IBGE

Instituto Brasileiro de Geografia e Estatística

Logistics and Means of Collecting Geo-Coded Statistical Data

Luiz Paulo Souto Fortes, PhD

Thematic Seminar on the 'Integration of Statistical and Geospatial Information: Issues and Challenges'

Seoul, October 27, 2011

Integration of statistical and geographic information

The close integration of <u>geographic information</u> in <u>statistical applications</u> yields large benefits to NSOs as it <u>reduces the cost and time</u> required to collect, compile and distribute information, and leads to a greater number of services and a much wider use of statistical information, thereby considerably increasing the return on investment in data collection.

IBGE = Geographic Information + Statistical Information

Aplications: Censuses/Spatial Data Infrastructures

Integration of statistical and geographic information

Example of application: after a disaster, satellite images can provide information on affect areas and, based on statistical related data, rescue and relief can be planned

Integrated and organized information

Agility in helping people – lives saved!

Evolution of Geotechnologies
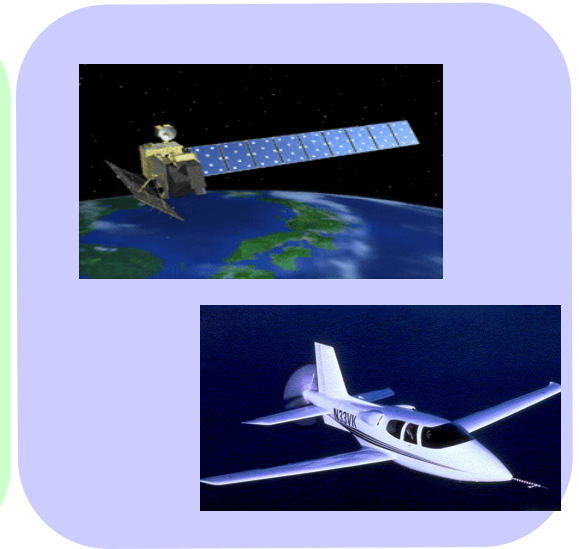
Evolution of Geotechnologies

IBGE

Handheld Devices (PDA)

GIS/Web
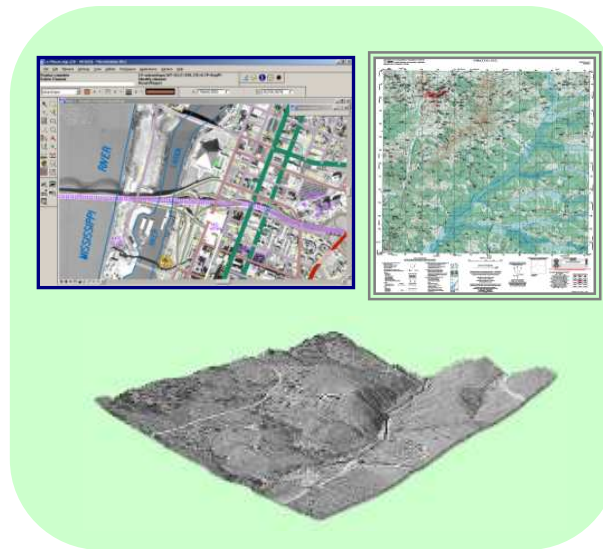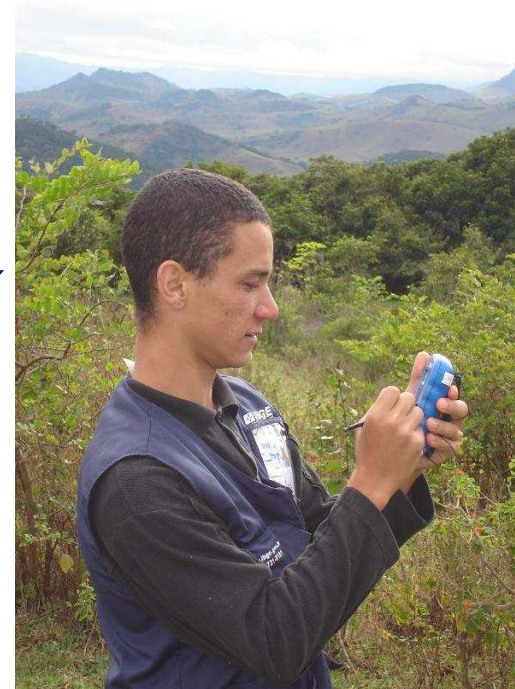
Low Cost Images

Increasing use at IBGE for both Geosciences and Statistics

Brazilian Geospatial Statistics

IBGE

Innovations in data collection

- First time PDA has been used

- Determination of agricultural establishments´ coordinates by PDA/GPS

- After collecting coordinates, the system automatically displayed the corresponding questionnaire

- 82,000 devices
- 68,000 enumerators
- 11,000 supervisors
- 162,770 enumeration sectors
- More than 5 million agricultural holdings

Innovations in data collection

Advantages

- Immediate **quality control** at the moment of data typing

- **Filling of all compulsory** items

- **Control of data filling** by automatic jumps in the form

- **Direct transmission** of collected data to central system

- **Dispensing the transportation** of high volumes of paper questionnaires and their handling in data collection centers

- **Location of the units surveyed** in rural areas (agricultural, health and educational establishments)

