SESSÃO 4. e-Science e Políticas Públicas de Sustentabilidade – 22/04

PARSEC : Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

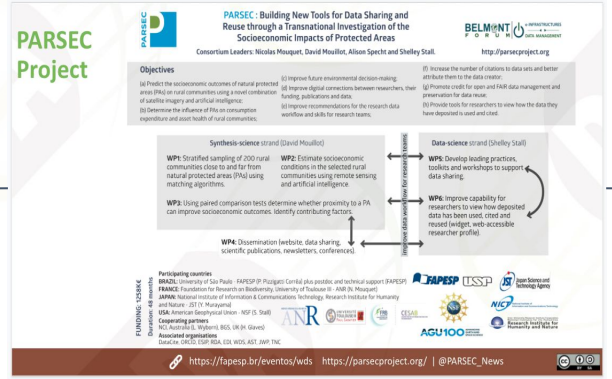
PARSEC

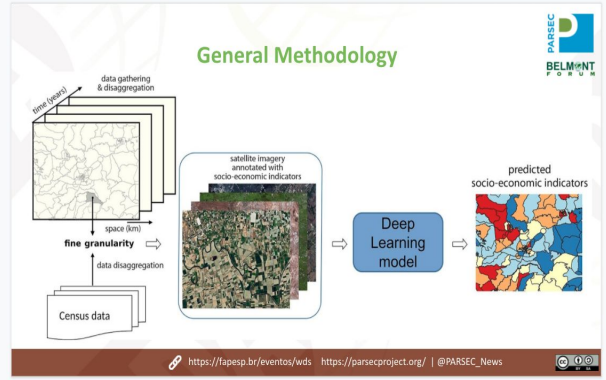
https://parsecproject.org

BELMONT FORUM | e-INFRASTRUCTURES DATA MANAGEMENT

Prof. Dr. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br
Digital Systems and Computer Engineering Department
Escola Politécnica da Universidade de São Paulo - EPUSP
Big Data and Data Science Research Group of EPUSP **wds.poli.usp.br**

Agenda





1st Case study of Brazilian team

Vale do Ribeira

2nd Case study of Brazilian team

Nexus area (Collab with INPE)

PARSEC Project

PARSEC : Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

Consortium Leaders: Nicolas Mouquet, David Mouillot, Alison Specht and Shelley Stall.

http://parsecproject.org

Objectives

(a) Predict the socioeconomic outcomes of natural protected areas (PAs) on rural communities using a novel combination of satellite imagery and artificial intelligence;

(b) Determine the influence of PAs on consumption expenditure and asset health of rural communities;

(c) Improve future environmental decision-making;

(d) Improve digitial connections between researchers, their funding, publications and data;

(e) Improve recommendations for the research data workflow and skills for research teams;

(f) Increase the number of citations to data sets and better attribute them to the data creator;

(g) Promote credit for open and FAIR data management and preservation for data reuse;

(h) Provide tools for researchers to view how the data they have deposited is used and cited.

Synthesis-science strand (David Mouillot)

WP1: Stratified sampling of 200 rural communities close to and far from natural protected areas (PAs) using matching algorithms.

WP2: Estimate socioeconomic conditions in the selected rural communities using remote sensing and artificial intelligence.

WP3: Using paired comparison tests determine whether proximity to a PA can improve socioeconomic outcomes. Identify contributing factors.

WP4: Dissemination (website, data sharing, scientific publications, newsletters, conferences).

improve data workflow for research teams

Data-science strand (Shelley Stall)

WP5: Develop leading practices, toolkits and workshops to support data sharing.

WP6: Improve capability for researchers to view how deposited data has been used, cited and reused (widget, web-accessible researcher profile).

FUNDING: 1258K€

Duration: 48 months

Participating countries

BRAZIL: University of São Paulo - FAPESP (P. Pizzigatti Corrêa) plus postdoc and technical support (FAPESP)

FRANCE: Foundation for Research on Biodiversity, University of Toulouse III - ANR (N. Mouquet)

JAPAN: National Institute of Information & Communications Technology, Research Institute for Humanity and Nature - JST (Y. Murayama)

USA: American Geophysical Union - NSF (S. Stall)

Cooperating partners

NCI, Australia (L. Wyborn), BGS, UK (H. Glaves)

Associated organisations

DataCite, ORCID, ESIP, RDA, EDI, WDS, AST, JWP, TNC

IA models

State-of-the-art deep-learning & poverty prediction

[Ayush, et al, 2016]

Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)
Special Track on AI for Computational Sustainability and Human Well-being

Generating Interpretable Poverty Maps using Object Detection in Satellite Images

Kumar Ayush[1*], Burak Uzkent[1*], Marshall Burke[2], David Lobell[2] and Stefano Ermon[1]

[Jean, et al, 2016]

Science

Combining satellite imagery and machine learning to predict poverty

Neal Jean,[1,2*] Marshall Burke,[3,4,5*†] Michael Xie,[1] W. Matthew Davis,[4] David B. Lobell,[3,4] Stefano Ermon[1]

2016 • VOL 353 ISSUE 6301

[Suel, et al, 2019]

scientific reports

Measuring social, environmental and health inequalities using deep learning and street imagery

Esra Suel, John W. Polak, James E. Bennett & Majid Ezzati

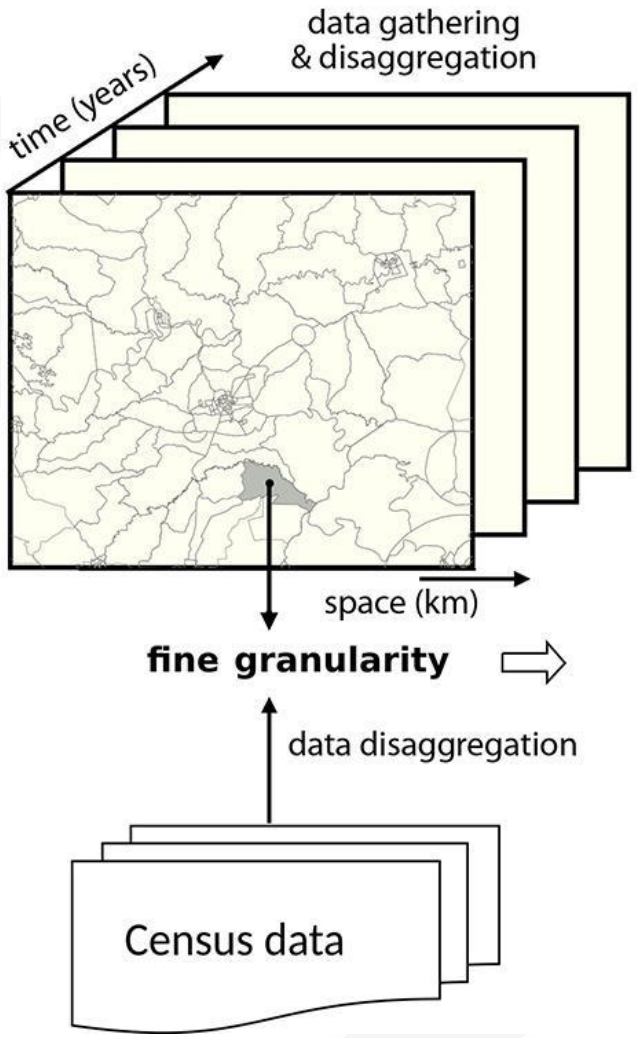
Scientific Reports **9**, Article number: 6229 (2019)

