Creating a National Data Architecture for Evidence-Based Policy

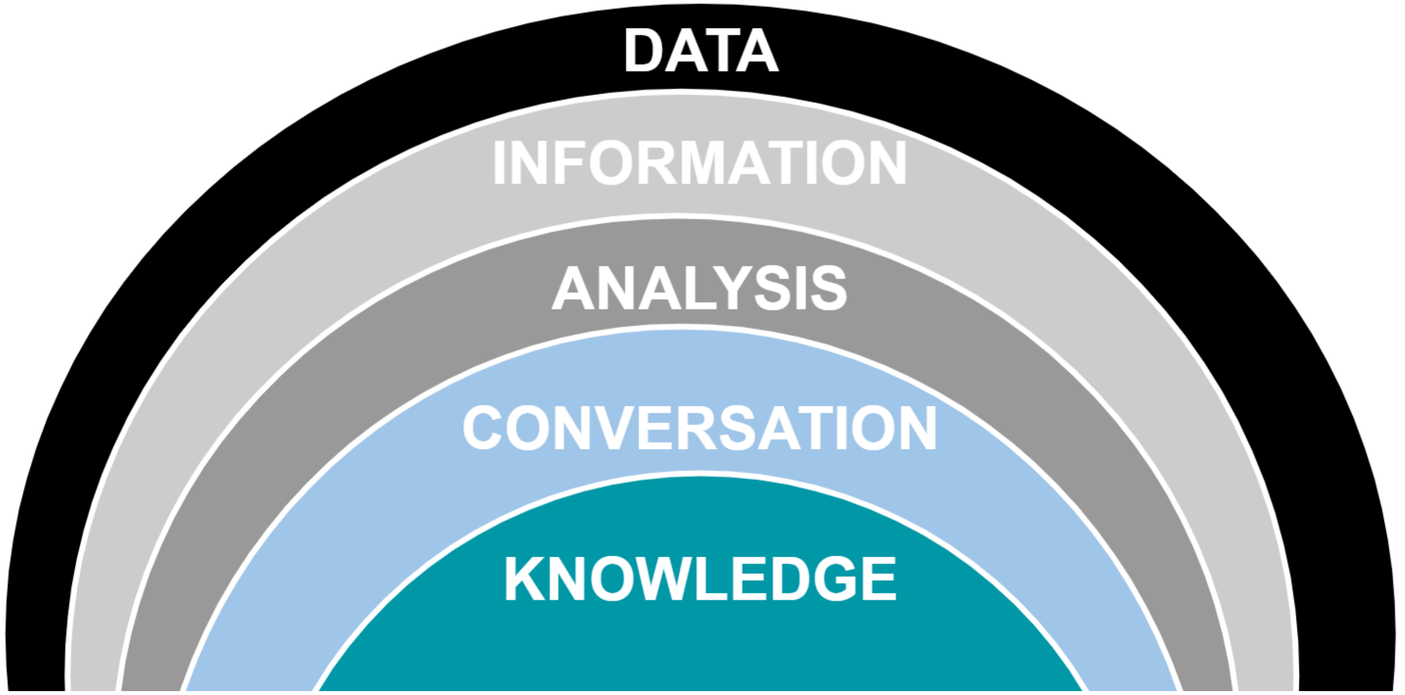
Andrea Fernandez
Deputy Director of Dissemination
andrea.fernandez@inegi.org.mx

March 10, 2020

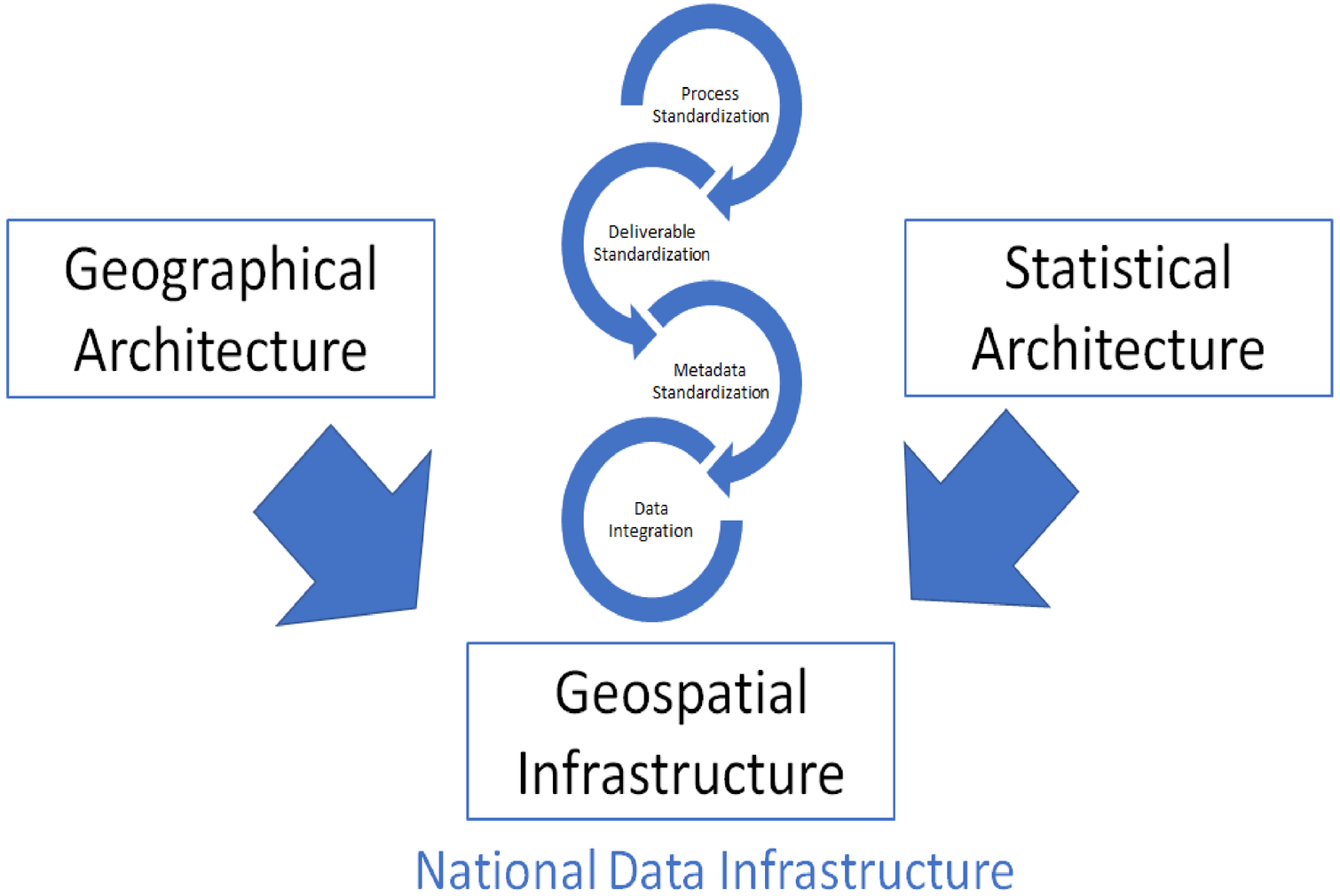
Table of Contents

01 Business Problem

- From data to knowledge
- From data producer to data user (policy maker)