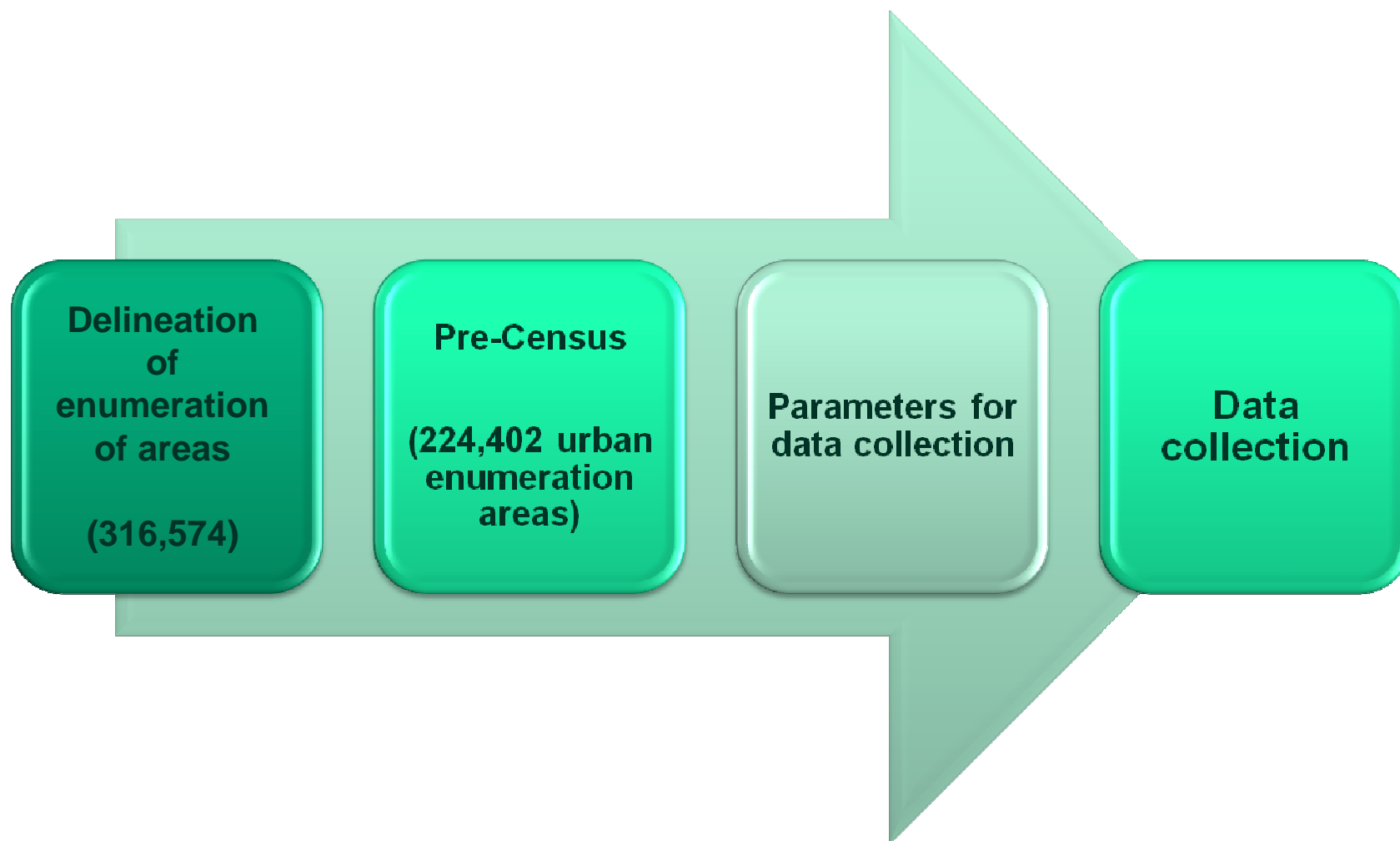## More than 5 million agricultural holdings georeferenced
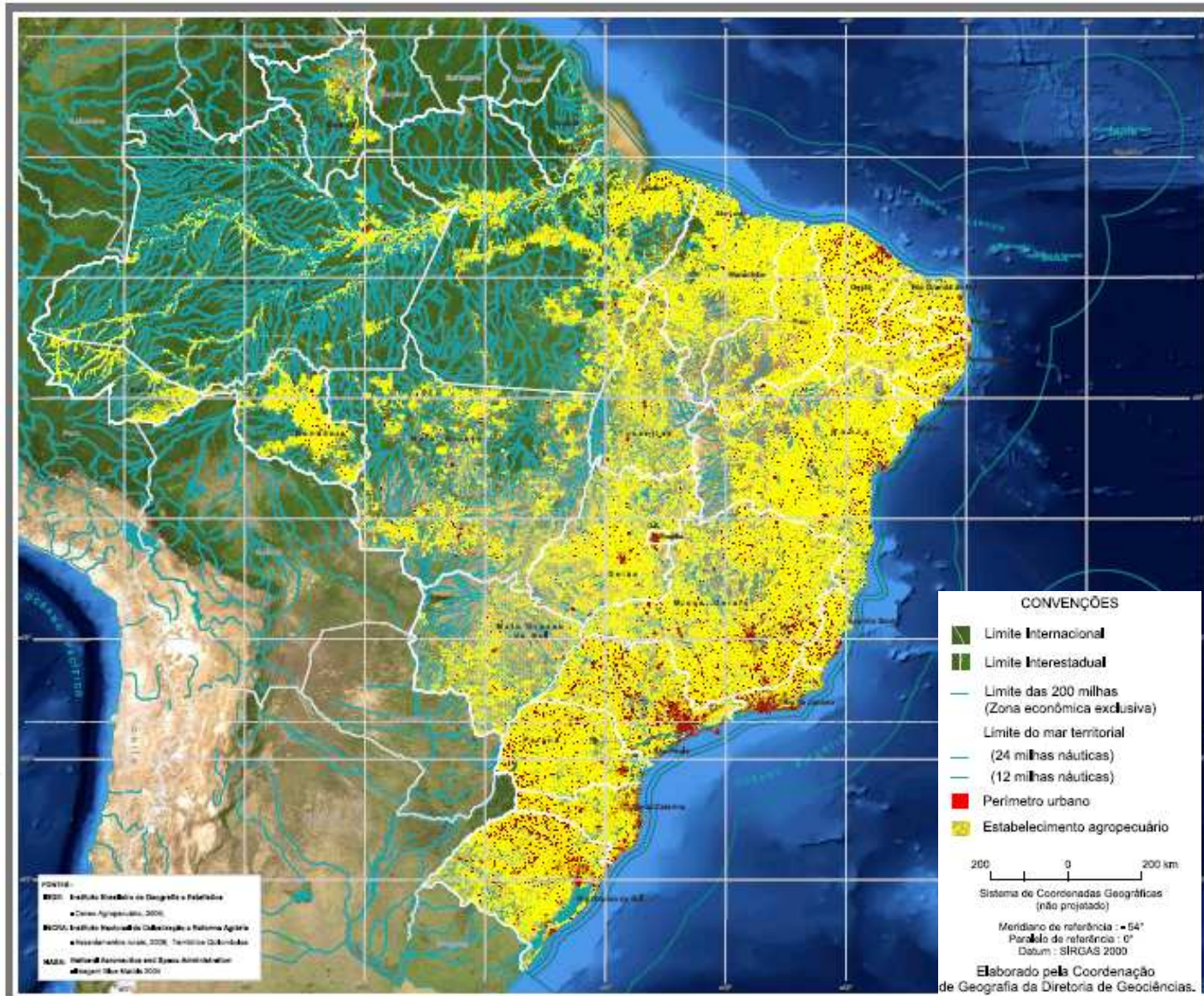
- Population Census produces information on the main characteristics of people and households in each of the 5,565 Brazilian municipalities

