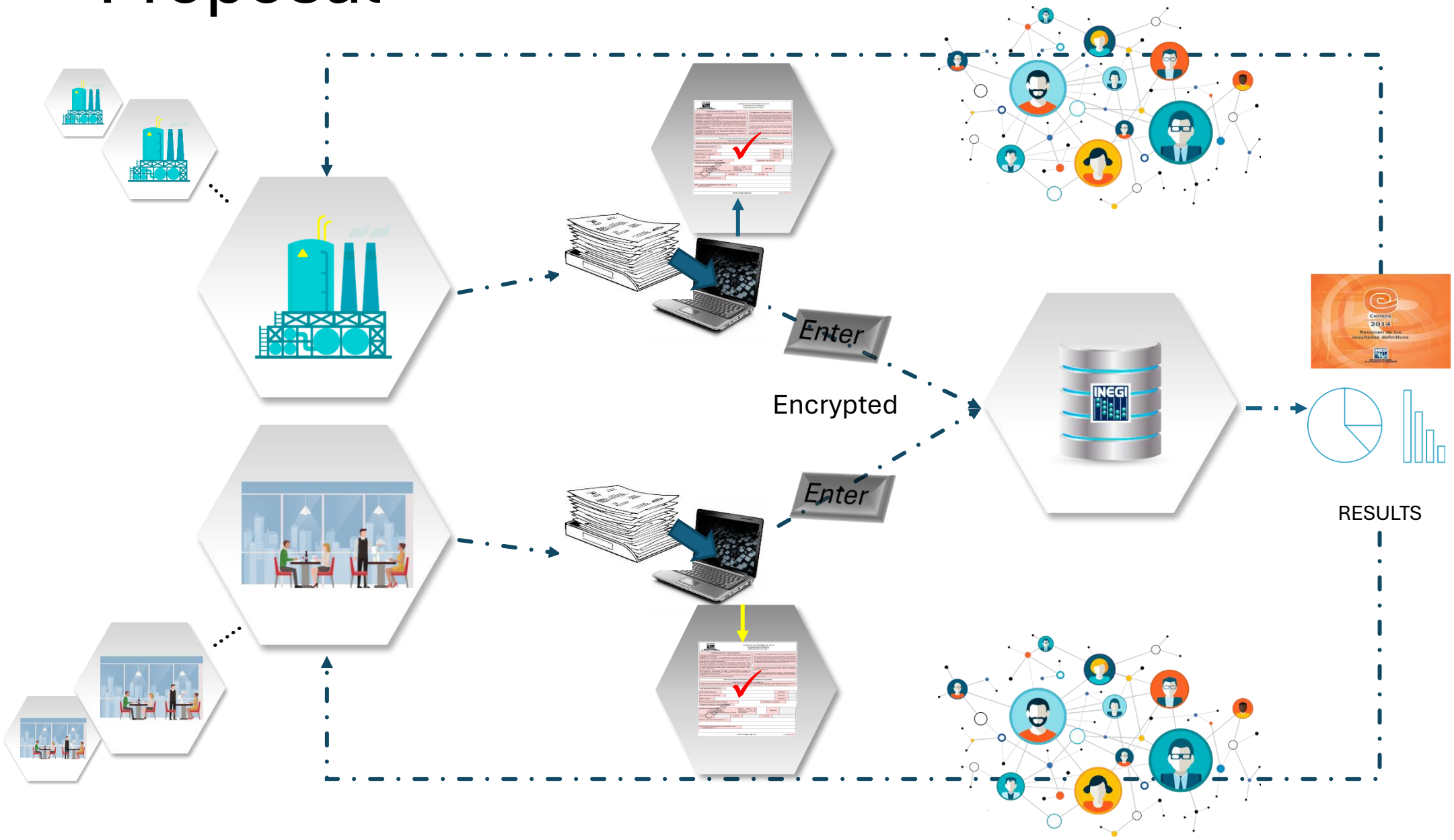
Framework to design and develop a computer architecture to automatically collect information for INEGI economic surveys using information technologies.

The objective of this 2013 project was to develop a model to automate the process of filling out economic survey and census questionnaires by collecting information from data contained in accounting computer systems.