01
General Approach



02
Basic descriptions

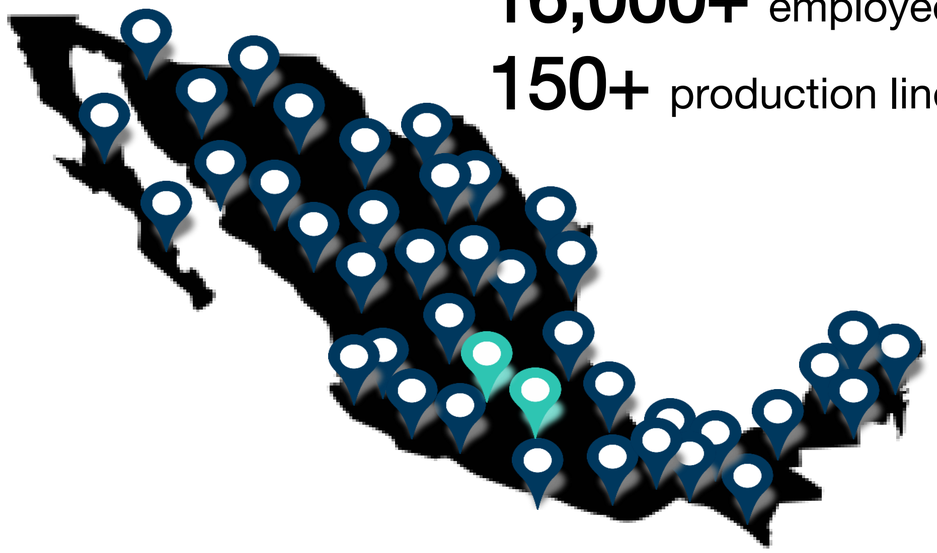
- National Institute of Statistics and Geography (INEGI)
- The National Statistical System
- Production of Statistical Information
- Production of Geographical Information
- Metadata
- Analysis taxonomy

❖ 02.1

National Institute of Statistics and Geography

A Federal Government Organization

50+ offices nationwide
16,000+ employees
150+ production lines



Two roles

1. Coordination of the National Statistical and Geographical System (Information produced by government agencies that support the design and evaluation of public policy.)
2. Production of Official Statistics and Geographical Information

Technically Autonomous Institution

- Headed by five board members (1 President, 4 Vice presidents) appointed by the Senate.

4 Divisions that produce information:

- Household
- Business
- Government and Security
- Geographic Information

❖ 02.2

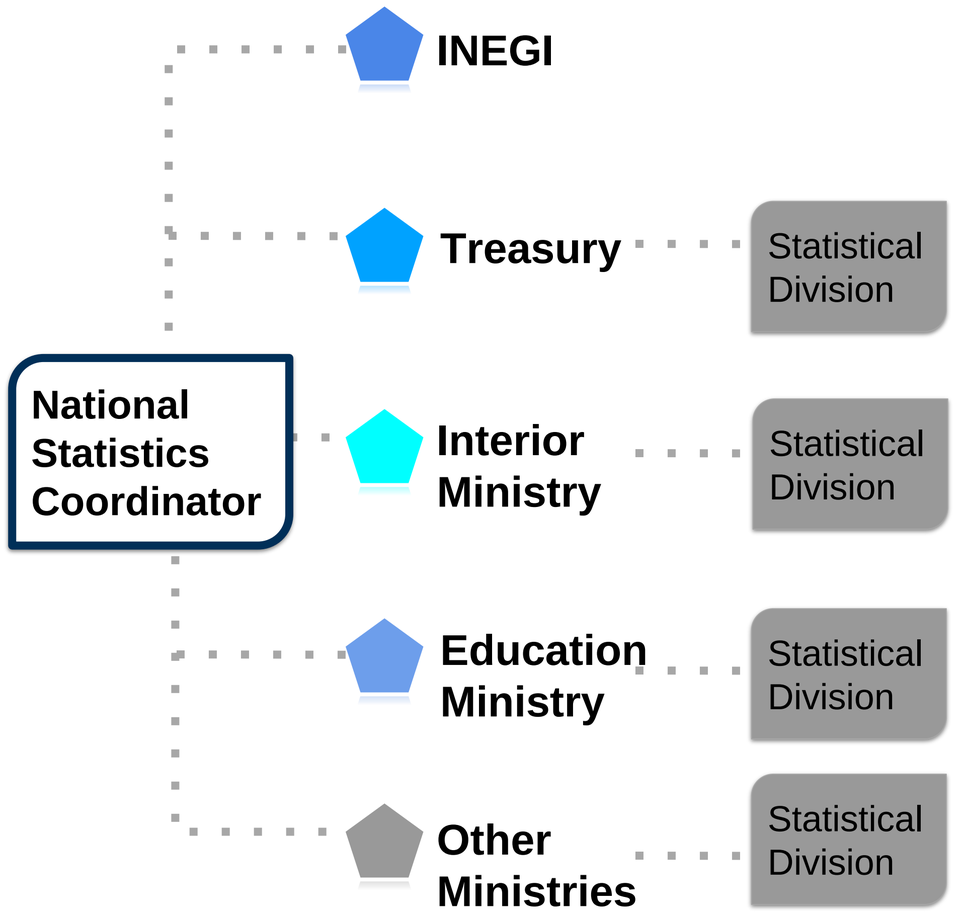
The National Statistical System

Mandate

To produce information that supports the design and evaluation of public policy

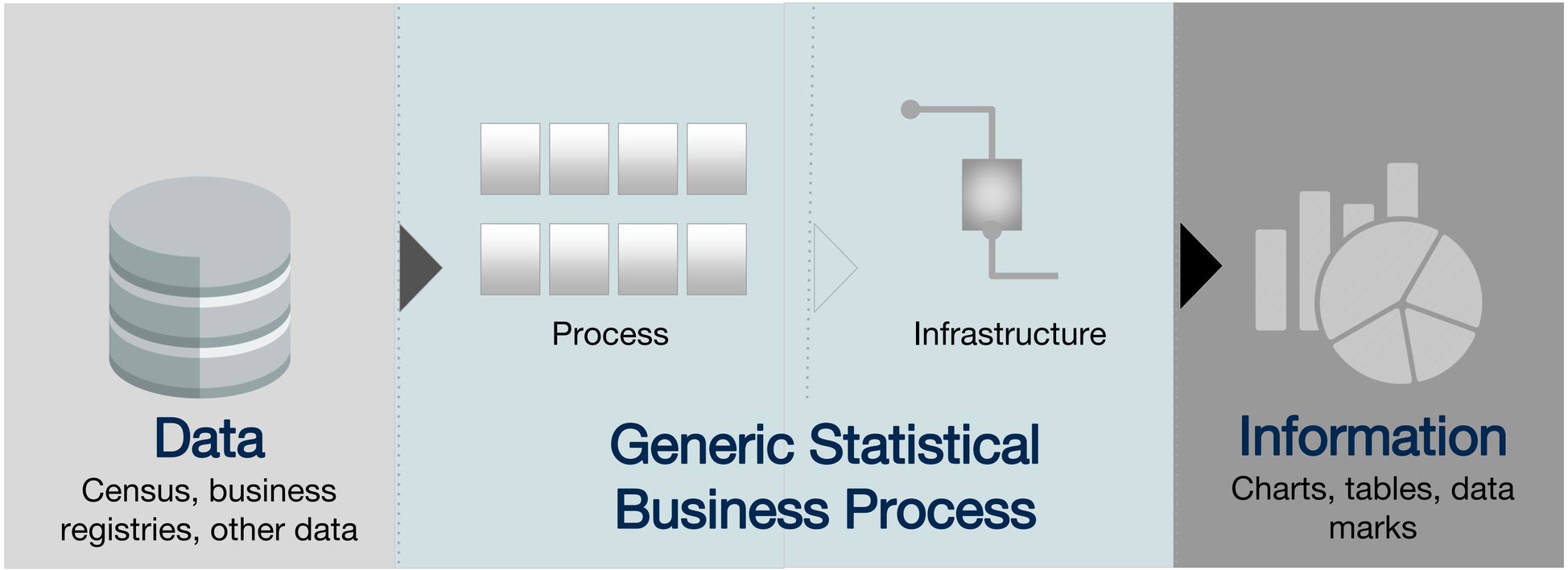
Status quo:

Multiple data producers with dispersed systems of records, various types of data, and conceptual frameworks.