- Population Census encompassed approximately 67 million housing units

2010 Population Census – Main steps

IBGE

Delineation of enumeration of areas

(316,574)

Pre-Census

(224,402 urban enumeration areas)

Parameters for data collection

Data collection

# Delineation of 316,574 Enumeration Areas (EA)   IBGE

- For 2010 Census, IBGE developed a Census Mapping Project to integrate Municipalities' urban and rural areas in a full digital format

- Massive use of aerial and orbital imagery

- Characteristics of 2010 Census Enumeration Areas
  - Each enumeration area is the working field of each enumerator
  - EA is a **continuous area** located at rural or urban places
  - Adherent to Boundaries of municipalities, districts and sub-districts are
  - Urban EA: covers 300/350 housing units
  - Rural EA: covers 150 agricultural holdings (maximum 500 km^2)

Enumeration Areas: Urban

Enumeration Areas: Rural
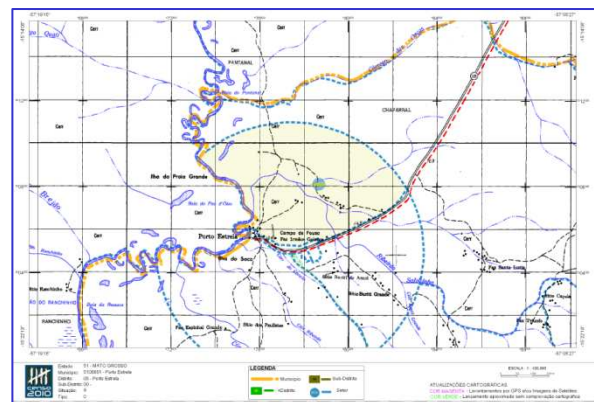
Integrated Census Mapping to support 2010 Census

• The "National Address File for Statistical Purposes – CNEFE" file associated to blocks and block-face in urban enumeration areas

• Geometrical integration of urban and rural limits for each Municipality
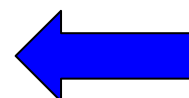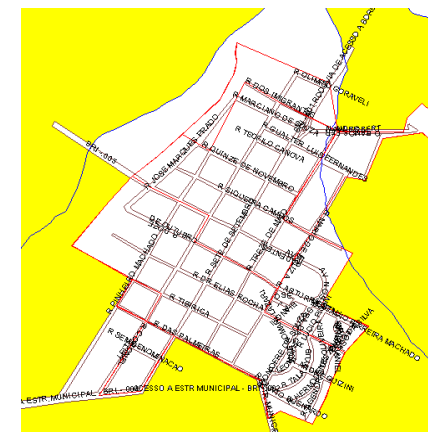
SISMAP – Brazilian System for Census Mapping

• IT tool developed for municipal mapping in a seamless spatial database

• Input data from several sources - vector and imagery data, like GIS, GPS, satellite imagery, digital and aerial photography
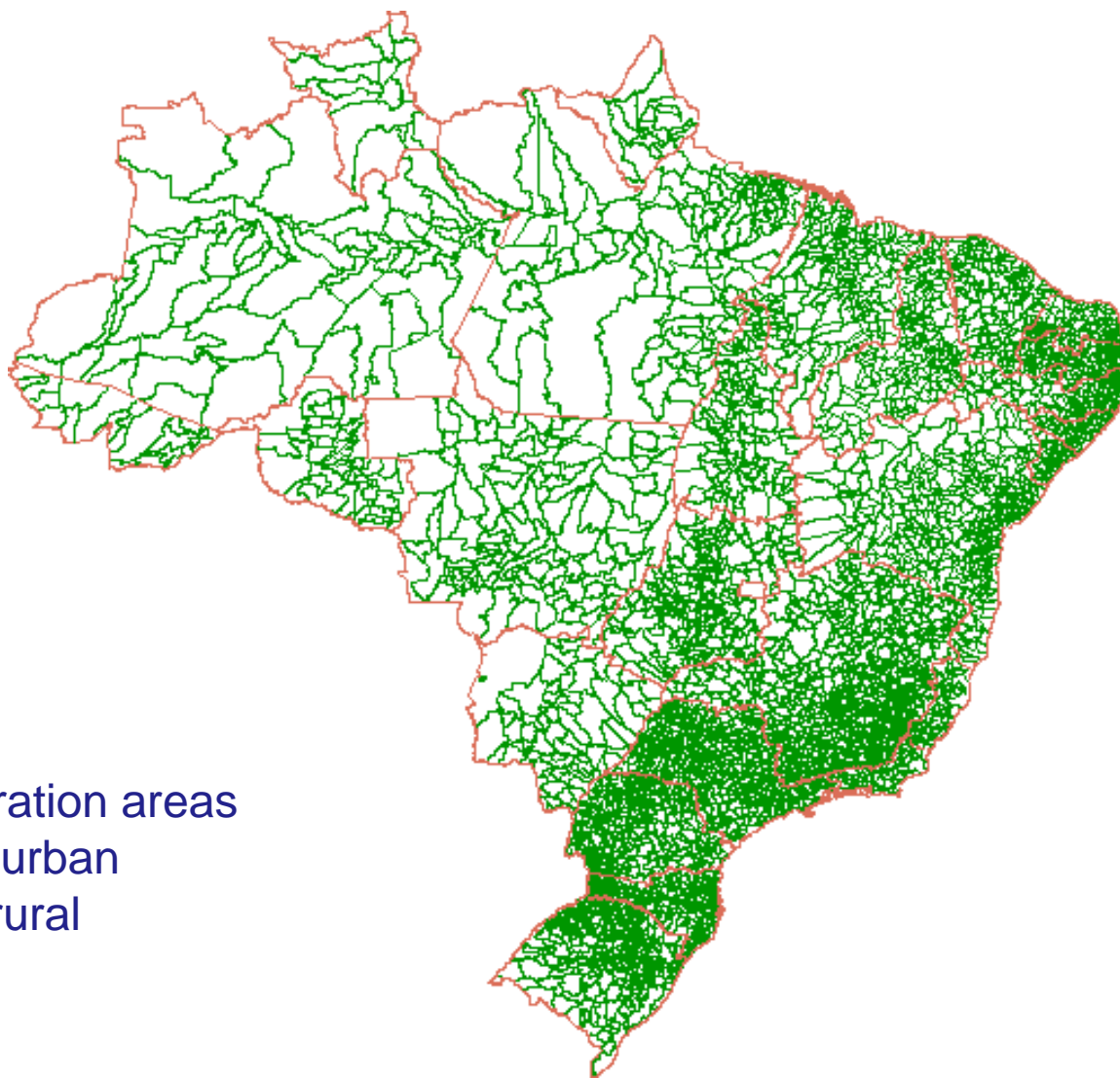
2010 Population Census