Current situation



Proposal

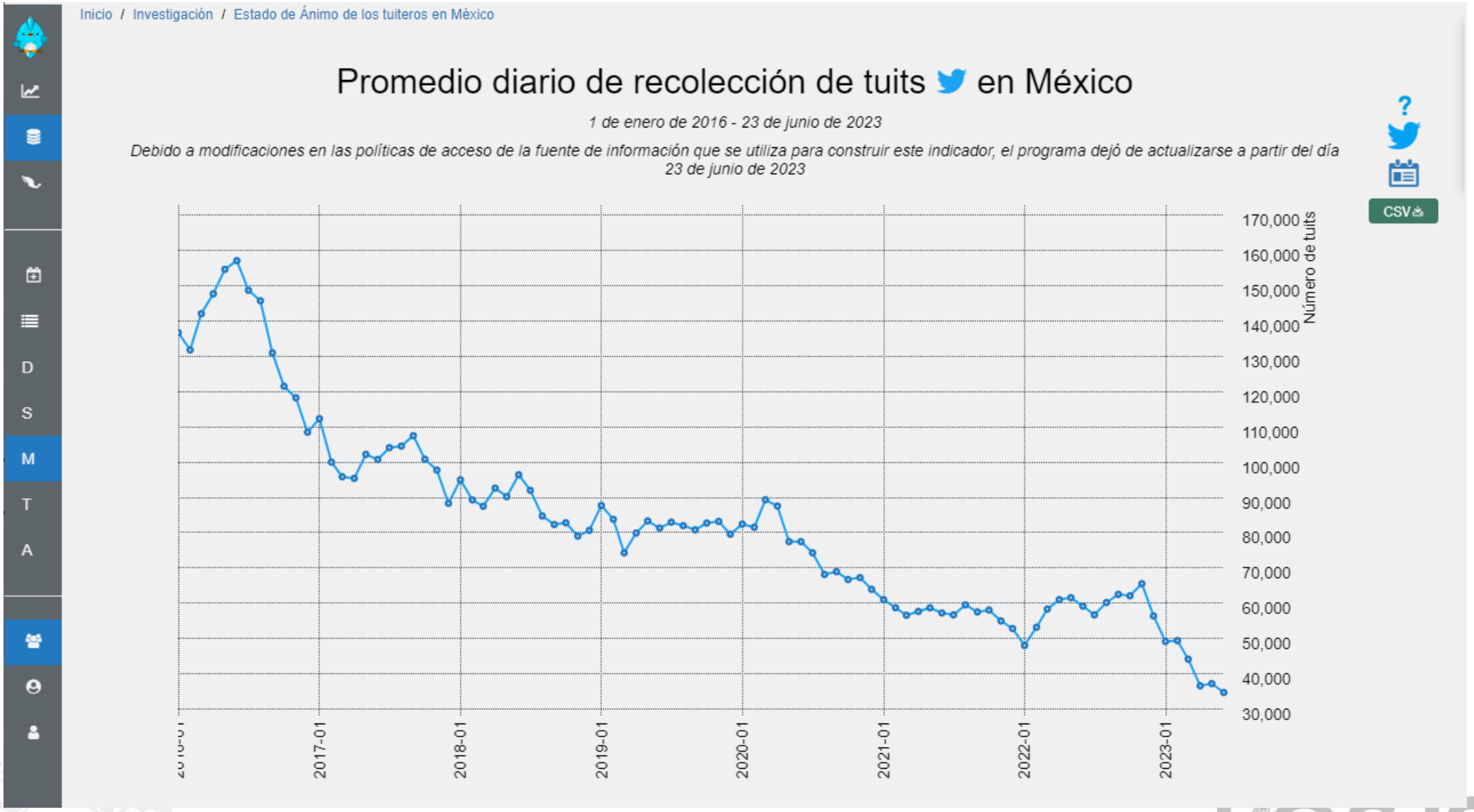


Tweeters' mood in Mexico

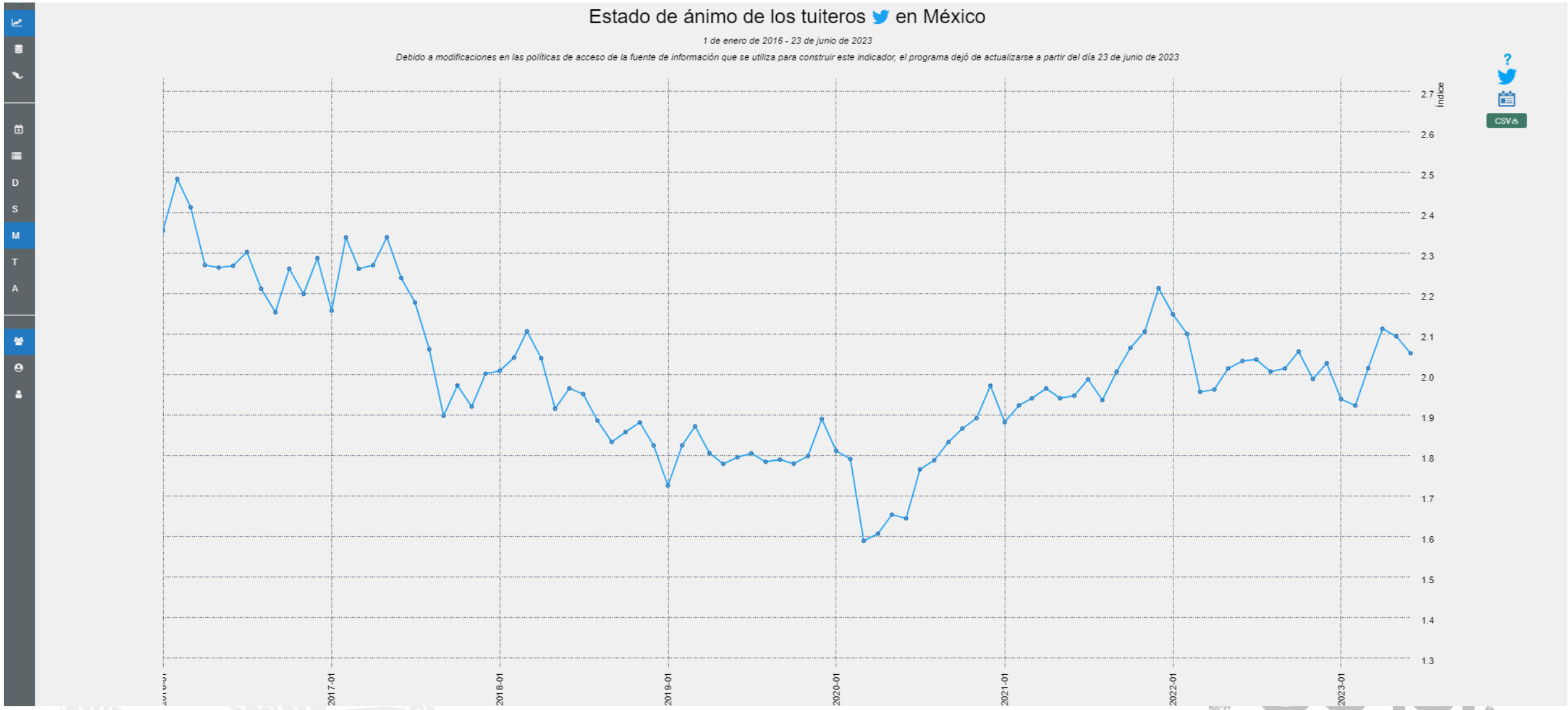
IOS-ISI 2024
MEXICO CONFERENCE

Mood of tweeters in Mexico.

In 2015, INEGI, INFOTEC and the Geo Center, and support from the Positive Psychology Center, of the University of Pennsylvania (PPC-UPenn), and the Universidad Tec Milenio (UTM), published the first version of the Tweeters' Mood.
A step towards using Big Data sources to generate new experimental statistics for the field of Subjective Well-being.
On-line results ranged from collection of data, to processing of each post, to the visual representation of results.

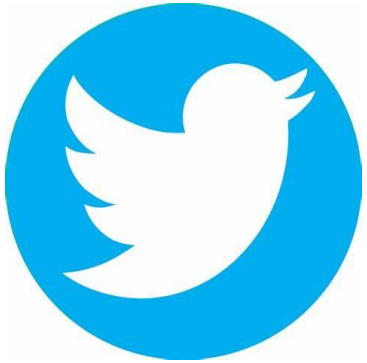


https://www.inegi.org.mx/app/animotuitero/#/app/collect



https://www.inegi.org.mx/app/animotuitero/#/app/multiline

Lessons learnt



A little bird told me:
Roadmap towards more frequent and representative official statistical information, through joint use of social network posts and survey data.