❖ 02.3

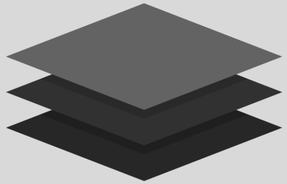
Production of Statistical Information



❖ 02.4

Production of Geographical Information

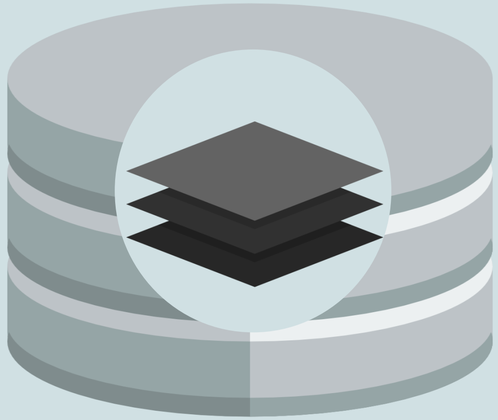
Lines
Points
Areas
Complex



Maps
Attributes
Images
Lidar
GPS

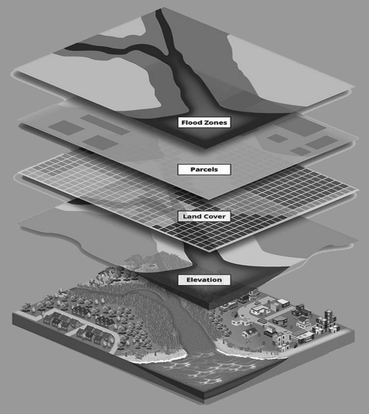
Data

Hard copy
Digital form



Geographical Information Systems

(Standardized following GSBPM)



Information

Hard copy
Digital form

❖ 02.5

Metadata

Metadata describes data lineage and relevant information that provides context, contents, and meaning.

❖ Statistics

DDI DATA DOCUMENTATION INITIATIVE

sdmx Statistical Data and Metadata eXchange

GSIM

❖ GIS

ISO Attribute standards (metadata) to produce and disseminate information.

OGC Making location count.

ISO ISO/TC 211 19100

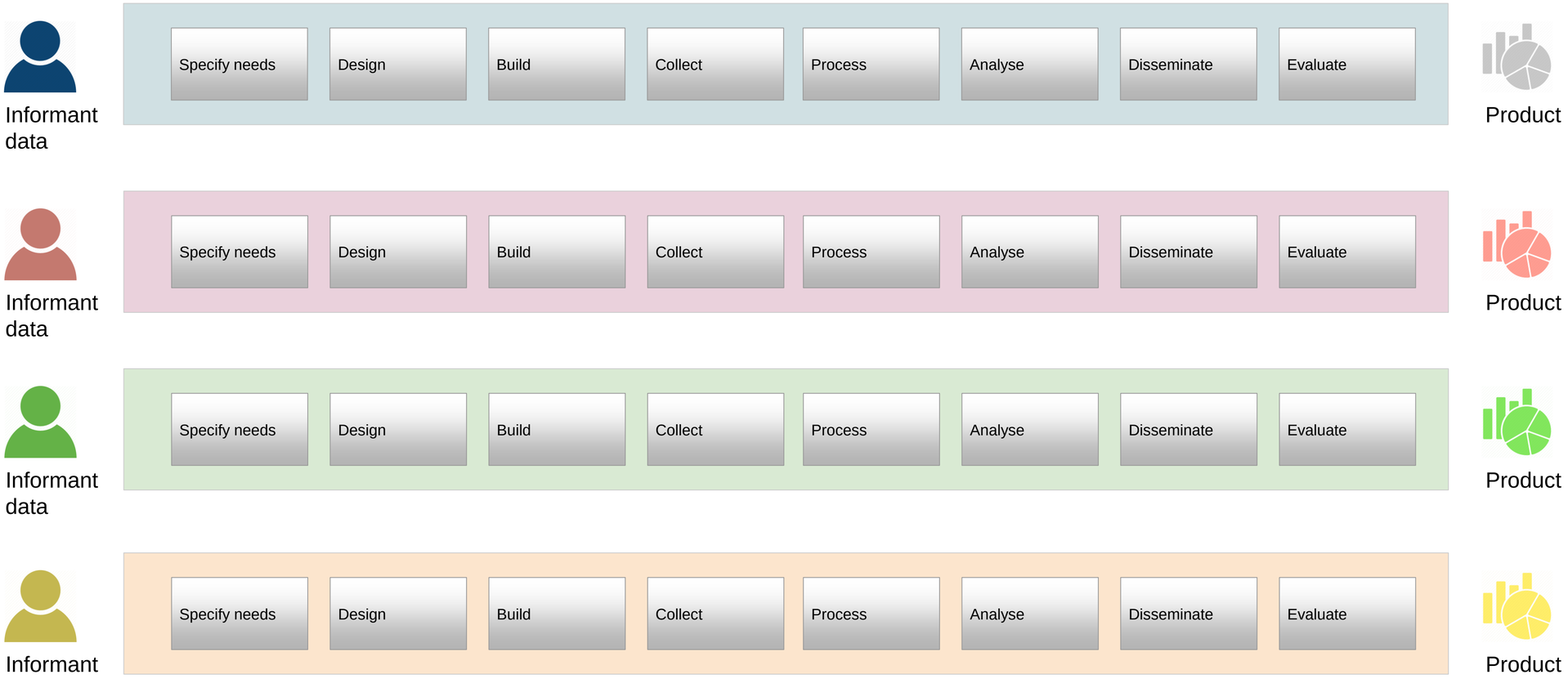
No standards for simultaneous production of statistical and geographical objects

03
Statistical Data Architecture

- Dominant Schema
- Standardized production line (life cycle).
- Statistical Data Domains
- Aggregate Data Architecture levels
- Statistical Metadata
- Opportunities