IBGE

Census Mapping

27 Federation Units
(States)

5,565 Municipalities

10,283 Districts

662 Subdistricts/R.A.

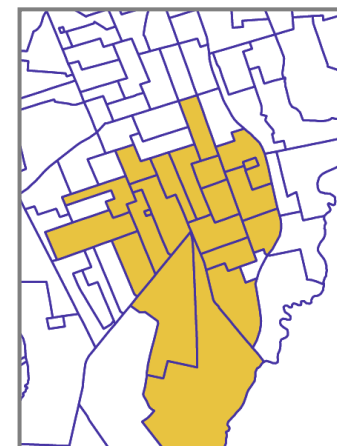316,574 Enumeration areas
240,382 urban
76,192 rural

IBGE

New ways of dissemination

Up to 2007 Census

Enumeration areas corresponded to the minimum area for statistical data analysis
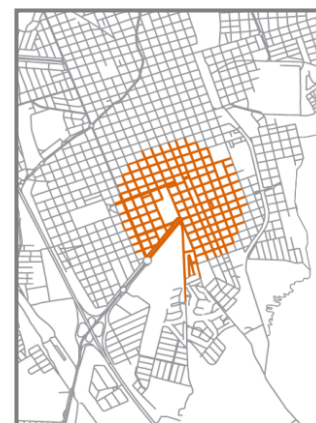
Area of interest

Selected area

As of 2010 Census

Block-faces correspond to the minimum spatial area for statistical data analysis

Need of dissemination rules to fulfill information confidentiality

Selected area

IBGE

Pre-Census (from March to June 2010)

✓ **Review of urban Enumeration Areas performed by 24,000 supervisors hired for the Census:**
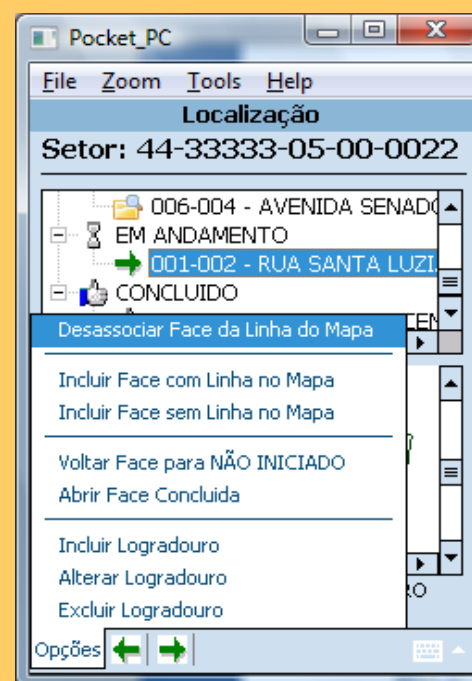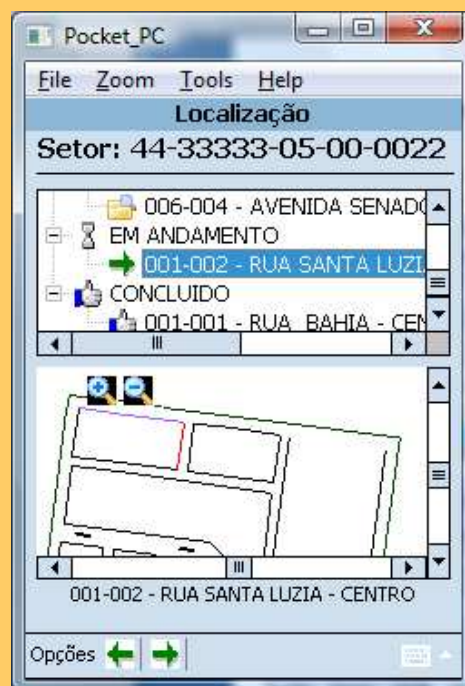
- Loading enumeration area maps and the list of streets and block-faces associated with the map
- Review/updating digital maps with PDA
- Review/updating association of addresses to digital maps
- Work unit: block-face

✓ **Duties of Supervisors:**

- For each street: confirm, delete, include, change the name and/or change name's spelling
- For each block-face: include, exclude, confirm or classify as "NAR" (no addresses in the block-face) and collect surrounding characteristics (electricity, sanitation, security, etc.)
- Tracking the shape of block-faces not represented on the map with GPS

IBGE

Pre-Census Application

- Personal Digital Assistant (PDA) application:
  - The application was developed in Windows Mobile, using SQLServer CE 2005, Framework 2.0 and Geopad
  - Loaded in 30,000 PDAs

Pre-Census Application

PDA application: Some forms to collect data

IBGE

- Based on information of the Enumeration Areas and data collected during Pre-Census, some acceptance parameters were created for questionnaires collected by the enumerators during the Census which started in 1st August 2010

- These parameters were loaded into the data collection controlling and monitoring IT System used by supervisors to assess the work of Census enumerators

- Address list file updated by Pre-Census was used for enumerators routing and coverage control

- Data Quality Control was carried out for each Enumeration Area, including ratio of species of housing units, average number of persons per housing unit, ratio of people by gender, age, etc.