Proposal for National Statistical Offices (NSOs) to produce representative information on multiple topics, with greater frequency, jointly using household survey data and social media posts, using AI tools.

IOS-ISI 2024 MEXICO CONFERENCE

- First stage:
  - Select a random sample of a few thousand households or individuals: S_i. To be termed, the *small sample*.
  - Collect topic and sociodemographic (SD) data: $Q(S_i)$.
  - Identify social media users in the household and ask for their usernames, $S_i^{\mathbb{X}} = S_i \cap \mathbb{X}$.
  - For each user, tag recent tweets with his/her survey information: $Q\left(S_i^{\mathbb{X}}\right) \otimes P\left(S_i^{\mathbb{X}}\right)$.
  - Train and evaluate algorithms: $Q_i\left(S_i^{\mathbb{X}}\right) = \mathcal{A}_1\left(P\left(S_i^{\mathbb{X}}\right)\right)$
  - Choose those that are suitable.
- Second stage:
  - Automatically tag all recent public posts at your disposal: $\hat{Q}_k\left(P(\mathbb{X})\right) = \mathcal{A}_k\left(P(\mathbb{X})\right)$.
  - Collect all tagged posts from the same user.
  - Apply rule to tag all users with own posts: $\hat{Q}_k(\mathbb{X})$.
  - Use their SD tags to correct for selection bias. This is termed the *large sample*: $\mathbb{X}^{(\omega)}$.
  - **The Holly Grail of databases**: $\hat{Q}\left(\mathbb{X}^{(\omega)}\right) = (\mathcal{W}_{SD}, \quad \hat{Q}_1(\mathbb{X}), \quad \hat{Q}_2(\mathbb{X}), \quad \cdots, \quad \hat{Q}_k(\mathbb{X}))$
  - Publish daily, weekly, monthly, etc. results whenever precision allows, until next survey round.

IOS-ISI 2024 MEXICO CONFERENCE

Roadmap towards more frequent and representative official statistical information, through joint use of social network post and survey data.

Survey	Random sample's answers	Identify respondents who are $\mathbb{X}$ users	Tag their posts $P(S_j^{\mathbb{X}})$, form training (T) and testing (t) sets. Socio-demographic (SD) variables, too	Train and test algorithms. Use sampling design?	Tag all posts in time interval, including SD	Tag users, to form reweighted database	Infer to whole population by topic or across them	
Topic 1	$Q(S_1)$	$S_1^{\mathbb{X}} = S_1 \cap \mathbb{X}$	$T_1(S_1^{\mathbb{X}}) = Q(S_1^{\mathbb{X}}) \otimes P(S_1^{\mathbb{X}})$ $T_1 = T_{1T} \cup T_{1t}$	$Q_1(S_1^{\mathbb{X}}) = \mathcal{A}_1\left(P(S_1^{\mathbb{X}})\right)$	$\hat{Q}_1(P(\mathbb{X})) = \mathcal{A}_1(P(\mathbb{X}))$	$\hat{Q}(\mathbb{X}^{(\omega)}) = (w_{SD}, \hat{Q}_1(\mathbb{X}), \hat{Q}_2(\mathbb{X}), \dots \hat{Q}_k(\mathbb{X}))$	$\hat{Q}_1(\mathbb{X}^{(\omega)})$	$\hat{Q}(\mathbb{X}^{(\omega)})$
Topic 2	$Q(S_2)$	$S_2^{\mathbb{X}} = S_2 \cap \mathbb{X}$	$T_2(S_2^{\mathbb{X}}) = Q(S_2^{\mathbb{X}}) \otimes P(S_2^{\mathbb{X}})$ $T_2 = T_{2T} \cup T_{2t}$	$Q_2(S_2^{\mathbb{X}}) = \mathcal{A}_2\left(P(S_2^{\mathbb{X}})\right)$	$\hat{Q}_2(P(\mathbb{X})) = \mathcal{A}_2(P(\mathbb{X}))$		$\hat{Q}_2(\mathbb{X}^{(\omega)})$	
…	…	…	…	…	…		…	
Topic k	$Q(S_k)$	$S_k^{\mathbb{X}} = S_k \cap \mathbb{X}$	$T_k(S_k^{\mathbb{X}}) = Q(S_k^{\mathbb{X}}) \otimes P(S_k^{\mathbb{X}})$ $T_k = T_{kT} \cup T_{kt}$	$Q_k(S_k^{\mathbb{X}}) = \mathcal{A}_k\left(P(S_1^{\mathbb{X}})\right)$	$\hat{Q}_k(P(\mathbb{X})) = \mathcal{A}_k(P(\mathbb{X}))$		$\hat{Q}_k(\mathbb{X}^{(\omega)})$	

IOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI

International Statistical Institute

IAOS
IMPROVING OFFICIAL STATISTICS

IOS-ISI 2024
MEXICO CONFERENCE

Win-Win Partnerships – Public Statistics Meets Academia

Pedro Silva (SCIENCE, Brazil)

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

Context (1)

Producing quality and trusted public statistics is a complex endeavor, requiring substantial legal, political, subject matter, methodological, and operational knowledge and expertise from the stats agencies.

The demand is increasing for more timely and detailed data, and for data on new topics, to support evidence-based decision-making.

Society's *datafication* is increasing pressure on stats agencies, where resources are not increasing in proportion to demand.

Context (2)

Official/public statistics evolve slowly, since its core sources such as censuses and surveys typically require lengthy development (user consultation, experimentation, and implementation) to provide data meeting the stringent applicable quality standards.

Society's data needs, on the other hand, evolve ever more quickly, driven by innovation, technology, and a thirst for data to guide decisions.

How to cope?

Innovation will be key to enable stats agencies to succeed.

And speeding innovation requires partnerships with:

- Other stats agencies;
- Other data providers;
- Academia.

Stats agencies partnerships with academia can lead to win-win situations.

Why partner with academia?

Because stats agencies can:

- Gain access to expertise that is impossible to have in-house;
- Influence research agendas;
- Influence education and development of future professionals it may later recruit;
- Develop a public voice to support its legitimacy and trustworthiness;
- Learn from the dialogue entailed;
- Speed up innovation and adaptation.

Why would academia want to partner?

Academia would benefit by:

- Participating in the decision-making on what data is to be produced;
- Gaining firsthand access to what data is being produced and what is learned from it;
- Learning how best to explore and analyze data;
- Engaging in relevant problem-solving and getting ideas on good topics for research;
- Adapting educational projects and practices to what is needed and expected.

IOS-ISI 2024 MEXICO CONFERENCE

Some examples worth considering

I will now highlight some examples that may be worth learning from, reflecting experiences in four countries:

- Brazil,
- Canada,
- UK and Europe.

Brazil

IOS-ISI 2024
MEXICO CONFERENCE

ENCE - National School of Statistical Science

Brazil's NSO (IBGE) is home to ENCE, founded in 1953.

ENCE is the **research and education** branch of IBGE, providing:

- Undergraduate degree in Statistics (over 2,000 graduates);
- MSc and PhD degrees in Population, Territory and Public Statistics (over 440 graduates);
- Training and continued professional development for IBGE's staff, including distance learning (161 courses with over 144,000 people/course in 2023).

IOS-ISI 2024 MEXICO CONFERENCE

ENCE's contributions to IBGE

- Conducting research in topics of relevance to IBGE's activities.
- Organizing seminars such as IBGE's Methodology Seminar series, co-organizing the NIC.br Methodology Seminar series and more.
- Participating in a range of capacities in working groups, committees and other survey / census planning, execution, analysis and evaluation.

ENCE's contributions to IBGE

- Mediating collaboration with other academic institutions, such as in the publication of the *Revista Brasileira de Estatística* (established in 1940 and now jointly published with the *Associação Brasileira de Estatística*).

RBEs
Revista Brasileira de Estatística

- UN Big Data Regional Hub in Brazil

UN Big Data Regional Hub in Brazil

IOS-ISI 2024 MEXICO CONFERENCE

Canada

IOS-ISI 2024
MEXICO CONFERENCE

Stats Canada's partnerships

- Advisory Committee on Statistical Methods
- Survey Methodology Journal
- 2024 International Methodology Symposium
  - "Shaping the Future of Official Statistics"
- "Maintain collaboration and exchanges with the scientific, academic and business communities through participation of employees in professional associations and other peer gatherings."