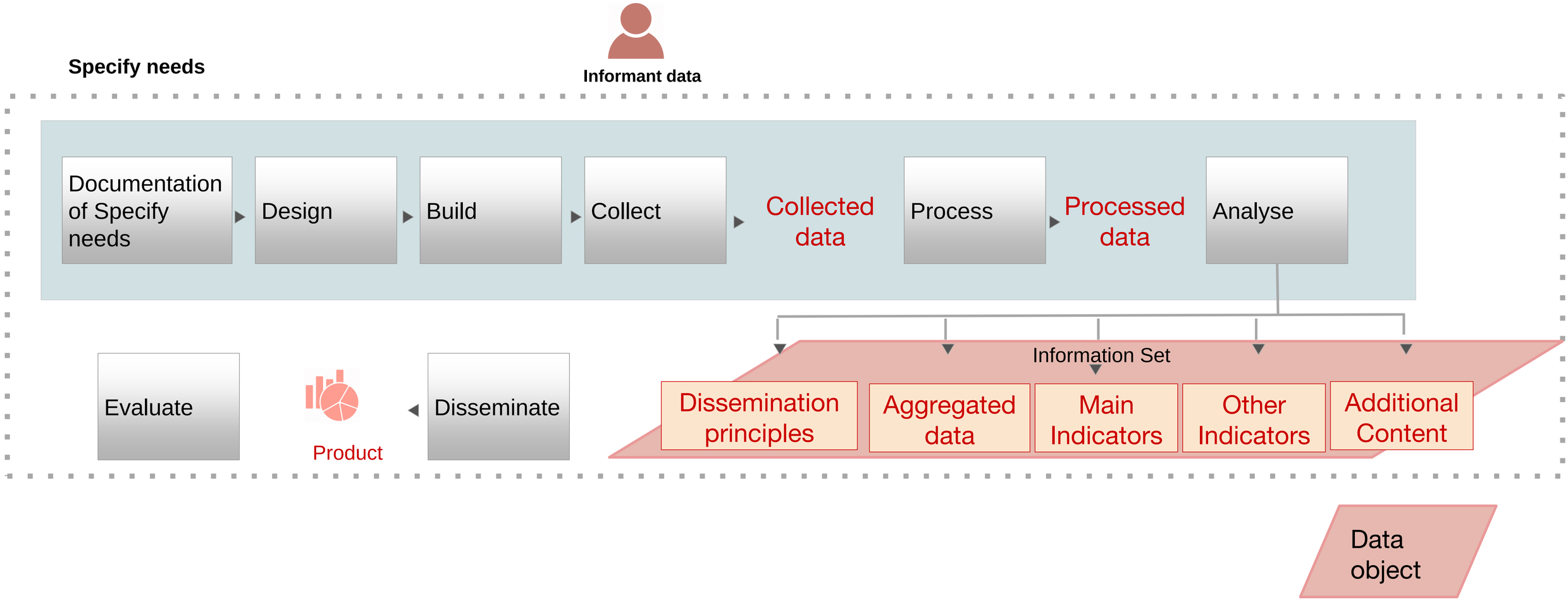
❖ 03.1

Dominant Schema

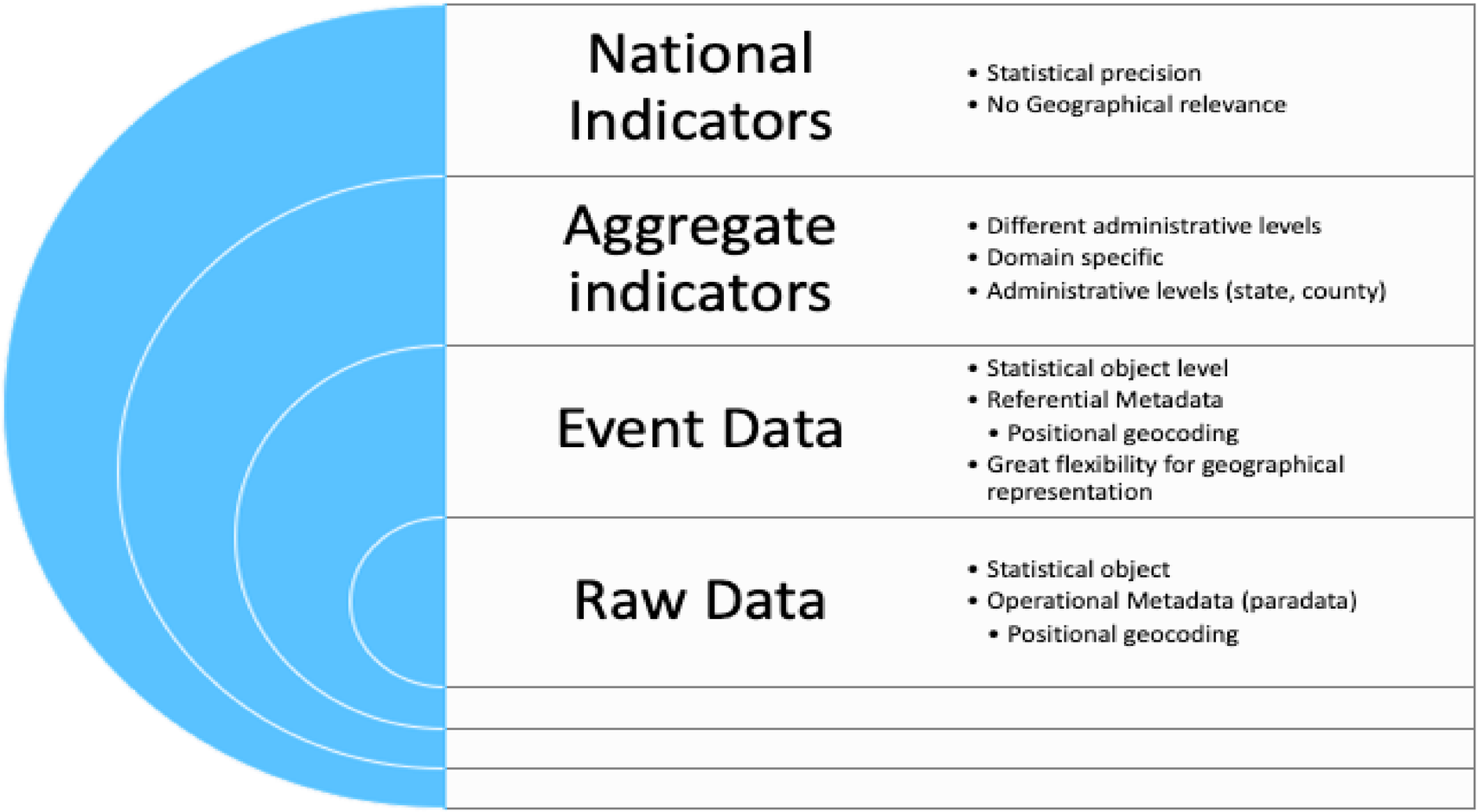


❖ 03.2

Standardized Production Line



Aggregate Data Architecture levels



❖ 03.5

Statistical Metadata

Metadata Standards

Conceptual Metadata

- Purpose
- Coverage
- Analysis unit
- Concepts
- Universe
- Variables
- Classifications
- Categories

Data collection Metadata

- Methodology
- Sampling
- Collection strategy

Processing metadata

- Entry
- Coding
- Editing
- Derivation
- Weighting

Support: Data discovery Data analysis Data distribution Data access Data availability

Opportunities with statistical information

- Despite the heterogeneity in the production of information, every collected statistical data is either georeferenced or geocoded.

- Other statistical information attributes (metadata) can be stored and managed in a geospatial infrastructure.

- Consolidation in a Geospatial Infrastructure enables the analysis of cross-discipline domains enhancing traditional statistical analysis.

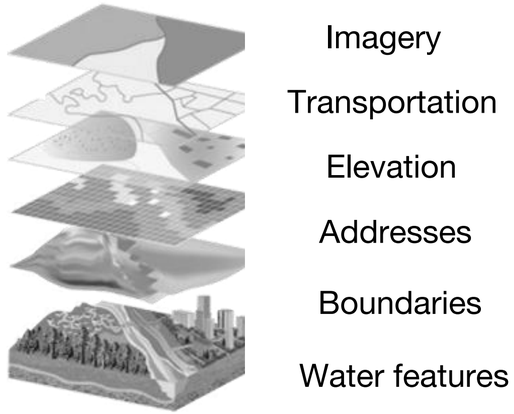
04

Geographical Data Architecture

- Geographical Information Systems (GIS)
- Geocoding vs. Georeferencing
- Schemas for Geospatial Data
- Grid-based representation
- Administrative boundaries
- Geographical Metadata

Geographic Information System (GIS)

A framework for gathering, managing, and analyzing data.