Questionnaires

All **67 million occupied housing units** were surveyed to collect characteristics of the dwellings and people living there

IBGE used two questionnaires:

1. **short form**, with 24 questions for housing unit and 13 for people (**37 in total**)

2. **long form** (**sample**) with 38 questions for housing unit and 69 for people (including same questions of the short form) (**107 in total**)

The choice of questionnaire was done **randomly** in the PDA during the Census collection

- The average sampling rate is 11%. The rates depend on the number of inhabitants of the municipality and varies from 5% (more then 500,000) to 50% (less then 2,500 inhabitants).

- **190,000 PDAs were used by enumerators to collect data**

- **Data confidentiality –** enumerators did not have access to data stored on the the PDA. After ending an interview, data were encrypted and could only be transferred to IBGE through a secure network

IBGE

- **Data collection:**
  - Running on the PDA in order to record the interviews, Internet token and update the List of Addresses

- **Data collection/Internet**
  - WEB application to fill out the forms by households

- **Supervision:**
  - Running on the PDA for quality control of the enumerators work

- **SIGPC - Management Information System at the Data Collection Local Station**
  - Running on notebooks, supported all decentralized operations, including administrative and operational tasks and the communication with the Central System

Data Collection Questionnaires Application

System developed in Windows Mobile and TotalCross

Data Collection Supervision Application

IBGE

- Developed in Windows Mobile, SQLServer CE 2005, Framework 2.0, Geopad
- Management Indicators and results of the previous visits

SIGPC - Management Information System at the Data Collection

Local Station

✓ Local application at the data collection station running on notebooks under Windows 7

✓ Essential tool for the administrative and operational organization and quality control of data collection, including:

- registration of temporary work force

- task assignment to staff

- loading the application and Enumeration Area data into PDA

- receiving collected data from the PDA

- communication with Central Datacenter

- reporting of quantitative and qualitative evaluation of data collected

Software Applications at the Central Datacenter

IBGE

✓ SIGC - Data Collection
  Management Indicators System

- Monitoring of data collection
  (coverage, quality and time of
  collecting in each
  Enumeration Area)

- WEB application running in
  the main Datacenter
  generated management
  reports, summaries,
  indicators and cartograms

✓ Database system

- Specialized data model to
  store the summaries of the
  data collection, data of
  supervision, administrative
  data and List of Address File
  data

ICT for Brazilian Census Data Collection
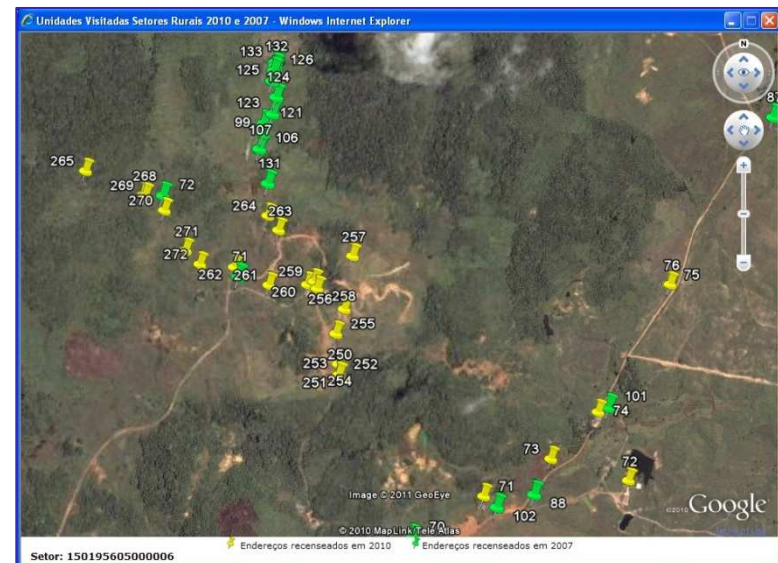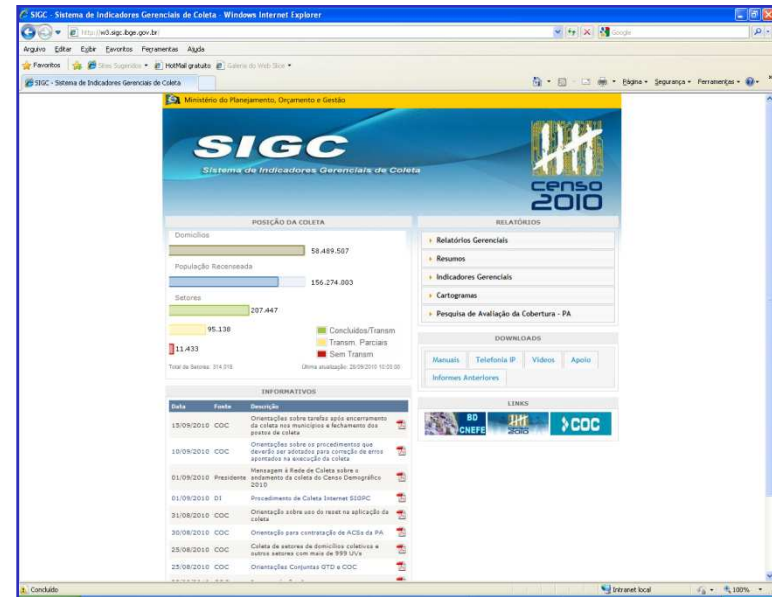
Information and Communication Technology for 2010 Census was developed by staff team of IBGE

✓ Data Collection

  - 7,000 Census Data Collection Local Stations (Census Local Offices), spread over 5,565 municipalities in Brazil
  - 220,000 PDAs for supervision and data collection

✓ Regional Coordination

 - 27 Regional (States) Coordinators for each subject (technical; operational; administrative and **ICT)**
  - 220 Operational Area Coordinators
  - 1,281 Subarea Coordinators
  - 6,000 Municipality Coordinators
  - 220 ICT Sub-Area Coordinators to deal with all ICT local subjects at the Data Collection Local Stations

ICT for Brazilian Census Data Collection

- ✓ All data collection made with Personal Digital Assistant - PDA (there was no paper questionnaire).

