Measuring Internet access using Mobile Phone Data

UN Big Data Regional Hub in Brazil

Why use Mobile Phone Data?

General context of NSOs

- To produce new data or complement existing data
- To produce data with high frequency of updates and with greater spatial disaggregation
- To lessen the burden of respondents

Information Society context

- The number of households with computers is decreasing and at the same time the number of internet accesses via cell phones is increasing
- During the pandemic and pos pandemic periods, it is essential to know who is and who is not connected to the internet

Case study

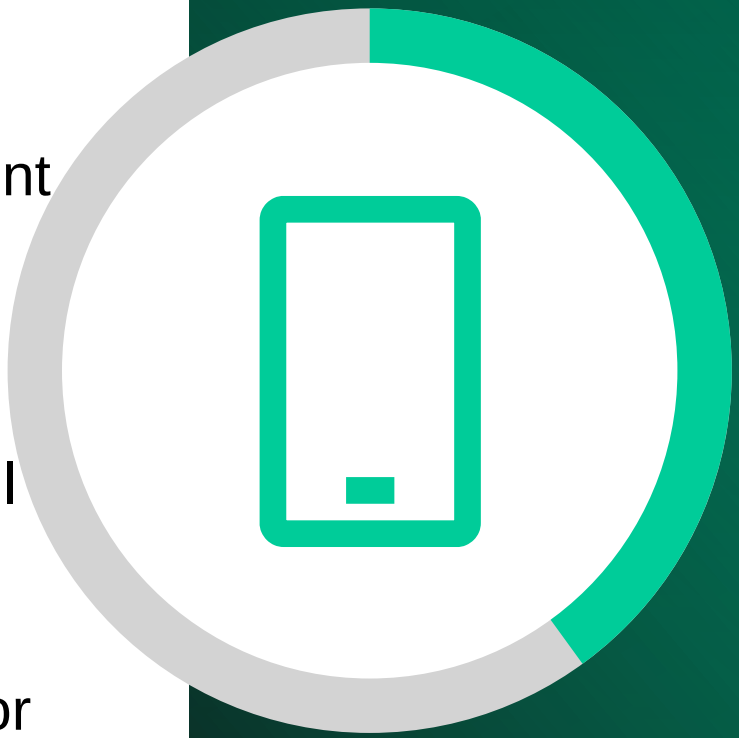
- 24 municipalities in Rio de Janeiro state ~ Metropolitan Region
- 13 million inhabitant → second most populated area of the country
- SDG Indicator 17.8.1: Proportion of people using the internet
- Reference: household survey – PNAD Contínua

Mobile Phone Data

- One operator of four → 40% of regional market share
- Signature of a data sharing agreement
- Period: 2 months, 2019
- Data provided
  - CDR and DDR → time stamp, cell tower identification, user*, technology

    *user pseudonimized by operator
  - Cell tower

    → identification, coordinates, active technology



Curiosity

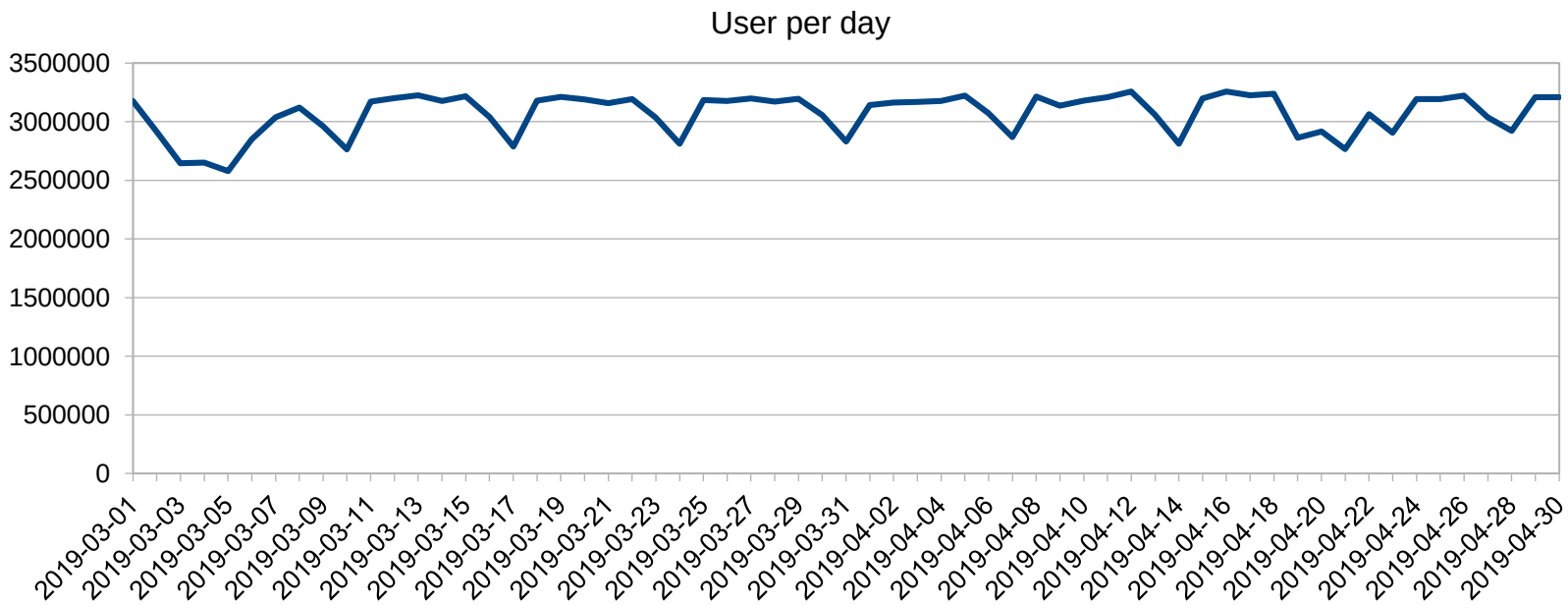
Rio Metropolitan Region / 2 months / 1 operator = 90 GB

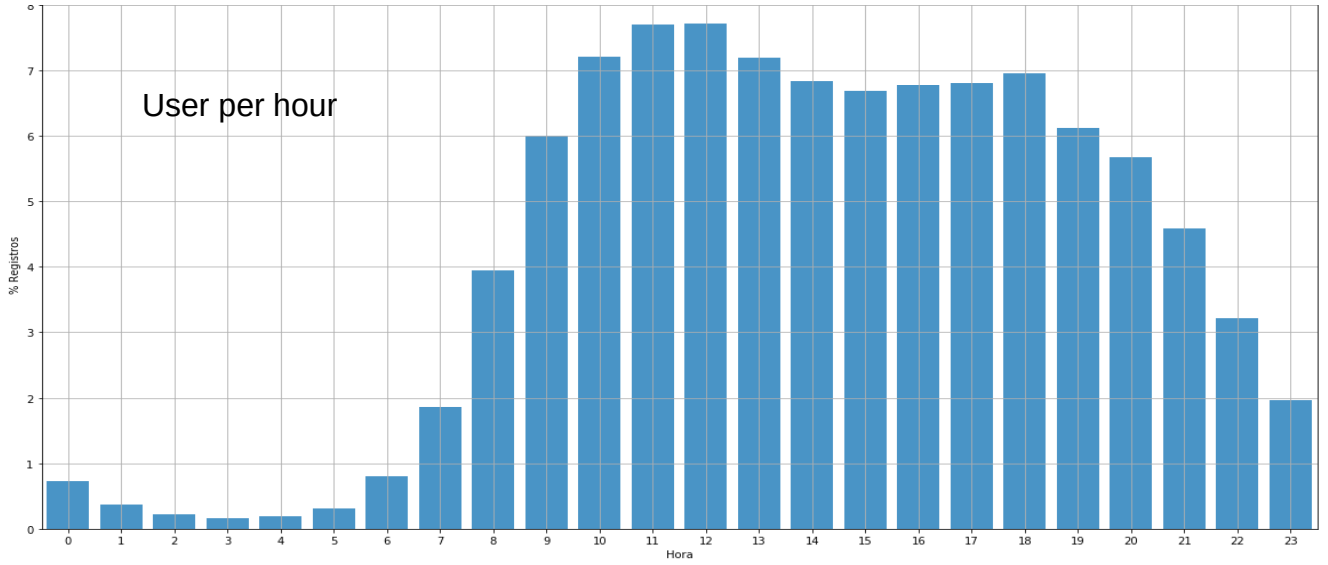
Brazil / 2 months / 4 operators = 6 TB

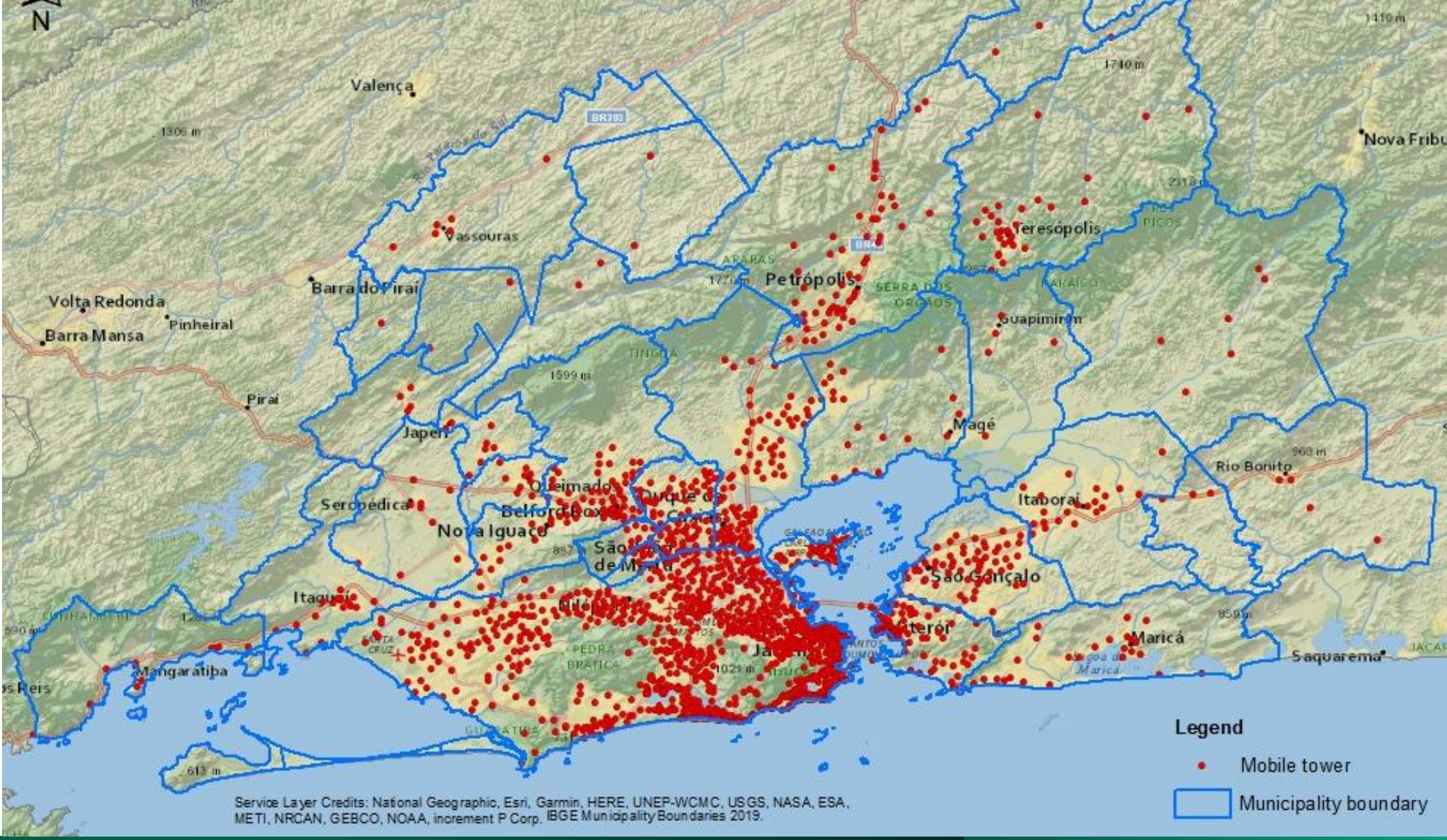
Brazil / 1 year / 4 operators = 36 TB

Data quality evaluation

Events	Day / user without data
	Events per day
	Users per day
	Distribution of events during the day
	Distribution of events per week
	Average number of events per day and per user
Cell towers	Geographic distribution of cell tower
	Number of events per cell tower


User per day
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
2019-03-01
2019-03-03
2019-03-05
2019-03-07
2019-03-09
2019-03-11
2019-03-13
2019-03-15
2019-03-17
2019-03-19
2019-03-21
2019-03-23
2019-03-25
2019-03-27
2019-03-29
2019-03-31
2019-04-02
2019-04-04
2019-04-06
2019-04-08
2019-04-10
2019-04-12
2019-04-14
2019-04-16
2019-04-18
2019-04-20
2019-04-22
2019-04-24
2019-04-26
2019-04-28
2019-04-30


User per hour
% Registros
Hora
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

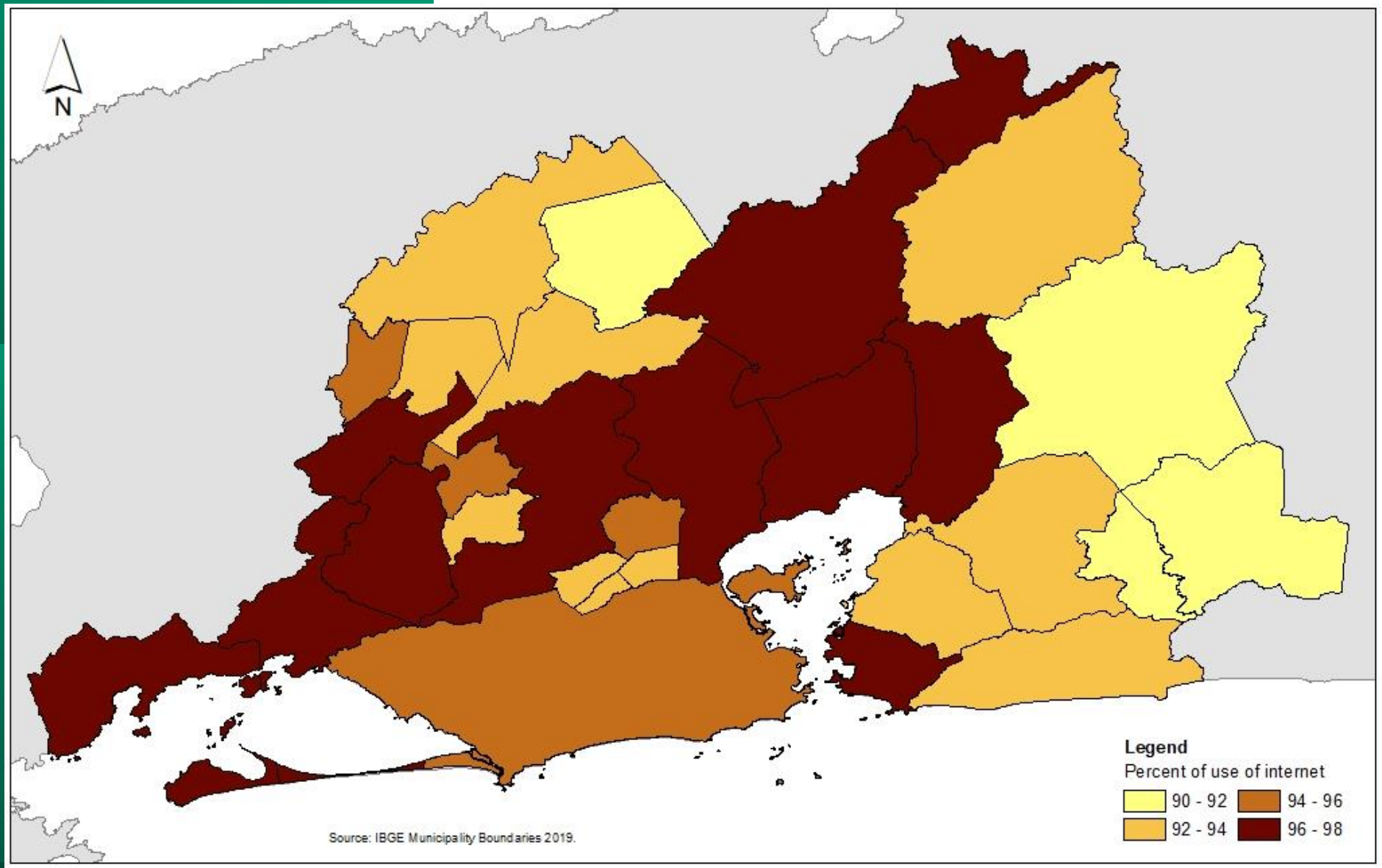


Geographic distribution of cell towers

Data processing

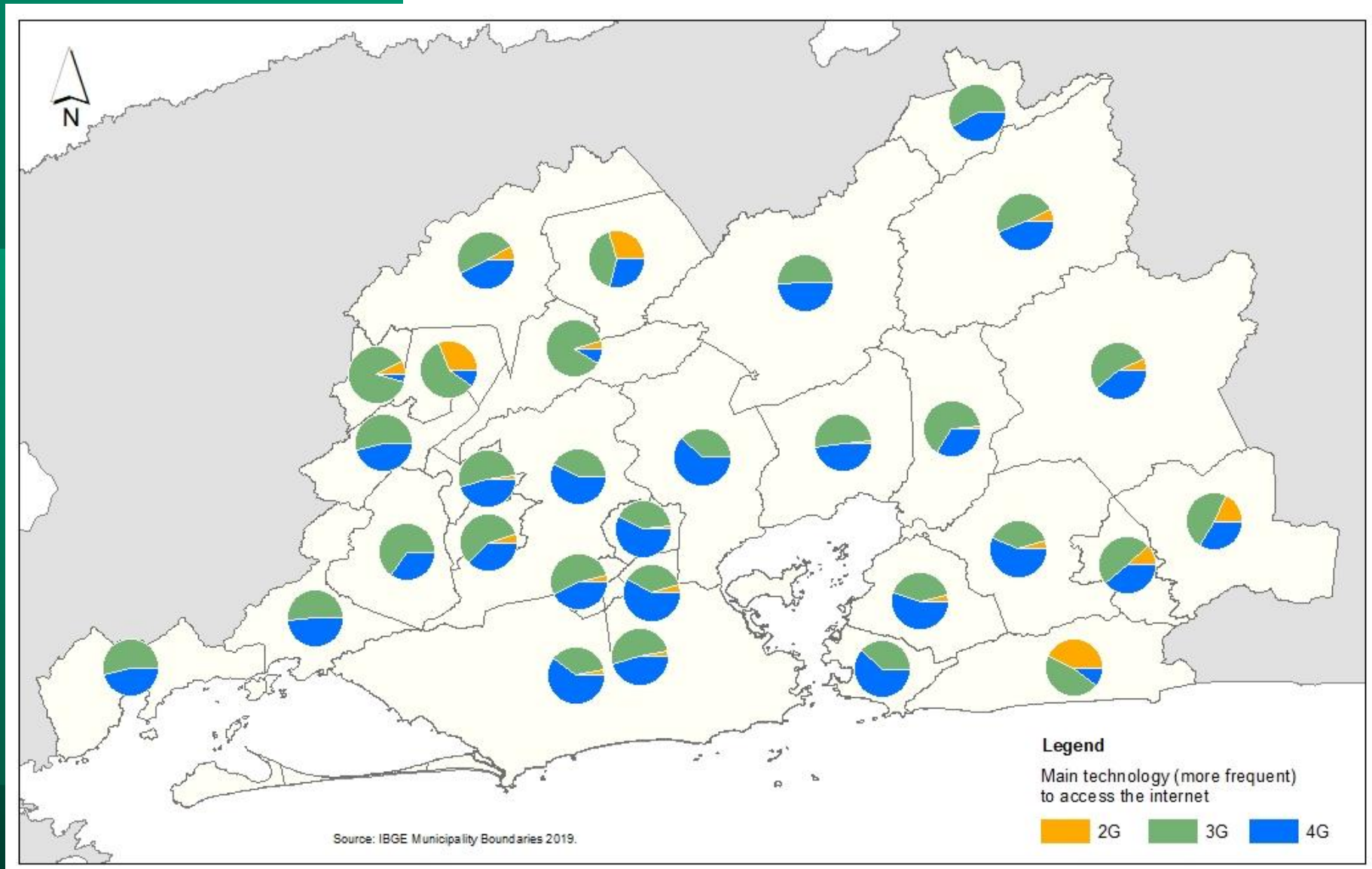
1. Place of residence inference based on cell phone use = anchoring
   - From midnight to 5 am
   - From 5 am to 8 am
   - From 9 pm to midnight

   → the cell tower most frequently used in one of these three time periods, from Monday to Thursday, was considered the place of residence
2. Totalization
   1. Total users, per residence cell tower, with only voice events, only internet events or both
   2. Total, per residence cell tower, of the most frequent technology for internet access


N
Legend
Percent of use of internet
90 - 92
92 - 94
94 - 96
96 - 98
Source: IBGE Municipality Boundaries 2019.

Results

Results


N
Legend
Main technology (more frequent)
to access the internet
2G
3G
4G
Source: IBGE Municipality Boundaries 2019.

Results

AREA	INTERNET ACCESS USING CELL PHONE (%)		
	MPD	PNAD Contínua	Diference (p.p.)
Study area	93,91	93,89	0,02
Rio de Janeiro Metropolitan Region	95,04	94,01	1,04
City of Rio de Janeiro	94,87	95,57	-0,70

Results

- The results for internet use obtained using MPD are very close to that obtained in the PNAD Contínua, showing that the methodology is robust

- The results can be used as a proxy for SDG Indicator 17.8.1 with the advantage of being more spatially disaggregated

- The results of the type of technology using MPD need to be better evaluated considering that there are many factors that interfere on them

Acknowledgments