Geocoding vs. georeferencing

Geocoding

Requires the latitude and longitude of every statistical object

Substantial historical data is not geocoded

Georeferencing

Fixed polígons
Widely available
Less flexible

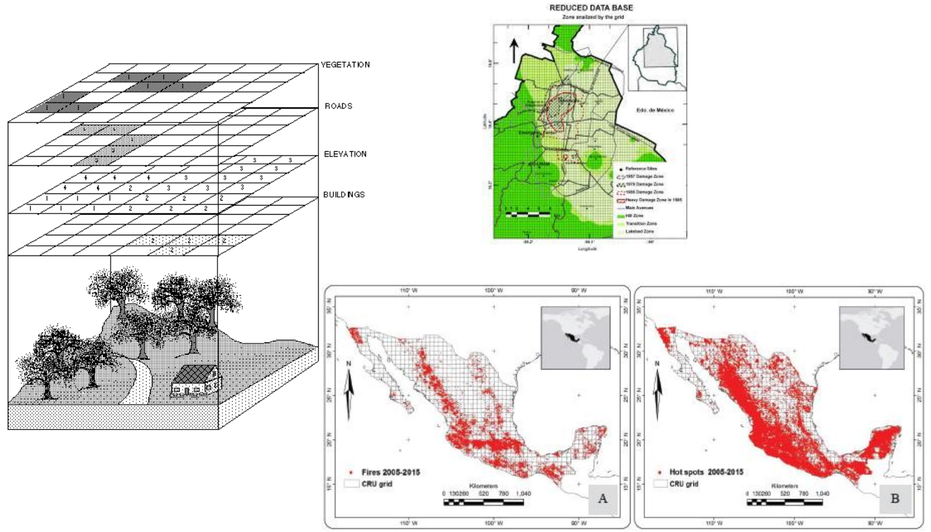
Most historical data was georeferenced

❖ 04.3

Schemas for Geospatial data

Grid-Based

Requires geocoding of every statistical observation
Some historical data was not geocoded



Administrative boundaries

Fixed polígons
Widely available
Less flexible

Grid-based representation

- Spatial units with equal size and even distribution.
- It offers flexibility in size.
- It is not population-centric.
- It can be applied across boundaries.
- Suitable for overlaying and spatial analysis.

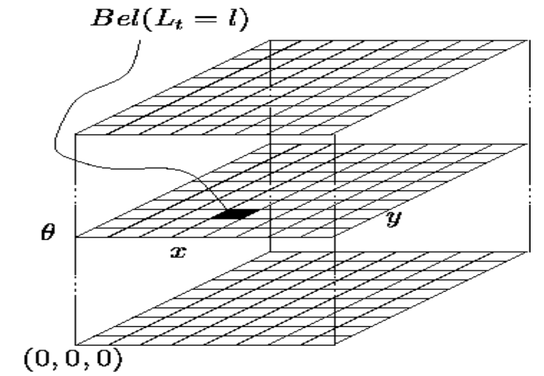
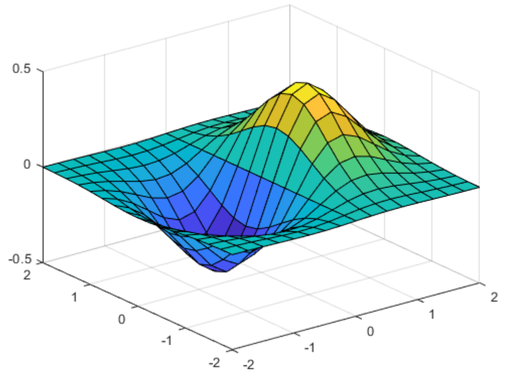


Fig. 10. Grid-based representation of the state space



❖ 04.5

Administrative Boundaries

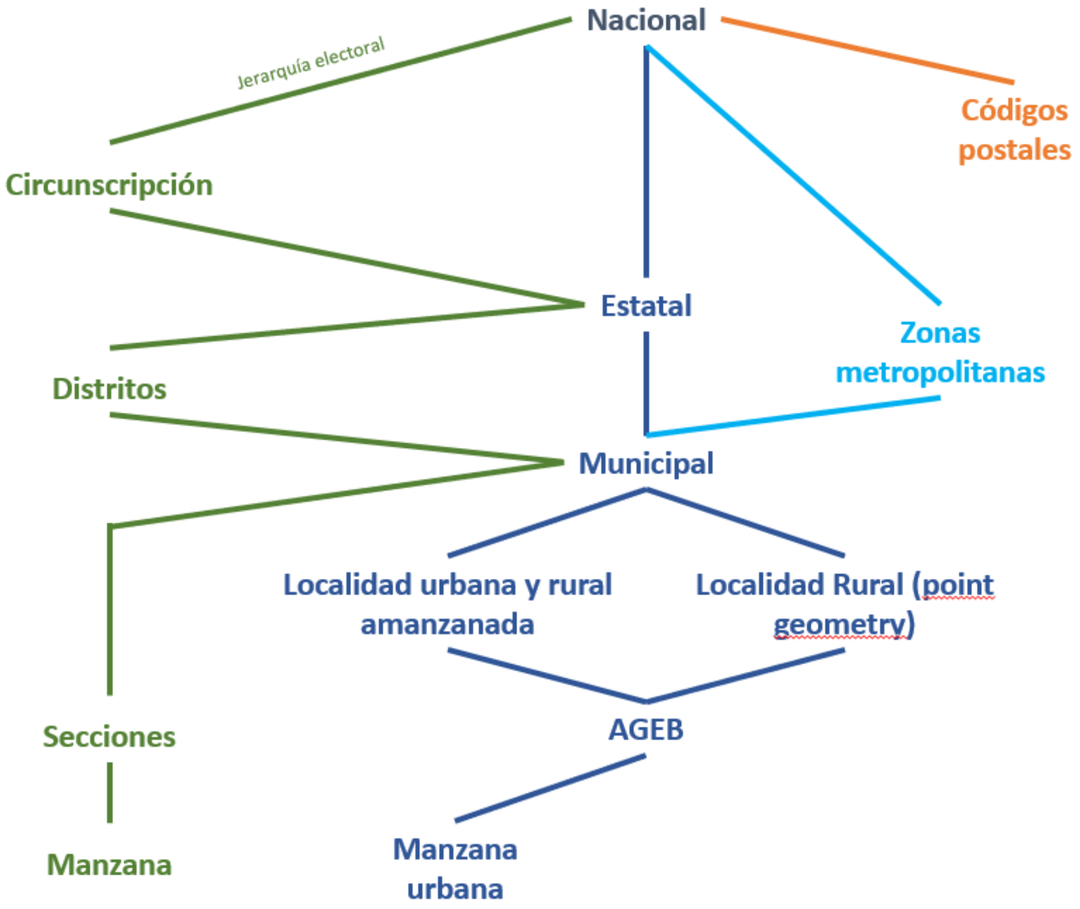
Country

State

Municipality

Region

Enumeration Area



- Delimitación de zonas metropolitanas (CONAPO, quinquenal)
- Servicio Postal Mexicano
- Marco Geoelectoral (INE, irregular)
- Marco Geoestadístico (INEGI, irregular)

Geographical Metadata

- Identification
  - Creation date, data author, contact information, source agency, map projection and coordinate system, scale, error, explanation of symbology and attributes, data dictionary, data restrictions, licensing.