  Quality Assurance Framework

IOS-ISI 2024 MEXICO CONFERENCE

Europe + UK

IOS-ISI 2024
MEXICO CONFERENCE

UK partnerships

- MSc in Official Statistics (MOffStat) – Partnership between ONS and the University of Southampton – now closed.
- ➔ MDataGov - Data Analytics for Government
- ➔ EMOS - European Master in Official Statistics

UK partnerships

MDataGov aims to:

- "… equip public sector employees with the skills required from a modern public sector data analyst, helping to increase data science capability in public sector organisations."

Four university partners deliver the MDataGov courses:

- Cardiff University
- University of Southampton
- Oxford Brookes University
- University of Glasgow

Data Science Campus

IOS-ISI 2024 MEXICO CONFERENCE

European partnerships

EMOS aims to:

- “advance postgraduate education in official statistics and data science at universities across Europe;”
- “educate students to become highly skilled statisticians with expertise in official statistics;”
- “facilitate the exchange of knowledge and experience between the European producers of official statistics and lecturers, researchers, and students in this field.”

UK partnerships

- ONS Data Science Campus
- Methodological Assurance Review Panel (UK Stats Authority)
- Research and consulting contracts
  - Southampton University

US

IOS-ISI 2024
MEXICO CONFERENCE

US partnerships

- Federal Statistics Research Data Centers
- NSF-Census Research Network
- “Provides support for a set of research nodes, each of which is staffed by a team of scientists conducting interdisciplinary research and educational activities on methodological questions of interest and significance to the broader research community and to the Federal Statistical System, particularly the U.S. Census Bureau”.

Concluding

IOS-ISI 2024
MEXICO CONFERENCE

Concluding remarks

- Given its benefits and all these success stories, it is surprising that we do not see more of these partnerships.
- Yes, it may be difficult to start or in the beginning, but it is worth the effort.
- Remember what is at stake: innovate or face the consequences…

IOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI

International Statistical Institute

IAOS
IMPROVING OFFICIAL STATISTICS

IAOS-ISI 2024
MEXICO CONFERENCE

Collaboration and Engagement with Academia: The Statistics Canada Experience

Erin Lundy
IAOS-ISI Mexico City
May 15th 2024

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

Statistics Canada

- Canada's central statistical office responsible for producing statistical data and insights to help Canadians better understand their country's population, resources, economy, society and culture.
- Over 7,000 employees across the agency.

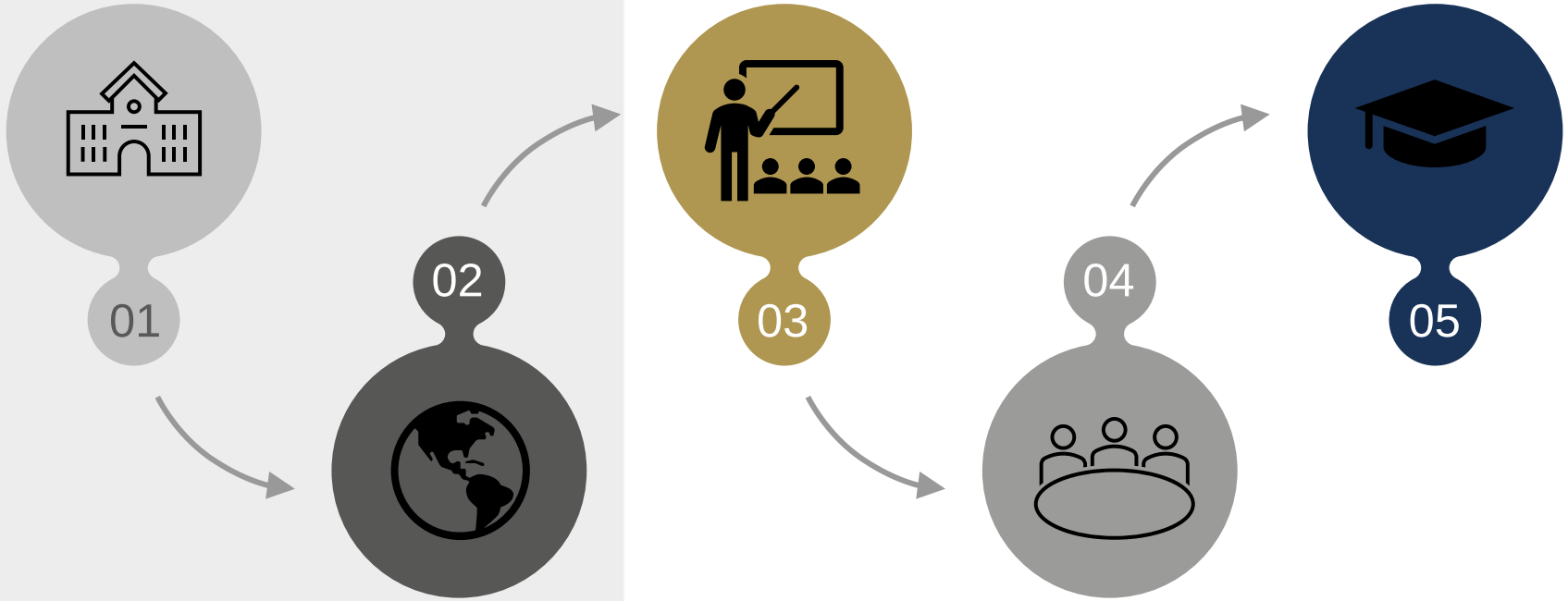
- Focus on collaboration experience within the Modern Statistical Methods and Data Science Branch.
  - ➢ Provide expertise to Statistics Canada statistical programs; conducts research and development of modern methods; teaches statistical and methodology courses; and consults, collaborates and builds capacity in international cooperation programs.
  - ➢ Approximately 400 employees, primarily methodologists and data scientists.