- ✓ Each PDA was equiped with Enumerator Area Map, Questionnaires (short and long) and List of Addresses

- ✓ 7,000 Census Data Collection Stations equipped with notebooks
    - Census Data Collection Stations were equipped to operate autonomously, connected or not, to the Internet
    - Interaction of PDAs was made solely on the notebooks of the Collection Station, regardless of a network of synchronous communication with the Central Processing
    - In Collection Stations without Internet connection, communication with the Central Processing was done through USB drives (flash drives) physically carrying lots of information to some other point of Internet access
    - All ICT technical assistance (software, hardware and communication) were provided by 220 ICT Area Coordinators.

ICT model in the 2010 Census

ICT in Data Collection
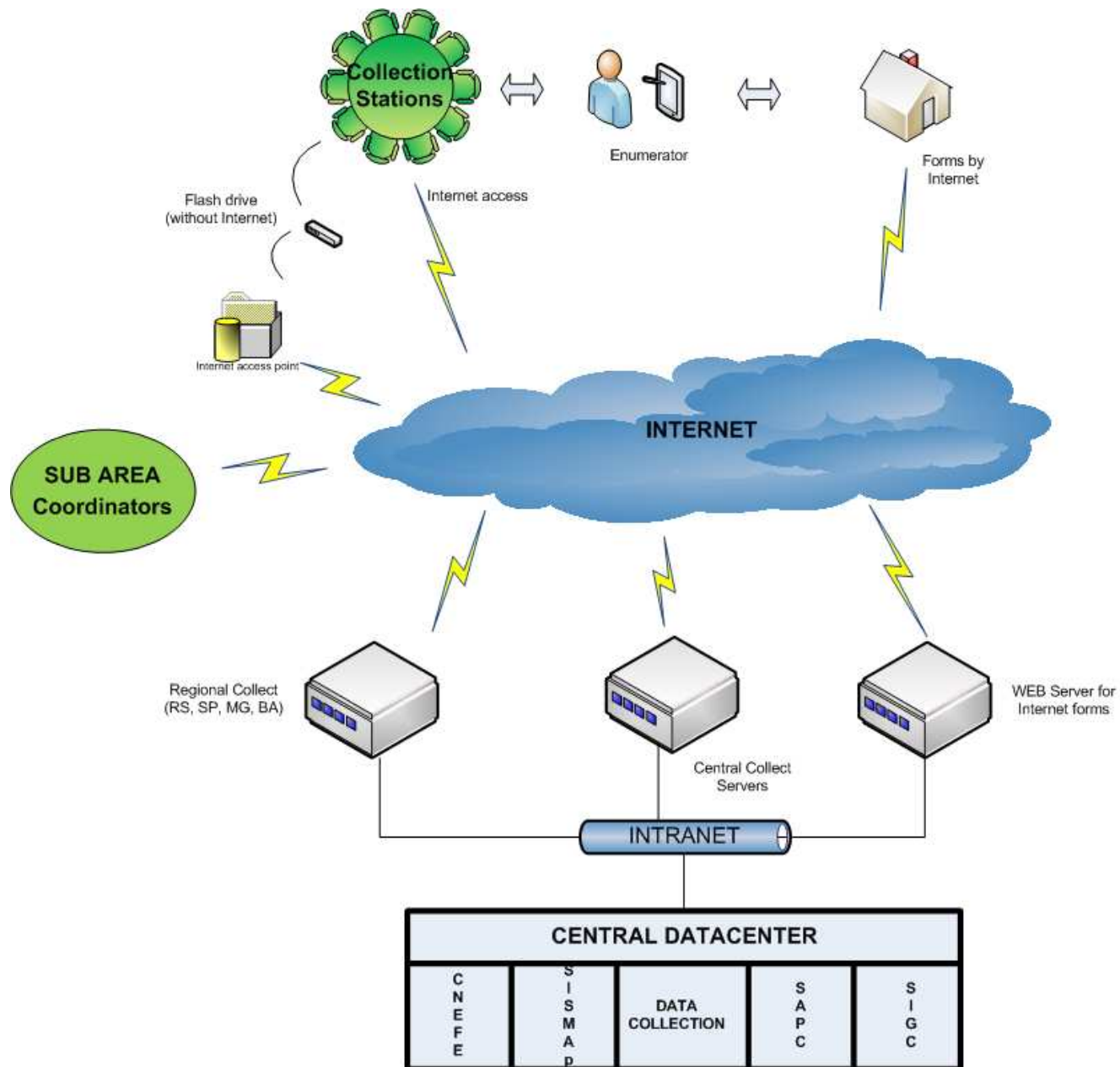
- ✓ 150,000 LG-750Q, smartphones locked (Phone, 3G) to be used solely as data collection equipment using programs certified by IBGE
- ✓ 70,000 PDA MIO-P550B (already used in 2007)

Data Collection Equipments

IBGE

PDA MIO P550B	LG GM-750Q
Number: 70,000	Number: 150,000
Processor: 400 MHz	Processor: 528 MHz
RAM: 64 MB	RAM: 512 MB
ROM: 128 MB	ROM: 512 MB
Bluetooth	Bluetooth
WiFi (200 eq.)	WiFi
Battery: lítio-ion (6 hs)	Battery: lí-ion 1000mAh (10 hs)
GPS	GPS
Slot MMC SD (2/4 GB)	Slot MMC micro-SD (2/4 GB)
Screen: touchscreen 3.5' , 320 x 240, 65k	Screen: TFT touchscreen 3.0' , 240 x 400, 65k
1 x USB	1 x USB
Windows Mobile 5.0	Windows Mobile 6.5

✓ Notebook (8,687 units)

- ▪ Blocked by "Content Advisor"

✓ Router: 1 per Local Office (7,000 WiFi routers)

✓ Printer: 1 per Local Office

✓ Internet (where possible) or flash drive for physical transport of information to another place with Internet access