- Assessment
  - Use constraints, access constraints, data quality, availability

- Access
  - On line, order, contact

05
Geostatistical Approach

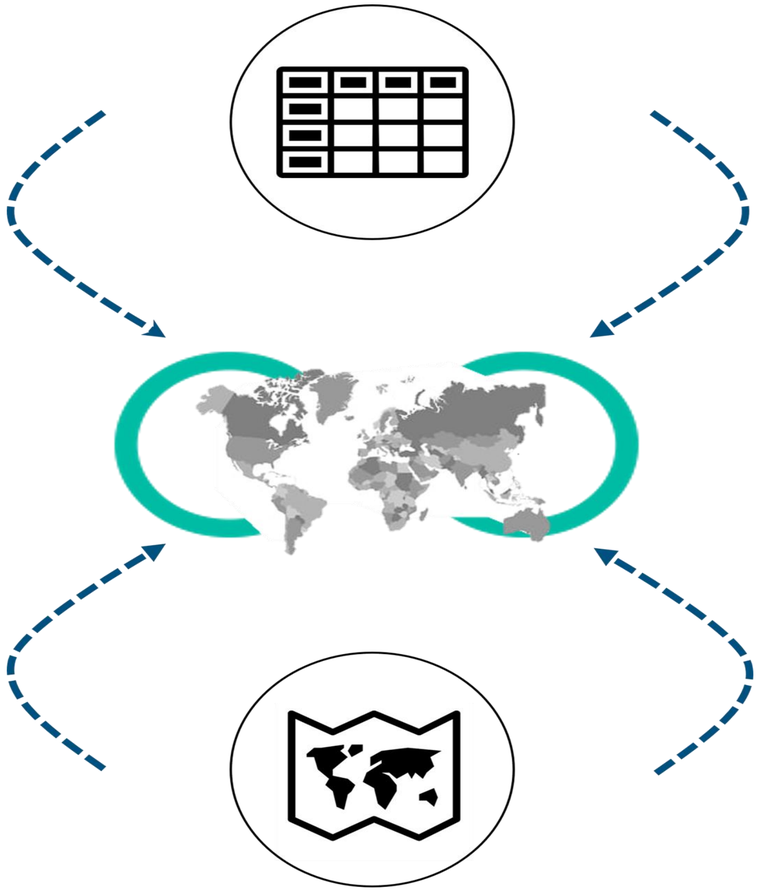
- Evolution to support of evidence-based analysis
- Geospatial Data Infrastructure
- Metadata Approach
- Implementation stages

❖ 05.1

Evolution to support evidence-based analysis

National Statistics Data

Statistical Analysis



Geostatistical National Data Architecture

Geostatistical Analysis

National Geospatial Data

Geospatial Analysis

❖ 05.2

Geostatistical Data Infrastructure

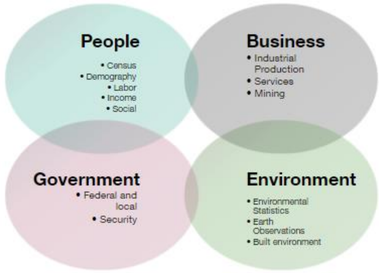
Statistical Data

Geospatial Data

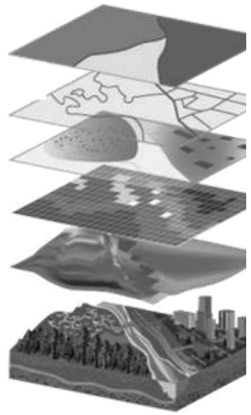
National Data Architecture

Geospatial Database

Stat Data Layers



Geo Data Layers



Imagery

Transportation

Elevation

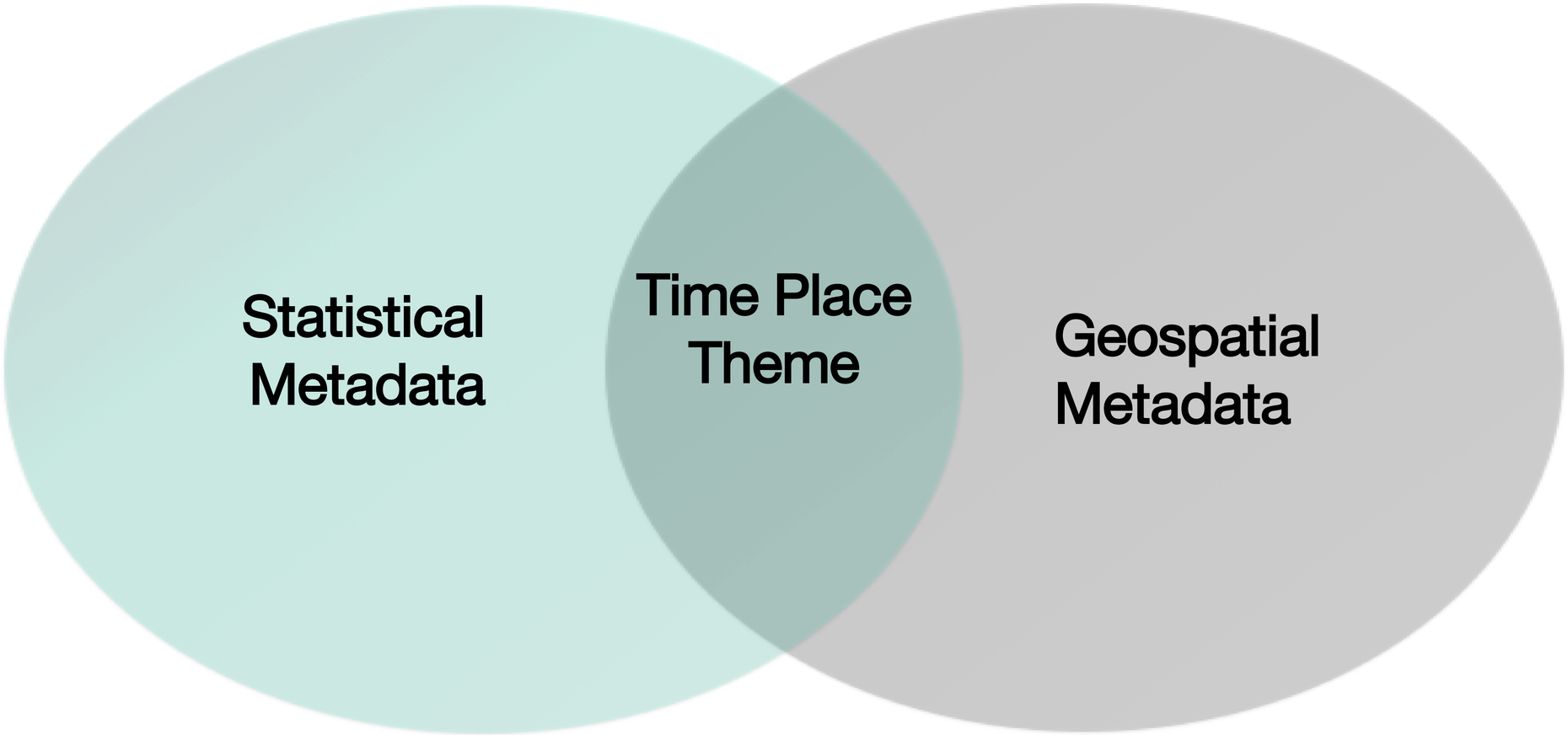
Addresses

Boundaries

Water features

❖ 05.3

Metadata Approach



05
Implementation Stages

❖ 05.4

Implementation Stages

0

Silos

Statistical Information

Informant data

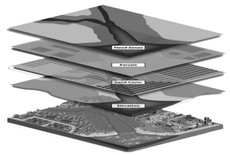
Informant data

Product

Product

Ad Hoc consolidation

Data Warehouse



Geographic Information

Informant data

Informant data

Product

Product

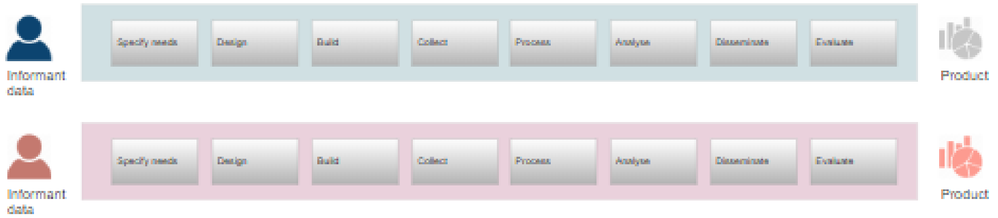
Geospatial Database

Implementation Stages

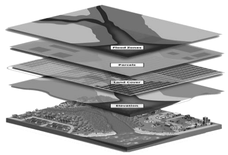
1

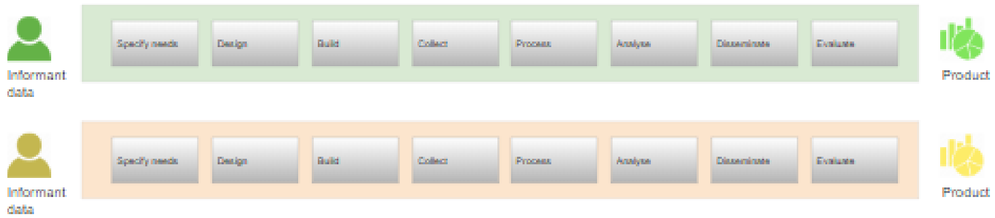
Standardization of processes and deliverables

Statistical Information



Geographic Information



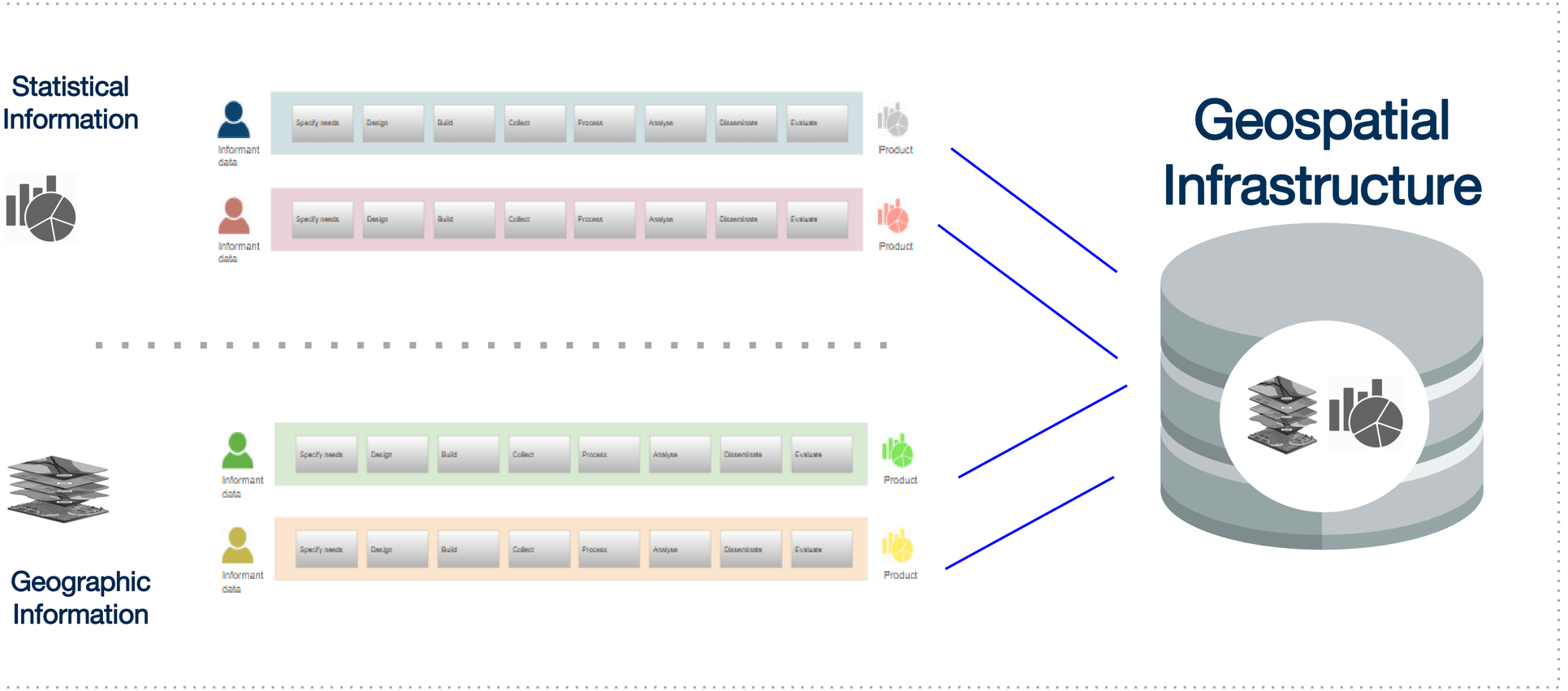


Both with standardized:
1)Output
2)Metadata
3)Paradata

Implementation Stages

2

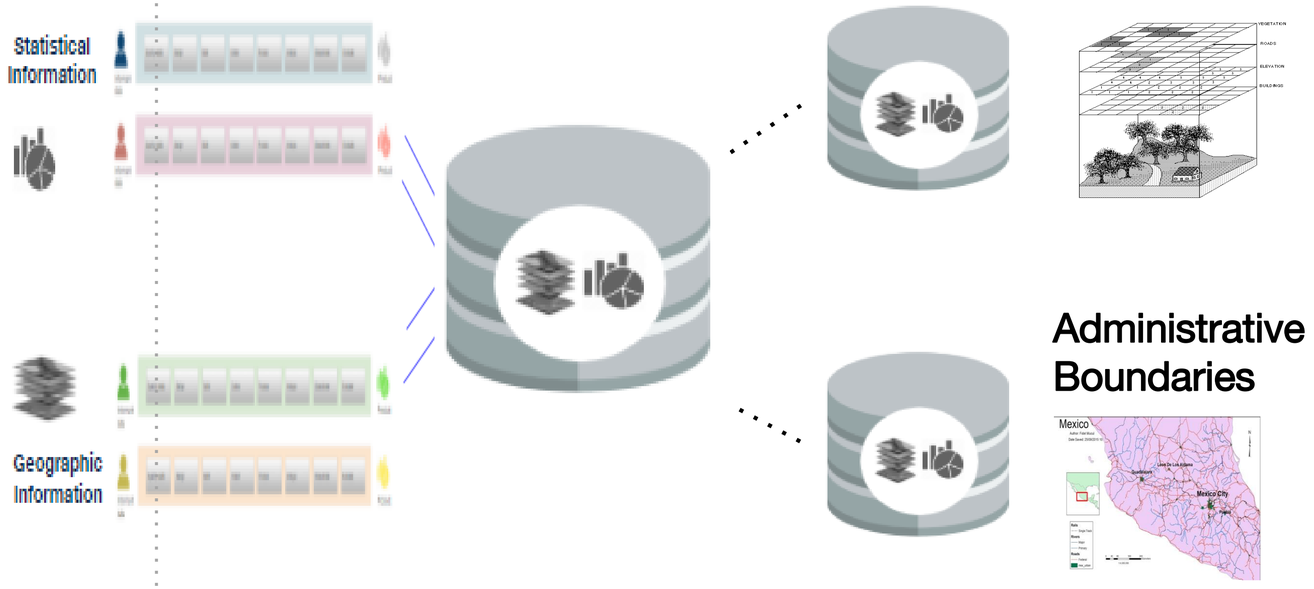
Integration into a Geospatial Infrastructure



Implementation Stages

3

Geospatial analysis



Representation

Mapping

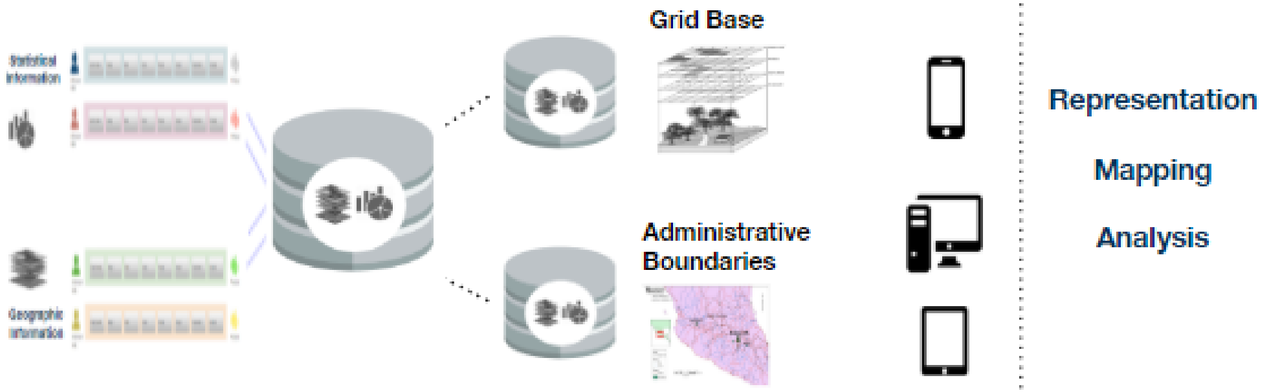
Analysis

❖ 05.4

Implementation Stages

4

Geospatial Network



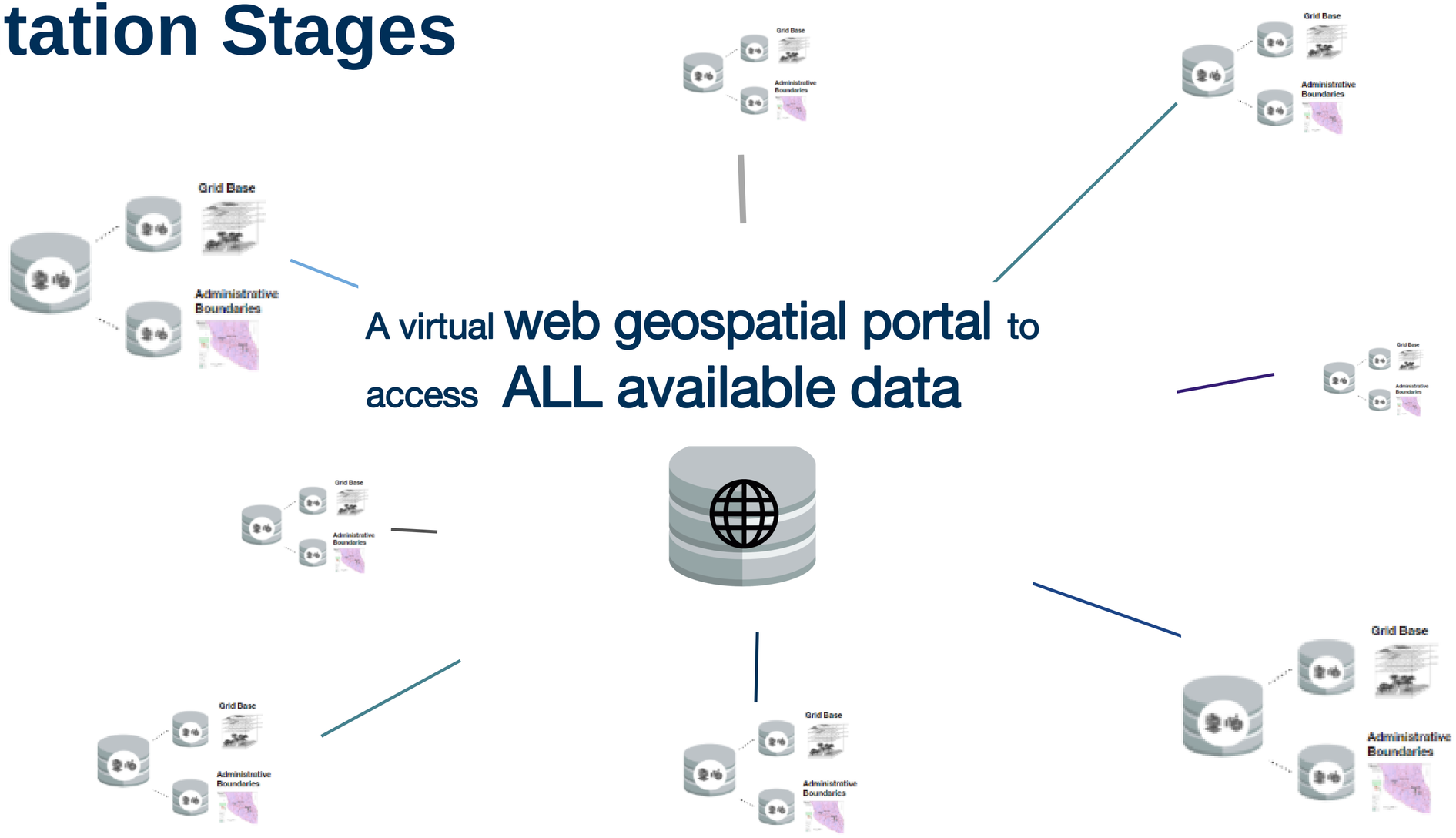
Implement the schema in every Federal Statistical Institution...

❖ 05.4

Implementation Stages

4

Geospatial Network



Summary

Statistical production has been supporting traditional evidence-based policymaking.

Most of the statistical production has been taking place within independent silos.

Metadata standards can enable common interfaces, but they do not provide a framework to define shared storage and use.

GIS provide a pool to concentrate statistical and geographical data into a common Geospatial Infrastructure.

The existence of standardized statistical and geographical metadata allows the consolidation of data and enhances the capabilities for representation, mapping and geospatial analysis.

Conociendo México

01 800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

INEGI Informa