Areas of collaboration

University faculty

Research forums

Students



International collaboration

Professional organizations and conferences

IOS-ISI 2024 MEXICO CONFERENCE

Why collaborate?

Development of innovative methods to address practical problems

Improvement of statistical methods and programs

Increased visibility for agency within the wider statistical community

Learning experience for Statistics Canada staff

Why collaborate ?

Sharing Knowledge

Research motivated by challenges in production of official statistics

Publications in peer-reviewed journals

Increased visibility for research within the wider statistical community

Learning experiences for students

IOS-ISI 2024 MEXICO CONFERENCE

University Faculty

IOS-ISI 2024
MEXICO CONFERENCE

Research collaboration with university faculty

- Theoretical developments motivated by practical problems at Statistics Canada.
- Collaboration with several Statistics Canada key initiatives.
- Many of these collaborations have resulted in publications in peer-reviewed statistical journals and implementation of innovative methods in our statistical programs.
- Several past and long-standing collaborations with well-known researchers in the area of survey sampling.

Research collaboration with university faculty

- Areas of research include:
  - Data analysis from complex survey designs
  - Estimation using auxiliary data
  - Estimation of variance due to imputation
  - Treatment of outliers in survey data
  - Re-sampling methods
  - Small area estimation
  - Nowcasting and time series modelling

Implementation in Statistics Canada programs

- Labour Force Survey
  - Composite estimation
  - Resampling methods
  - Small area estimation
- Small area estimation
  - Monthly Survey of Manufacturing
  - Visitor Travel Survey
  - Census net undercoverage
- Generalized systems
  - Generalized regression estimation
  - Edit and imputation
  - Variance estimation due to non-response and imputation

Advisory Committee on Statistical Methods

- Bi-yearly meeting, first held in Spring 1985.
- Purpose is to advise the Chief Statistician on matters related to the utilization of efficient statistical methods in the Agency's programs and its programs of research and development in statistical methods.
- Notable members past and present: Morris Hansen, J.N.K. Rao, Wayne Fuller, Graham Kalton, Mary Thompson, Changbao Wu, Carl-Erik Sarnal, Leslie Kish, Phillip Kott

Advisory Committee on Statistical Methods

Activities of the committee include:

- Review and comment on Statistics Canada's priorities in statistical research.
- Review and comment on methods used in particular programs.
- Review and comment on generic methods used widely in the Agency's programs
- Suggest functions or programs within the Agency that could benefit from the innovative application of statistical methods.
- Suggest means by which Statistics Canada can ensure its continuing leadership. role in the development of statistical methods.