✓ Supervisor and Enumerator: 1 PDA - MIO or LG

Internet
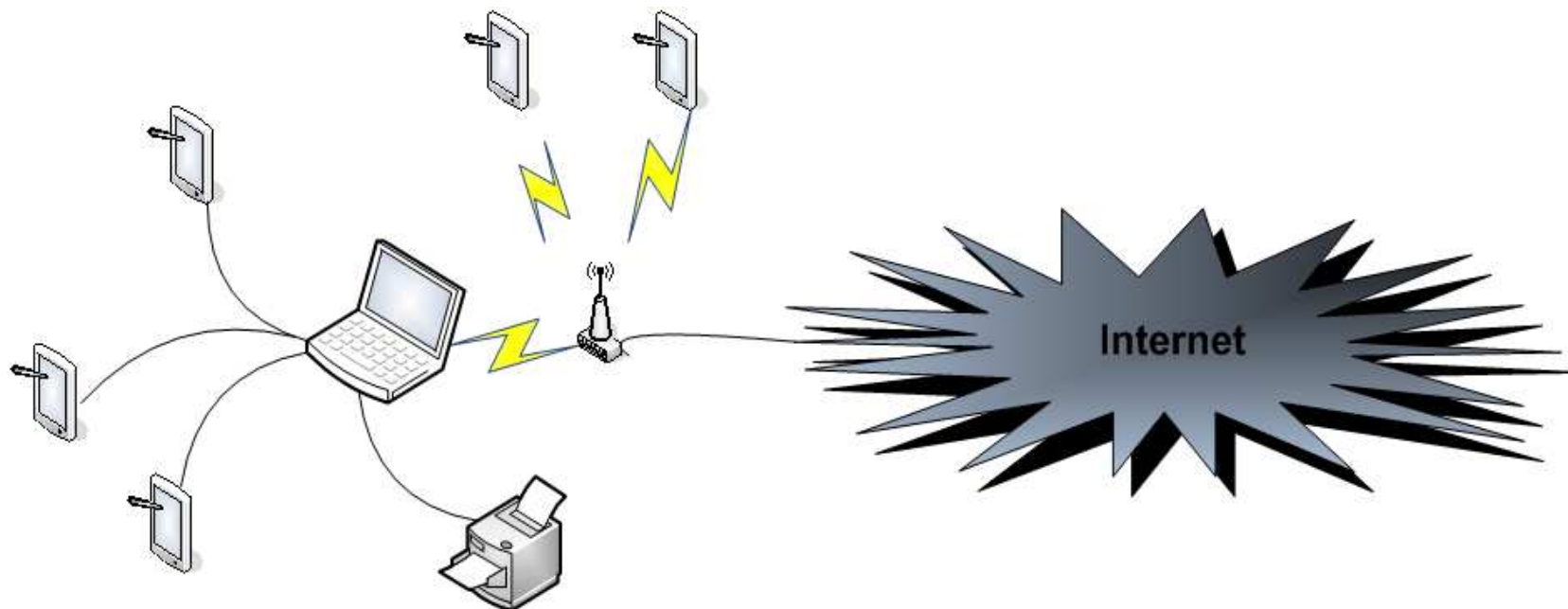
The 7,000 Collection Stations were supported and supervised by 1,281 Subarea Coordinators

One of the duties of these coordinators was to visit Local Offices under their supervision to provide quality control of collected data and organization

Equipments:

• Notebook: 1 per coordinator
    (1,300 notebooks)
• Desktop computer: 1 per Sub Area
• Router (with VoIP): 1 per office
• Multifunctional printer: 1 per office
• Broadband Internet
• 3G Modem: 1 per coordinator

Communication Resources

IBGE

- ✓ 2 links of 155 Mb/s for the exclusive operation of the Census 2010 in Rio de Janeiro (Headquarter)
- ✓ 2 links of 20 Mb/s in São Paulo and Minas Gerais
- ✓ 2 links of 12 Mb/s in Bahia and Rio Grande do Sul

- ✓ Subarea Coordination and Census Collection Stations:
  - 3,783 mini modems 2G/3G
  - 139 antennas VSAT (Hughes)
  - 40 antennas BGAN (Tesacom)
  - 4,000 ADSL connections (leased, or by courtesy of others)
  - 1,000 Radio connections
  - 300 other types of connections
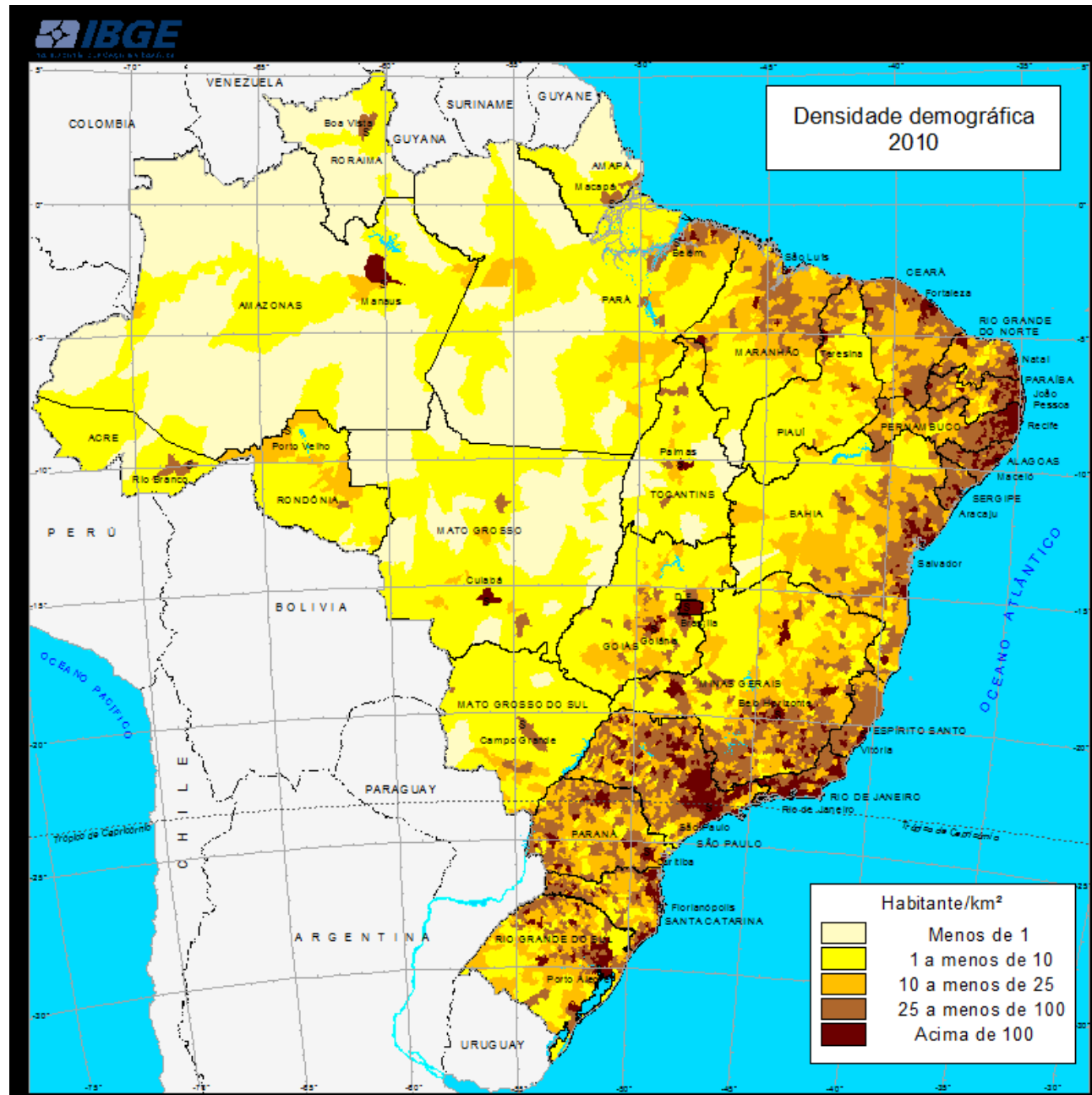
Datacenter for the 2010 Census

Data volumes

IBGE
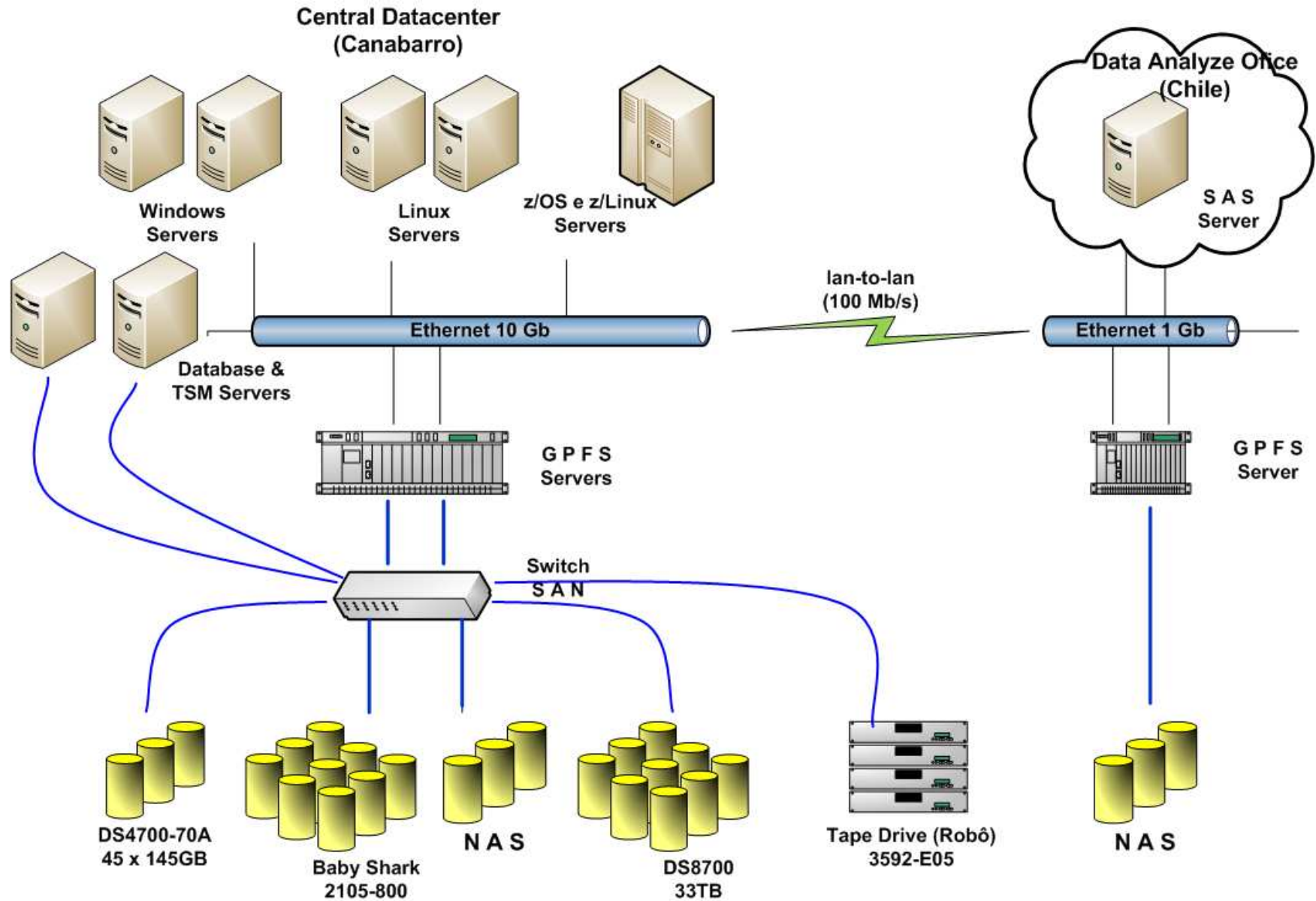
		25,029,015 (files)	3,069,097.91 MB (transmitted data)
TRANSMISSIONS			
Pre Census	6,818 stations	131,303	20,957.27 MB
Data Collection	314,017 E.A.	10,347,094	2,783,161.34 MB
Summary	6,831 stations	540,016	10,340.00 MB
Supervision	313,972 E.A.	13,658,485	251,090.00 MB
Post Census	1,921 stations	9,222	812.23 MB
Administrative	6,809 stations	312,816	2,680.00 MB
Internet forms	2,290 stations	30,079	57.07 MB
CENTRAL DATABASE		281 (tables)	397,517.00 MB (data store)

Demographic Density - 2010

Densidade demográfica
2010

Habitante/km²

	Menos de 1
	1 a menos de 10
	10 a menos de 25
	25 a menos de 100
	Acima de 100

Thank you!

Luiz Paulo Souto Fortes, PhD
luiz.fortes@ibge.gov.br