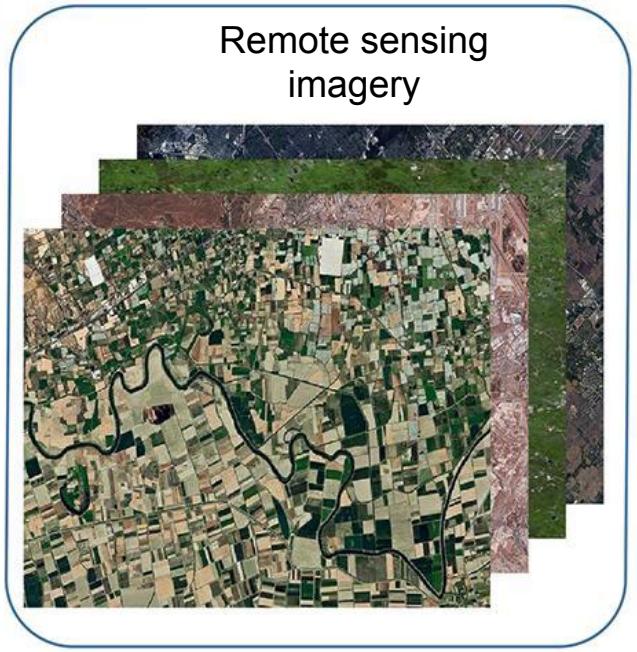
[Yeh, et al, 2020]

Using publicly available satellite imagery and deep learning to understand economic well-being in Africa

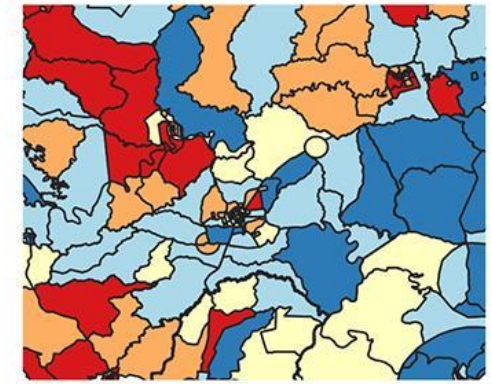
Christopher Yeh[1,7], Anthony Perez[1,2,7], Anne Driscoll[3], George Azzari[2,4], Zhongyi Tang[5], D
Stefano Ermon[1] & Marshall Burke[3,4,5,6]

General Methodology

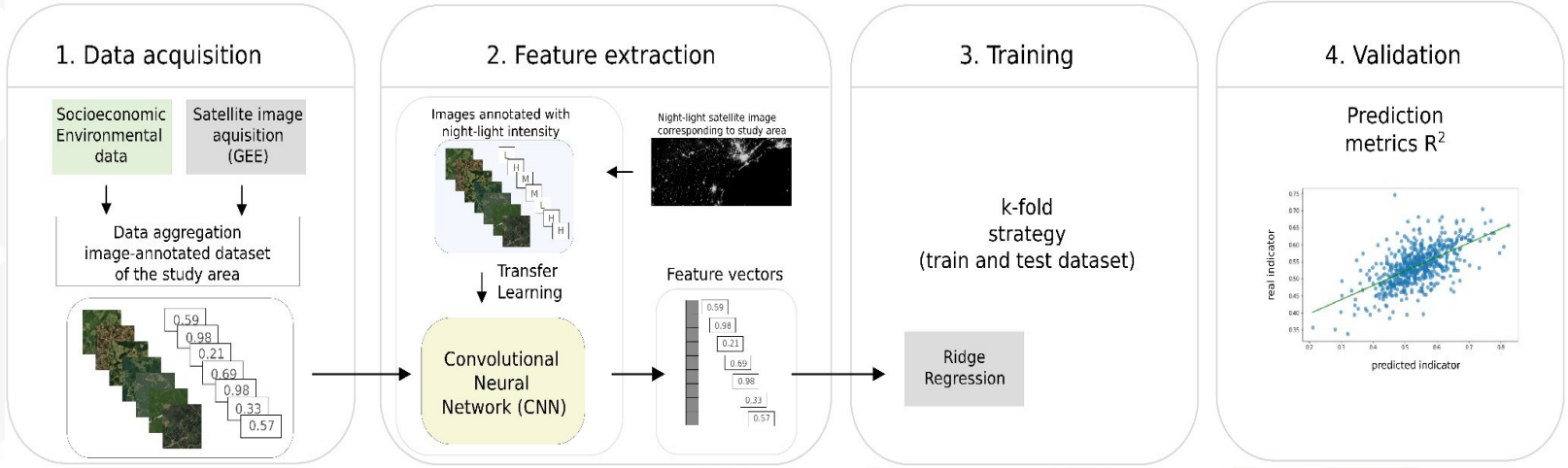




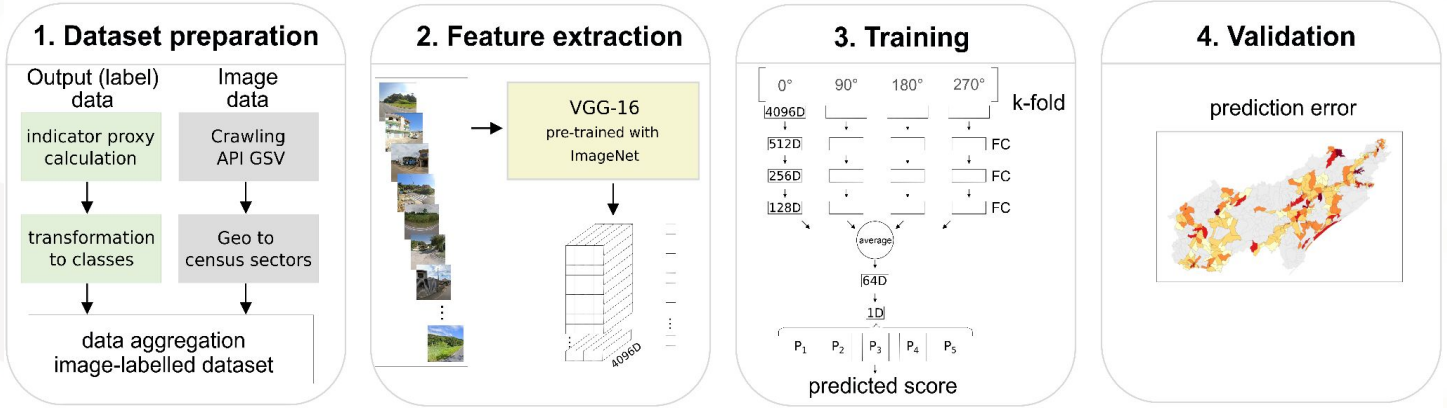
Deep Learning model



Methodology using satellite images



Methodology using Street view images



Types of Data Used and Generated

Raw Data: Inputs

Ground Truth Data
Local surveys

Satellite images

Processed Data: Outputs

Global gridded socio-economic information

Socio-economic dynamics in our study sites

Raw data

The DHS Program
Demographic and Health Surveys

USAID
FROM THE AMERICAN PEOPLE

Atlas do Desenvolvimento Humano no Brasil

+30.000 household surveys every 5 years in +90 countries

- Anemia - prevalence of anemia, iron supplementation
- Child Health - vaccinations, childhood illness, newborn care
- Domestic Violence (module) - prevalence of domestic violence and consequences of violence
- Education - literacy, attendance, highest level achieved
- Environmental Health - water, sanitation, cooking fuel
- Family Planning - knowledge and use of contraceptives

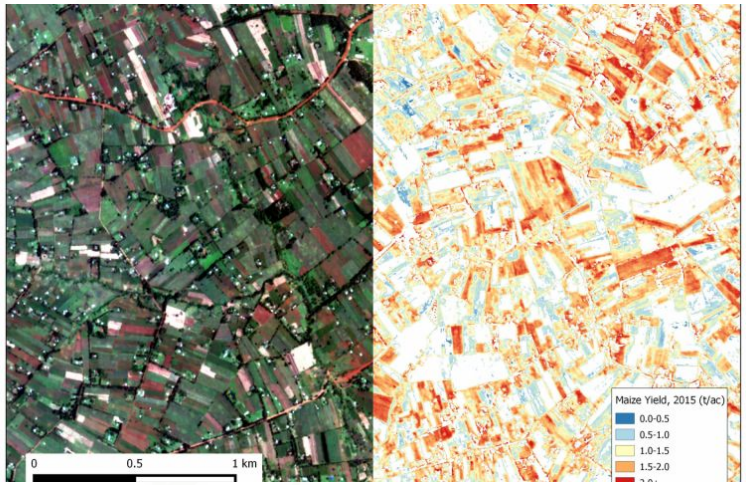
No restriction access for academic research
Can be stored locally or downloaded when necessary

Raw data

Satellite images

Many options

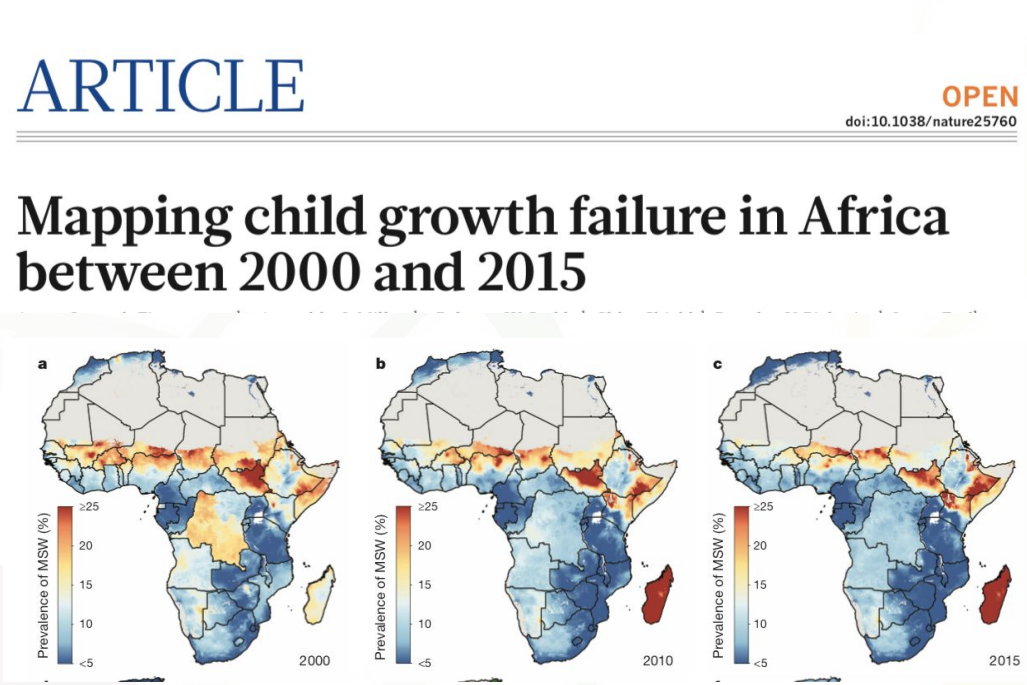
- Both free and fees required
- Various Resolutions
- Various time series
- Usability can be an issue for some datasets



Processed data

Global gridded socio-economic information

LSMS
Living Standards Measurement Study


ARTICLE

OPEN
doi:10.1038/nature25760

Mapping child growth failure in Africa between 2000 and 2015

Only some information/data have gridded at the regional scale

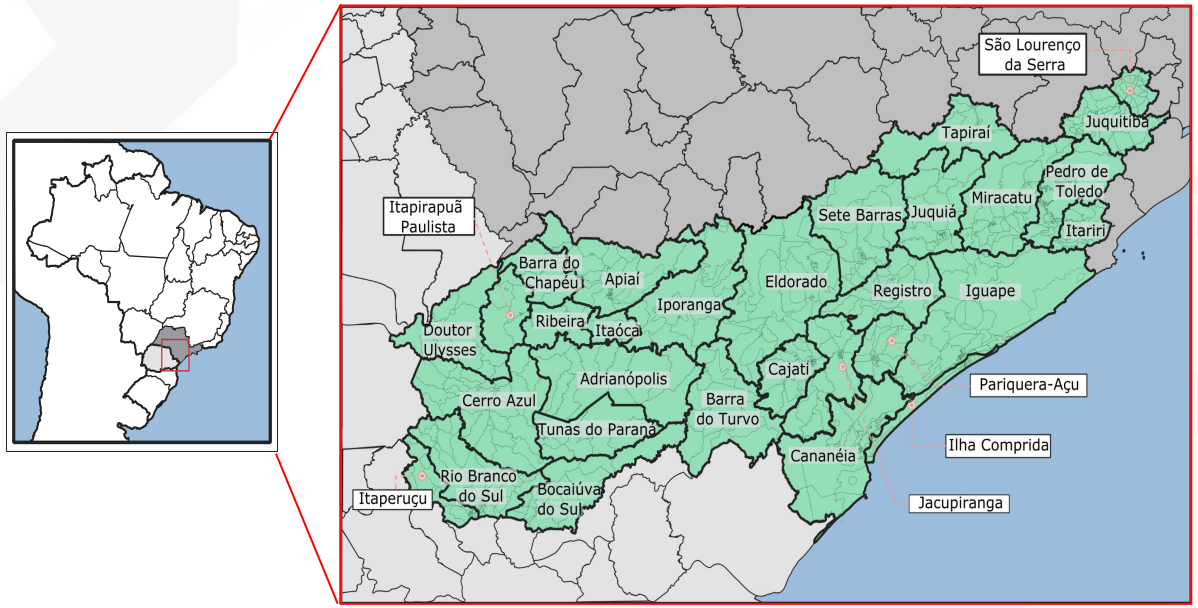
We plan to grid many of them globally and make them available

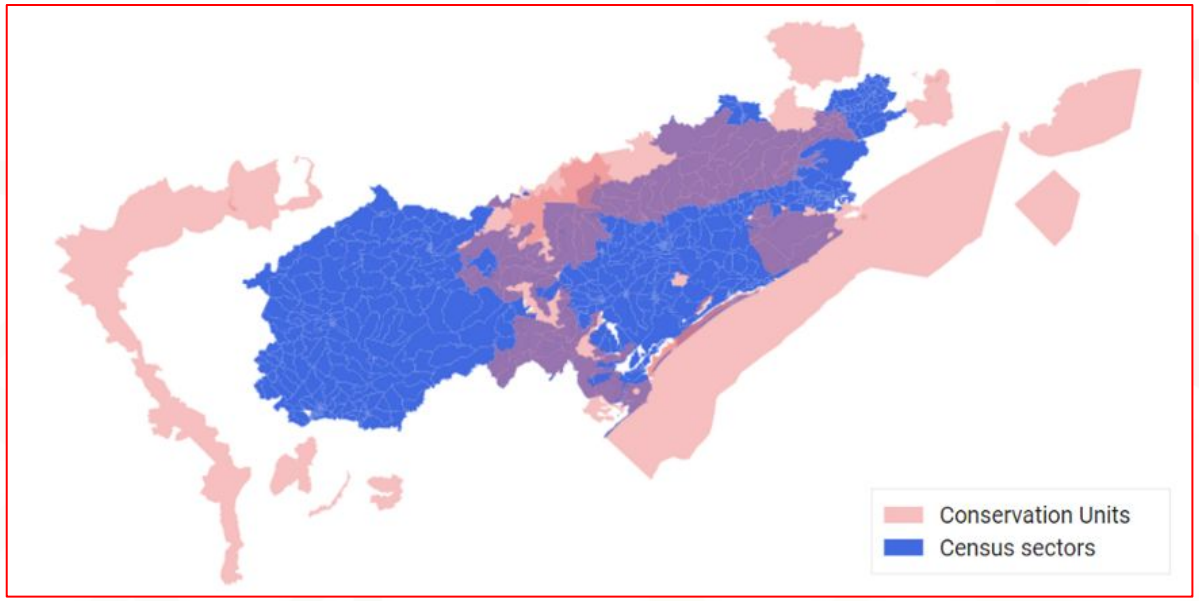
1st Case study of Brazilian team

Vale do Ribeira

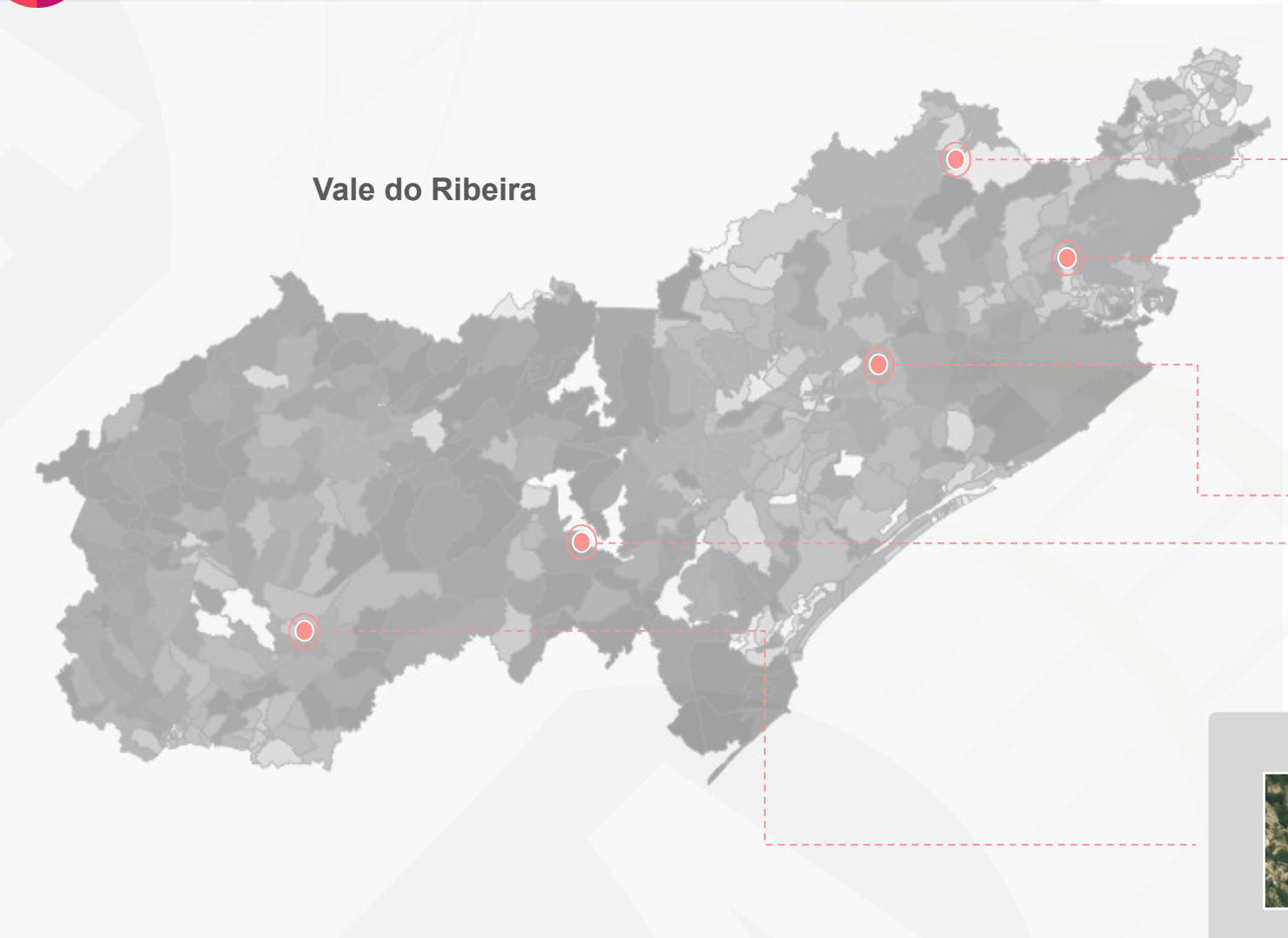
Surrounded by Protected Areas

Location

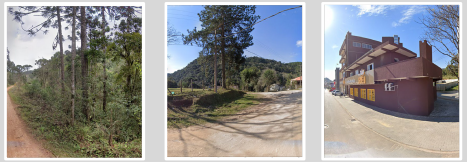




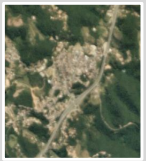
Estimate socioeconomic indicators

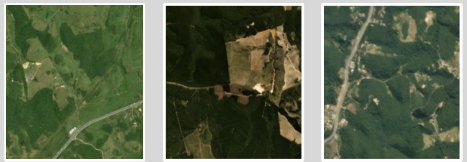


Street View Images



Satellite Images

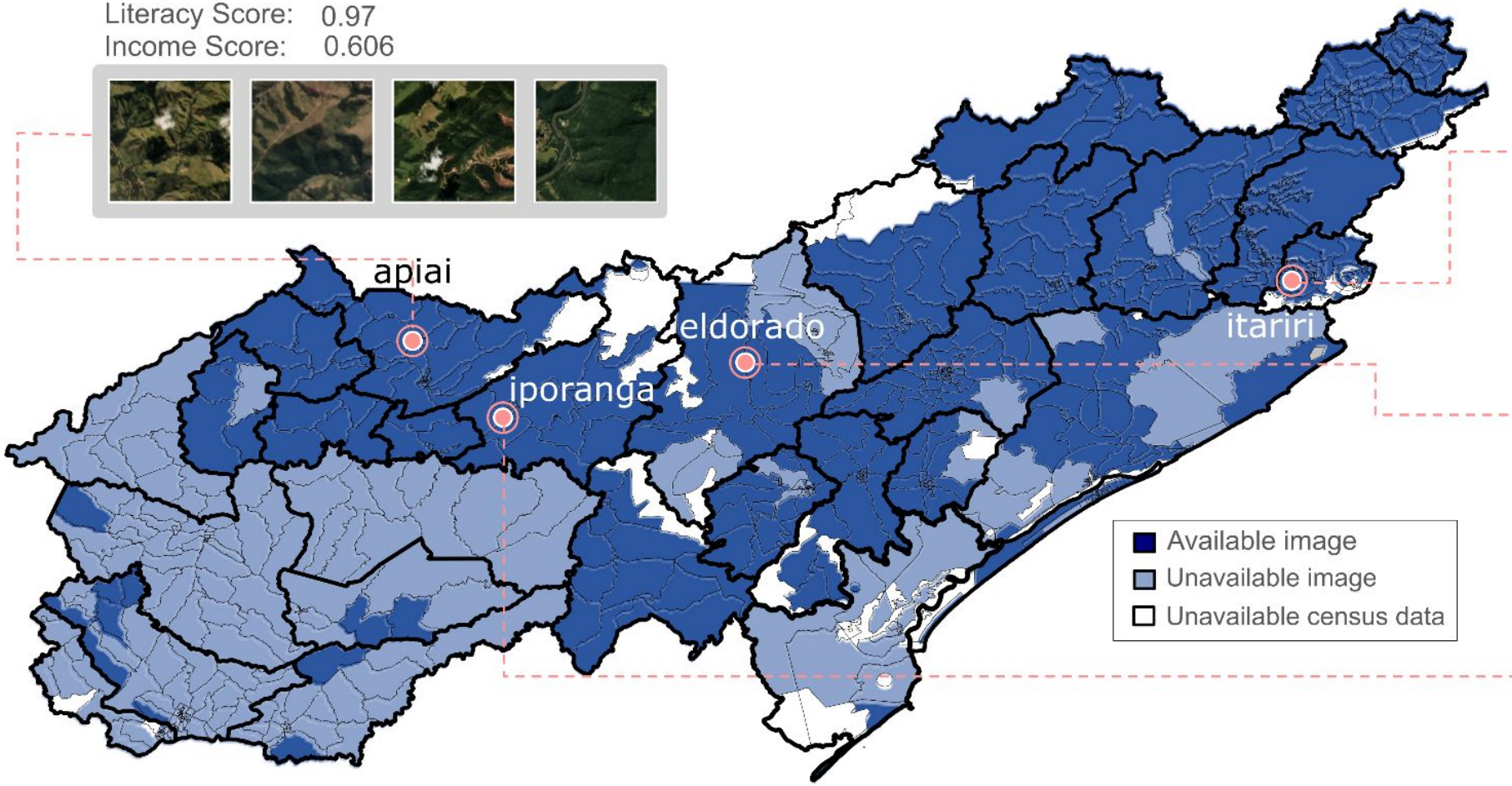




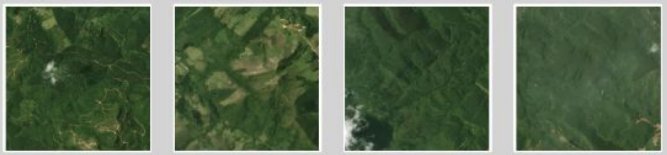
Satellite image availability

Longevity Score: 0.8
Literacy Score: 0.97
Income Score: 0.606

Longevity Score: 0.917
Literacy Score: 0.982
Income Score: 0.635

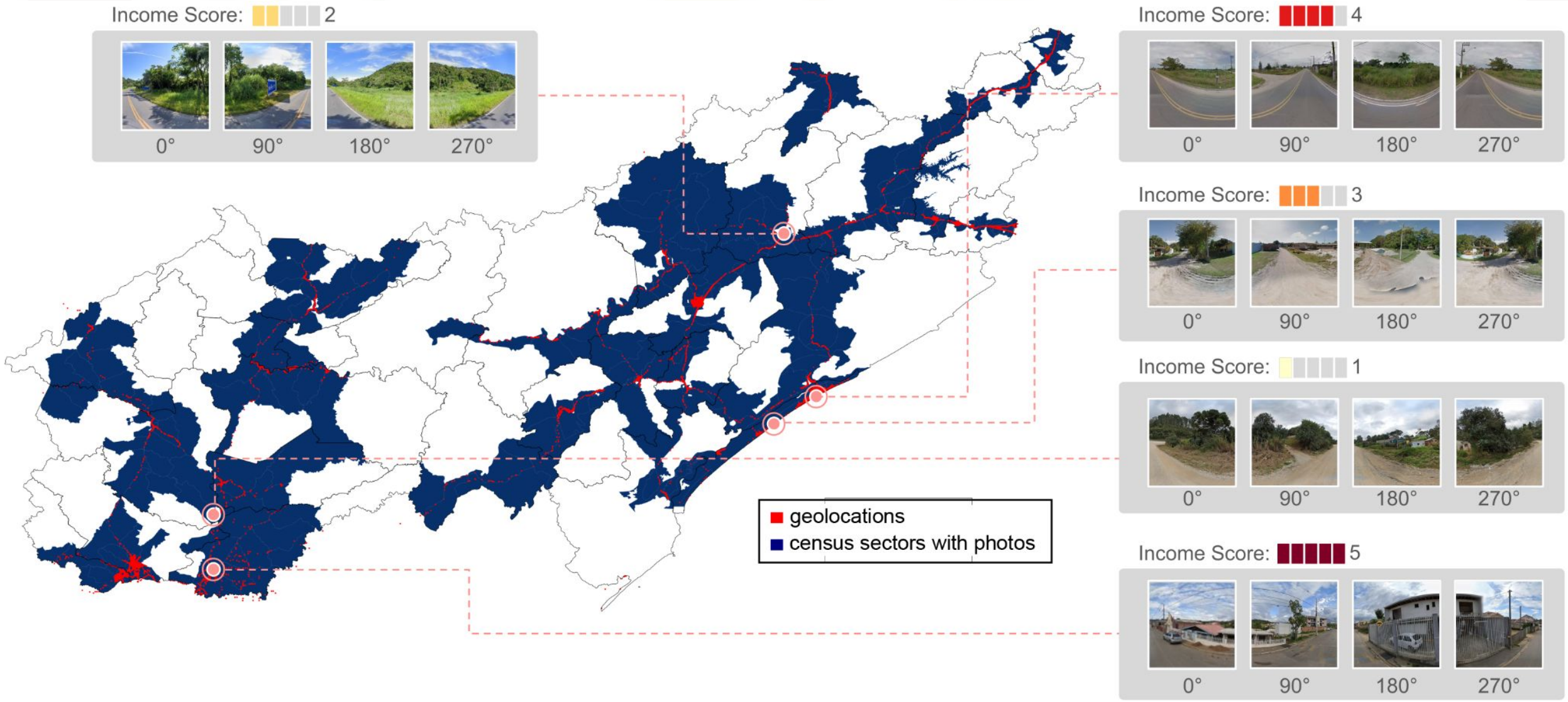


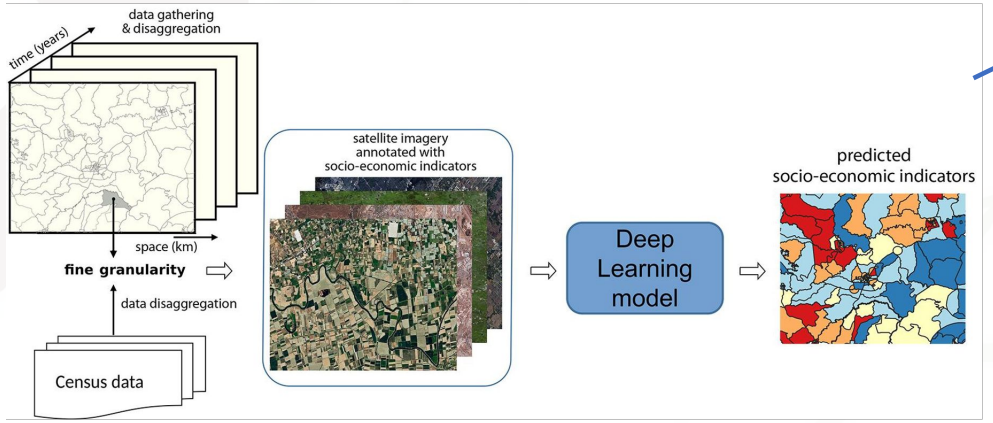
Longevity Score: 0.9
Literacy Score: 0.981
Income Score: 0.654



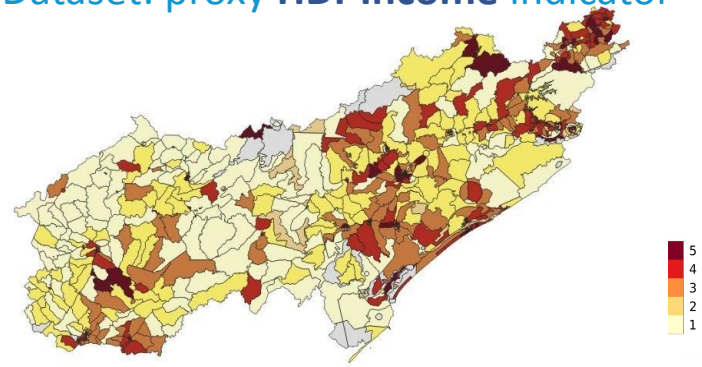
Longevity Score: 0.85
Literacy Score: 0.964
Income Score: 0.594

Street view images availability

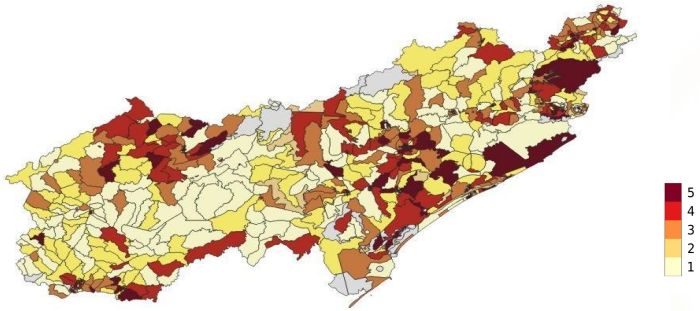




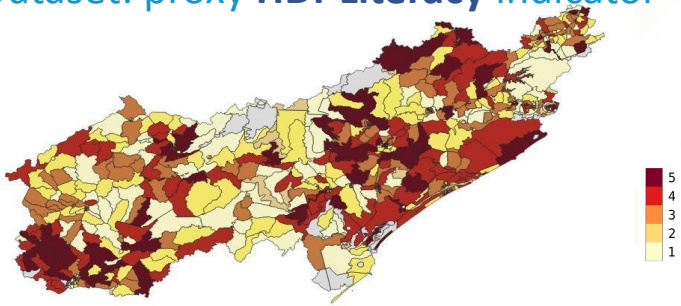
Dataset: proxy **HDI-Income** indicator



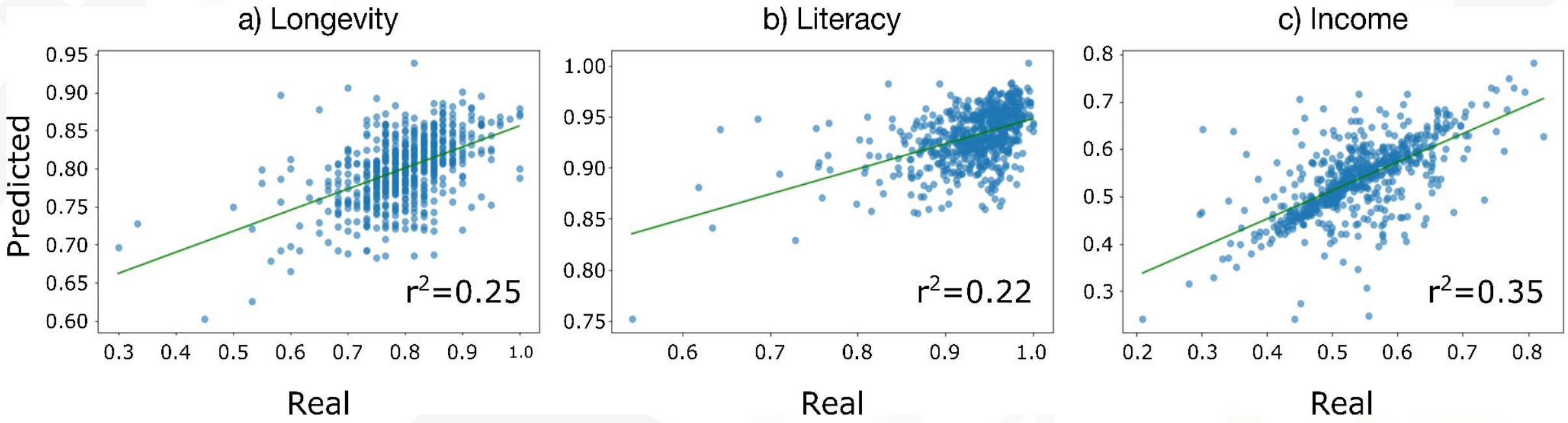
Dataset: proxy **HDI-Longevity** indicator



Dataset: proxy **HDI-Literacy** indicator

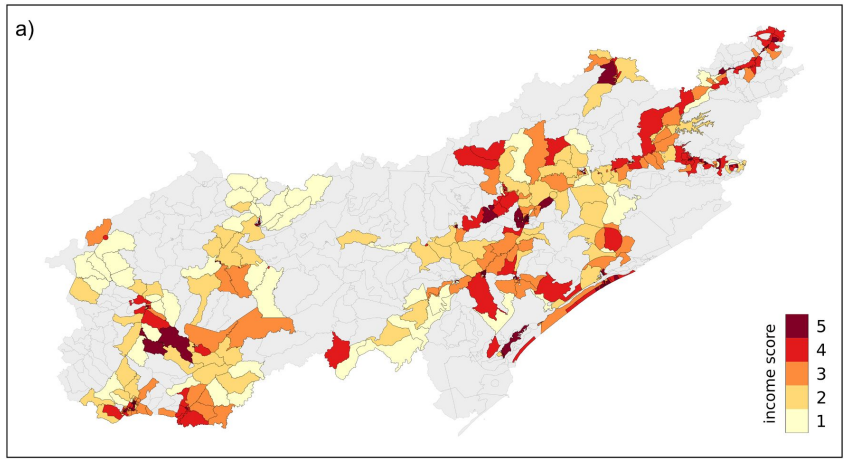


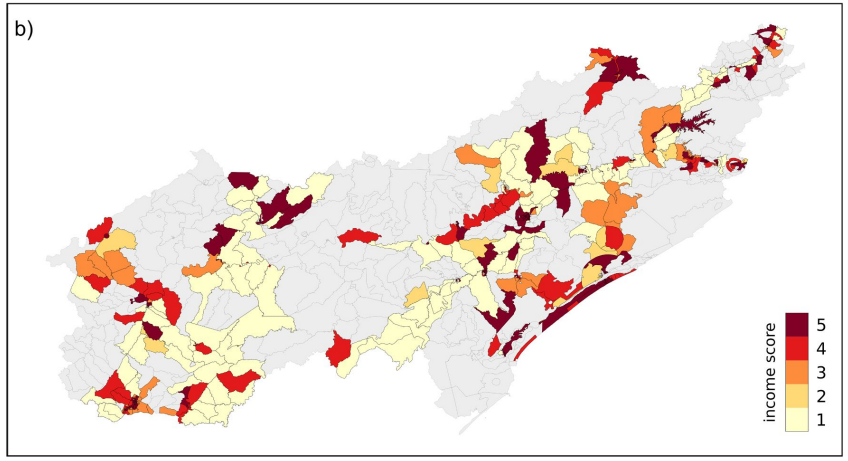
Prediction results using satellite images

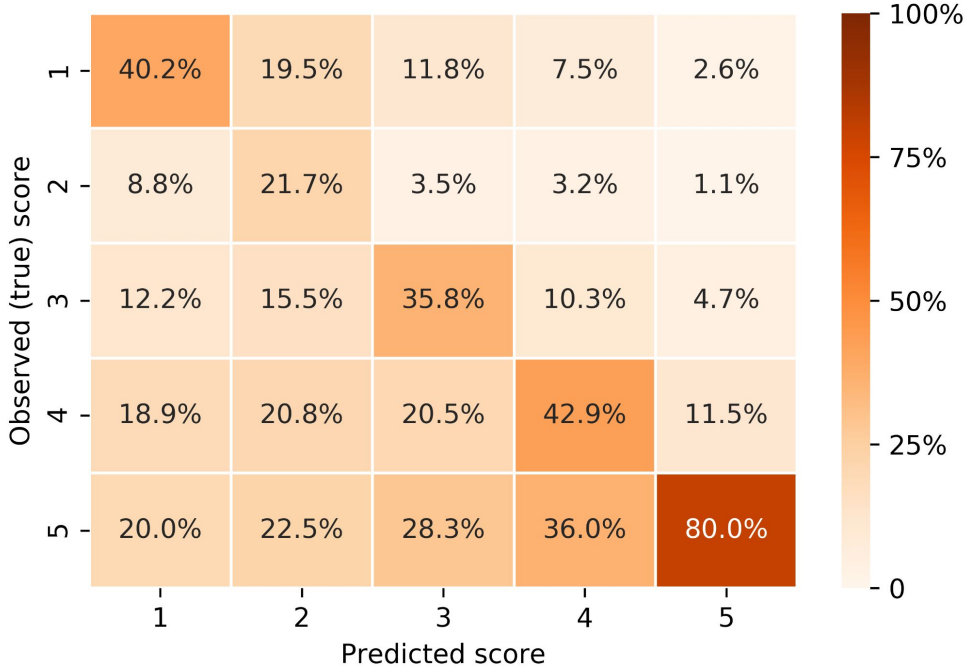


Prediction results using Street view images

Machicao, et al. A Deep-Learning Method for the Prediction of Socio-Economic Indicators from Street-View Imagery Using a Case Study from Brazil. *Data Science Journal*, 21(1), p.6. DOI: http://doi.org/10.5334/dsj-2022-006



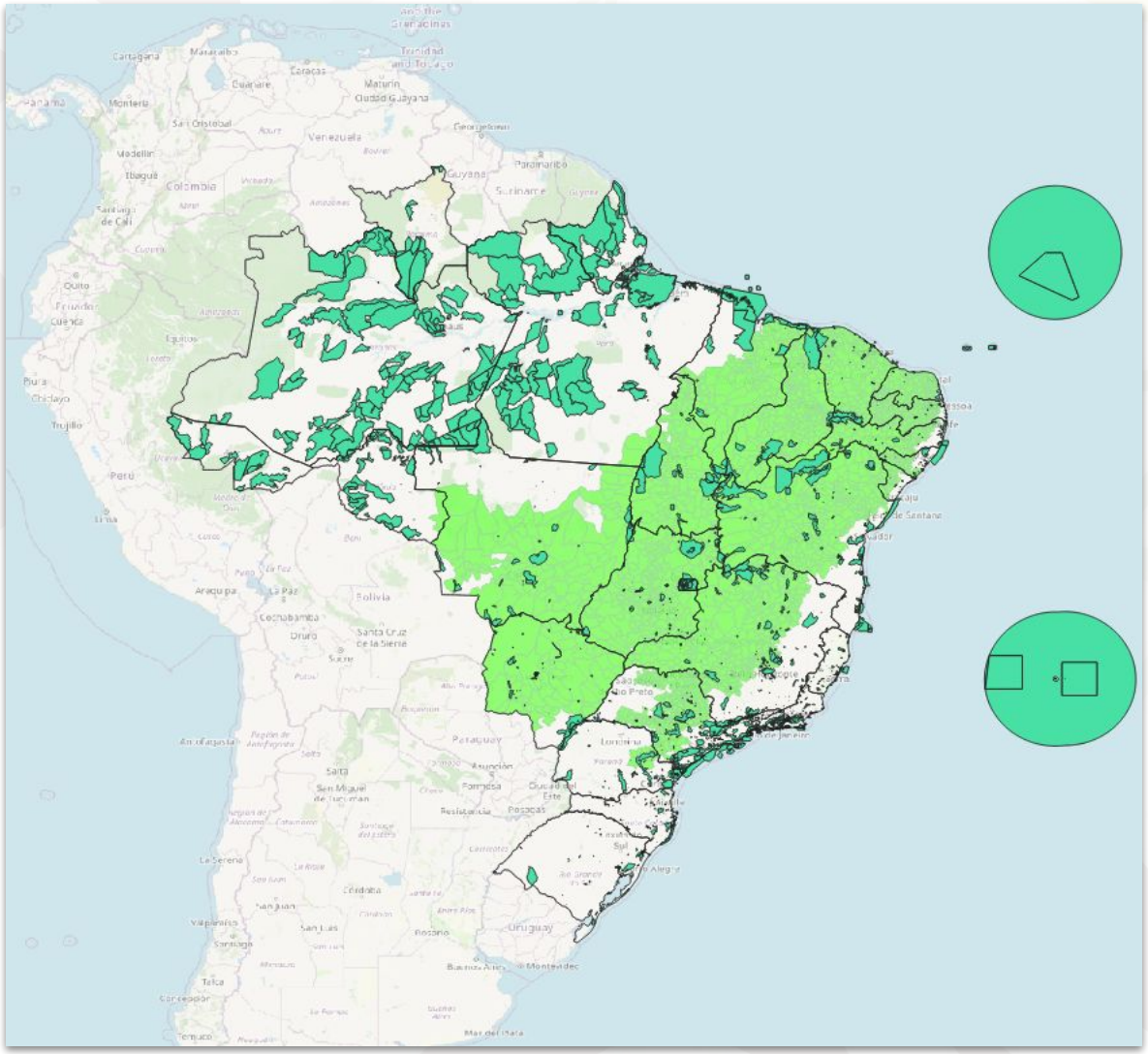




2nd Case study of Brazilian team

Nexus area (Collab with INPE)

NEXUS area



- 2513 municipalities
- A huge extension of Brazilian territory (~ 3,5 millions km2)
- Contains various biomes
  - Amazonia, Caatinga, Cerrado, Mata atlantica, Pantanal.
- Contains two important basins
  - São Francisco River Basin
  - Parnaiba Basin

PAs (pre-filtered) in Brazil

NEXUS area

Indicators examples

Health

- Occurrence of diseases linked to water source
- Doses applied for per capita immunization

Education & Culture

- school attendance rate
- Decrease in school dropouts, a measure adopted by the agency

Work & income

- Low income (households with up to one salary)
- Value of annual revenue obtained by family farming establishments

Infrastructure

- Proportion of people affected by climate-related natural disasters
- Inappropriate Households

governance

- Urban Policy Instruments
- Proportion of homicide of women and youth

Genre

- Women's Reference Center per capita
- Proportion of women in formal jobs with more than 5 minimum wages

Sanitation

- Proportion of population that is served with water supply
- Proportion of permanent private households with garbage collection service

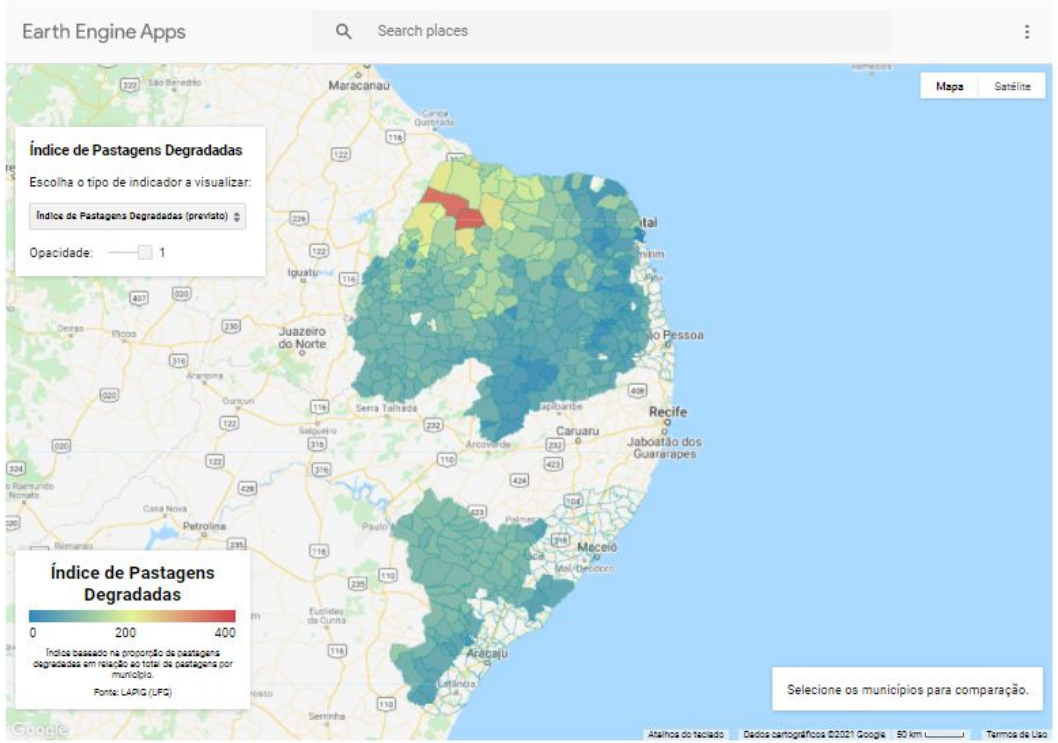
NEXUS have more than 50 socioeconomic and environmental indicators corresponding to 2015

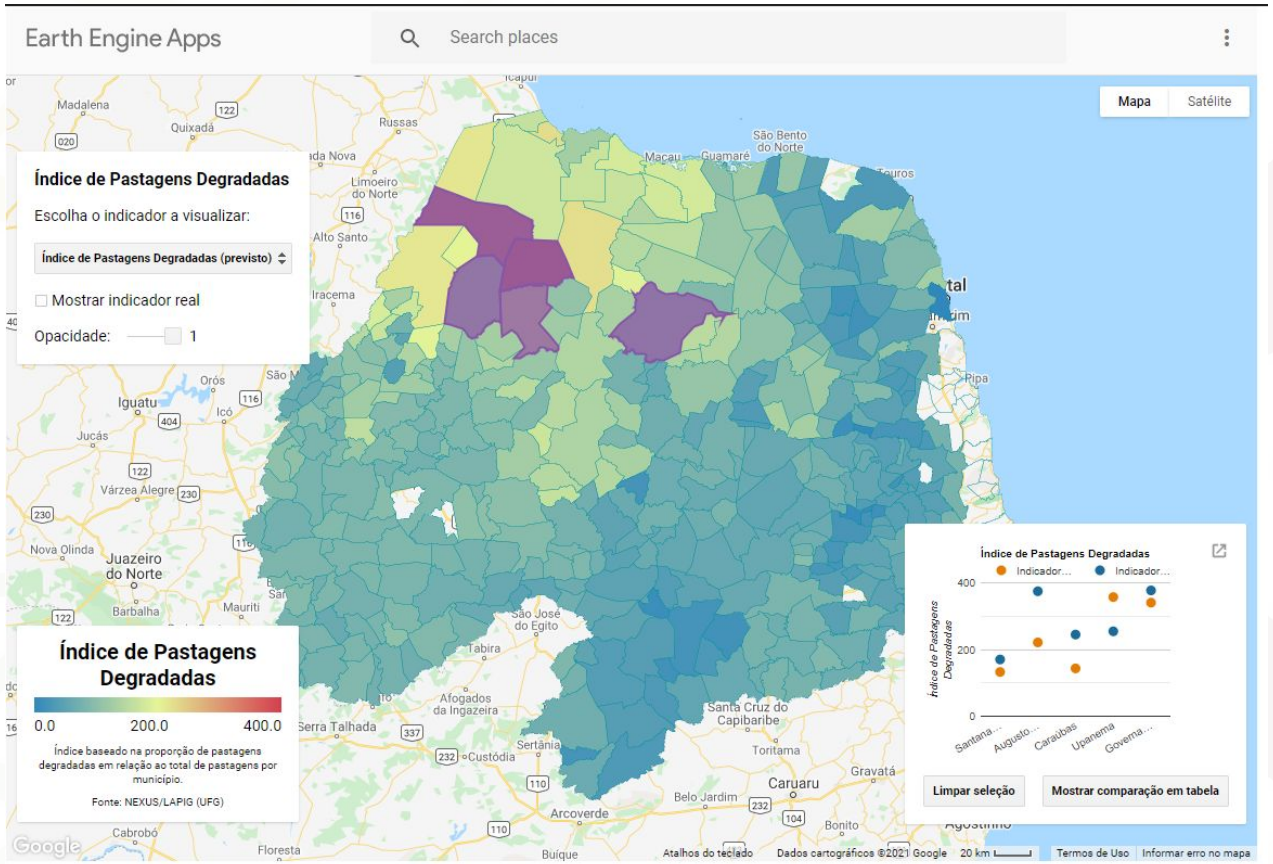
Prediction results

Nome do indicador socioeconômico	R^2
Domicílios com renda maior que um salário mínimo	0.347
Isolamento da população considerando a distância a corpos hídricos e estradas	0.328
PIB per capita	0.197
Domicílios Inadequados	0.101
Ocorrência de doenças veiculadas com fonte hídrica	0.073
Taxa de mortalidade em menores de 5 anos de idade	0.013
Proporção de Cadastramento de pessoas em serviços básicos de saúde	0.002
Taxa de mortalidade fetal (TMF) - óbitos fetais (Óbitos fetais)	-0.397

Platform

Acreditamos que o emprego da metodologia de aprendizado de máquina para análise de imagens de satélite pode potencializar o monitoramento socioambiental, facilitando iniciativas voltadas para a preservação ambiental e desenvolvimento sustentável no Brasil. Como parte do projeto Nexus, desenvolvemos um modelo baseado em redes neurais capaz de prever indicadores ambientais e socioeconômicos na Caatinga brasileira através de imagens de satélite.





LINK/QR Code para a página:
https://sites.google.com/usp.br/visualizanexus/

Data and Digital Outputs Management Plan (DDOMP) Guide

A Step-By-Step User Guide for Building a Successful Data Management Plan

Why is a Data and Digital Outputs Management Plan **(DDOMP)** important? Ensuring proper data management helps Belmont Forum researchers achieve the goal of supporting international transdisciplinary research to provide knowledge for understanding, mitigation, and adaptation to global environmental change, and is required of all Belmont Forum-funded projects. A DDOMP ensures that the data is organized, sharable, and reproducible, and helps researchers gain recognition and credibility through data sharing.

Benefits of a Workbook

Supports the iterative / flexible needs of the research team
Provides guidance on “what data”, “where to store”, “what to track”
Provides guidance on when actions are taken
Gives method for

- what to do **during** the project
- how to **preserve** your digital objects for **publication and sharing**

Checklist for your team to make it “super simple”
Validation task for the PI to ensure **compliance** and **consistency**

Details to follow

DDOMP Checklist – Team Resources

- Material development and temporary storage location
  - Google Drive
- Team communications and information decimation tools
  - Email, Slack
- Dataset storage location during the project
  - Open Science Framework (AWS integration)
- Software development platform
  - GitHub
- Data preservation (including derived products) repository
  - Environmental Data Initiative
- Software preservation repository
  - Zenodo
- Training, workshop material preservation repository
  - Zenodo

PARSEC

- PIs – 4
- Country Leaders – 6
- Funders – 4
- Researchers – 30
- Languages - 4

PARSEC Data and Digital Output Management Plan and Workbook

Further details can be found on the process and methods used for PARSEC:

Stall, Shelley, Specht, Alison, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Mabile, Laurence, Machicao, Jeaneth, O'Brien, Margaret, Wyborn, Lesley. (2020). PARSEC Data and Digital Output Management Plan and Workbook. Zenodo. 10.5281/zenodo.3891426

Use your DMP or DDOMP to make your own Checklist.

Special thanks to PARSEC Brazilian team

Researchers:
- Profa. Dra. **Katia** Maria Paschoaletto Micchi de Barros Ferraz: (ESALQ/USP);
- Dr. **Jean** Pierre Henry Balbaud Ometto: Instituto Nacional de Pesquisas Espaciais (INPE);
- Dra. Marina **Jeaneth** Machicao Justo - (EPUSP), (postdoc fellowship PARSEC);
- Dra. **Solange** Maria Dos Santos: (Scielo);
- Dr. **Silvio** Marchini - (ESALQ/USP);
- Eng. **Danton** Ferreira Vellenich (EPUSP), (TT fellowship PARSEC and Master Student);
- Prof. Dr. **Pedro** Luiz Pizzigatti Corrêa: (EPUSP) (Country Leader).

Support of Brazilian Institute of Geography and Statistics (IBGE):
- Dra. **Nadya** Maria Deps (IBGE);
- MSc. **Miguel** Suarez Xavier Penteado (IBGE).

Support of Support of **INPE**:
- Dr. **Pedro** R. Andrade;
- Dr. **Gustavo** Arcoverde;

SESSÃO 4. e-Science e Políticas Públicas de Sustentabilidade – 22/04

PARSEC : Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

PARSEC

https://parsecproject.org

BELMONT FORUM | e-INFRASTRUCTURES DATA MANAGEMENT

Prof. Dr. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br
Digital Systems and Computer Engineering Department
Escola Politécnica da Universidade de São Paulo - EPUSP
Big Data and Data Science Research Group of EPUSP **wds.poli.usp.br**