International Collaboration

IOS-ISI 2024
MEXICO CONFERENCE

International research collaboration

Statistics Canada staff have published articles in refereed journals with other researchers from:

- Other National Statistical Organizations
  - US Census Bureau
  - DESTATIS (Germany)
  - Istat (Italy)
  - Australian Bureau of Statistics
  - INSEE (France)
- La Poste (Postal Service in France)
- University of Southampton
- Toulouse School of Economics

Research Forums

IOS-ISI 2024
MEXICO CONFERENCE

Survey Methodology journal

- Peer-review journal published twice a year since 1975 by Statistics Canada.
- Publishes innovative theoretical or applied research papers on statistical methods relevant to National Statistical Offices or other statistical organizations.
- Editorial board includes academic statisticians, researchers from other National Statistics Offices as well as Statistics Canada staff with expertise in areas relevant to official statistics.
- Submissions from both official statisticians and academics.

International Methodology Symposium

- Statistics Canada organizes an annual conference to bring together analysts, researchers, academics, planners and policy- and decision-makers.
- 2024 Symposium - Shaping the future of official statistics.
- Past themes:
  - Data disaggregation: building a more representative data portrait of society
  - Adopting data science in official statistics to meet society's emerging needs
  - Combine to conquer: innovations in the use of multiple sources of data

IOS-ISI 2024 MEXICO CONFERENCE

International Methodology Symposium

- Includes workshops by university professors, researchers from other government organizations and industry professionals. Past examples:
  - ➢Methods for Multiple Frame Surveys
  - ➢Use of administrative data for socio-economic statistics
  - ➢ Statistical learning with applications for official statistics
  - ➢Survey data integration
- Hosted in Ottawa, opportunities to connect with academic statisticians.

Professional Organizations and Conferences

IOS-ISI 2024
MEXICO CONFERENCE

Statistical Society of Canada

- Active participation in the Statistical Society of Canada.
- Members of board of directors (past and present).
- Staff encouraged to present work at annual meeting.
- Particularly within the Survey Methods Section.
  - Aims to promote the development and use of the statistical methods in surveys and in official statistics.
  - Several staff have served as member of the executive committee of this section.

Canadian Statistical Sciences Institute

- National institute that advances the development, application and communication of innovative statistical research.
- Aims to build collaborations between academia, industry and government.
- Statistics Canada currently chairing the Board of Directors.
- Participation in Collaborative Research Teams projects which foster research and training initiatives that span disciplines and institutions to tackle complex problems of broad importance
- Recent projects:
  - Synthetic Data and Risk Measures for Statistical Disclosure Control
  - Modern Techniques for Survey Sampling and Complex Data

IOS-ISI 2024
MEXICO CONFERENCE

International conferences

- Organization of international conferences and workshops:
  - ➢Conference of International Establishment Statistics (ICES),
  - ➢European Network for the Betterment of Establishment Statistics (ENBES) workshop
  - ➢United Nations Economic Commission for Europe (UNECE) Edit and Imputation workshops
  - ➢Colloque Francophone sur les Sondages

Students

IOS-ISI 2024
MEXICO CONFERENCE

Formal Programs

- Co-op students
- Agreements with participating French universities for an internship-type program
- Recruitment campaigns
- MITACS program
  - Nonprofit national research organization that operates research and training programs
- Agreements with several universities for PhD student research

Informal Arrangements

- Graduate research motivated by challenges in production of official statistics
- Staff undertaking part-time graduate studies
- Several former graduate students of faculty collaborators currently employees at Statistics Canada

New Initiatives with Carleton University

- Part of long-standing relationship with Carleton University (Ottawa, Canada)
- Data science course
  - ➢Class project for graduate course on data science involving nowcasting of Canadian Labor market indicators
  - ➢Students to worked in groups under guidance of professor and shared findings
- Development and evaluation of statistical models for a combined census
  - ➢First step - undergraduate student honours project

Summary

- Multifaceted approach to collaboration and engagement with academia.
- Built on mutually beneficial relationships.
- Key outcomes:
  - Development of innovative methods to address practical problems
  - Improvement of statistical methods and programs
  - Increased visibility for research with the wider statistical community
  - Valuable learning experiences

IOS-ISI 2024
MEXICO CONFERENCE

Thank you

Merci

For more information, please contact

Erin Lundy
erin.lundy@statcan.gc.ca

Pour plus d'information, veuillez contacter

Erin Lundy
erin.lundy@statcan.gc.ca

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS