Making Global Connections in Earth and Environmental Science Data Infrastructures and Repositories

April, 21st – 2021

Earth and Environmental Science Data Infrastructures in Brazil

Prof. 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**

https://www.rd-alliance.org/plenaries/rda-17th-plenary-meeting-edinburgh-virtual/making-global-connections-earth-and

School of Engineering (Escola Politécnica – POLI) of University of São Paulo is the most complete and important engineering School in Latin America

"Train professionals comitted to the sustainable development of the country, with social, economic and environmental responsibilities (...)"

Poli Mission

Fonte: Website da Escola Politécnica (www.poli.usp.br), Times Higher Education

Institutional Data

- 15 Departments of Teaching and Research
- Built area: 141,500m² -
- 9 buildings Library: collection of 590,000 documents
- UnderGrad: ~ 4,500 students enrolled PhD: ~ 840 Master's and ~730 PhD students

Leadership position

- Poli/USP is the 105th best technology school in the world and the best in Latin America
- Largest graduate center in engineering in Brazil
- One of the largest trainers of entrepreneurs and executives in the country
- USP is responsible for more than 20% of the total national scientific production

Research on management and analysis of large volume of scientific data

Pedro Luiz Pizzigatti Corrêa:

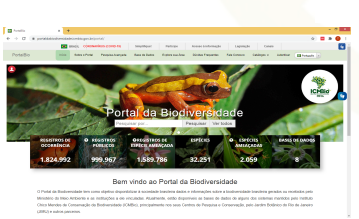
- Associate Professor (Univeristy of São Paulo - USP) – Department of Computer and Digital Systems Engineering and Coordinator Big Data and Data Science Research Group - Engineering.
- Education: Bachelor and Master of Computer Science (USP). PhD in Electrical Engineering (USP) and Post-doctorate in data science focusing on distributed databases – Univesity of Tem

Research projects involving scientific data management:

- Devlopment of new tools for sharing and reuse of data through transnational research on the socioeconomic impact of Conservation Units (PARSEC) - FAPESP/NSF/ANR/JST - BELMONT FORUM - Result – Data Science and Computational Models (that uses satellite images to generate socioeconomic indicators of communities close to Protect Areas) - https://parsecproject.org/
- FAPESP Thematic Project in the Climate Change program focusing on Data Management (Coordination Prof. Dr. Paulo Artaxo). Result under development: Model for Aerosol Data Quality Management Report (DQMR), Data Portal and Big Data Analysis based on Cloud Infrastructure (Partnership with ARM/ORNL/DoE/USA)
- E-Science Program - FAPESP - "Enabling Integrated Research through monitoring of biodiversity and climate measurements" – Result: Infrastructure of Big Data Analytics bioclimatic data that integrates biodiversity observation data and aerosols collected at different sites near the city of Manaus (Amazon - Brazil) – finished
- Brazilian Biodiversity Data Portal – Minitry of Environment - Brazil, 2015 in colaboration with Atlas of Living Australian (ALA) https://portaldabiodiversidade.icmbio.gov.br/

https://parsecproject.org/

https://portaldabiodiversidade.icmbio.gov.br/



Big Data and Data Science Research Group of Engineering wds.poli.usp.br

Center of Data Science (C2D) – Itaú-Unibanco - http://c2d.poli.usp.br/

Collaboration:

USGS science for a changing world

The University of Tennessee Knoxville

redhat

Oak Ridge National Laboratory

INPE Unidade de Pesquisa do MCTIC

UNIFESP Universidade Federal de São Paulo

ICMBio MMA

https://www.rd-alliance.org/plenaries/rda-17th-plenary-meeting-edinburgh-virtual/making-global-connections-earth-and

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)

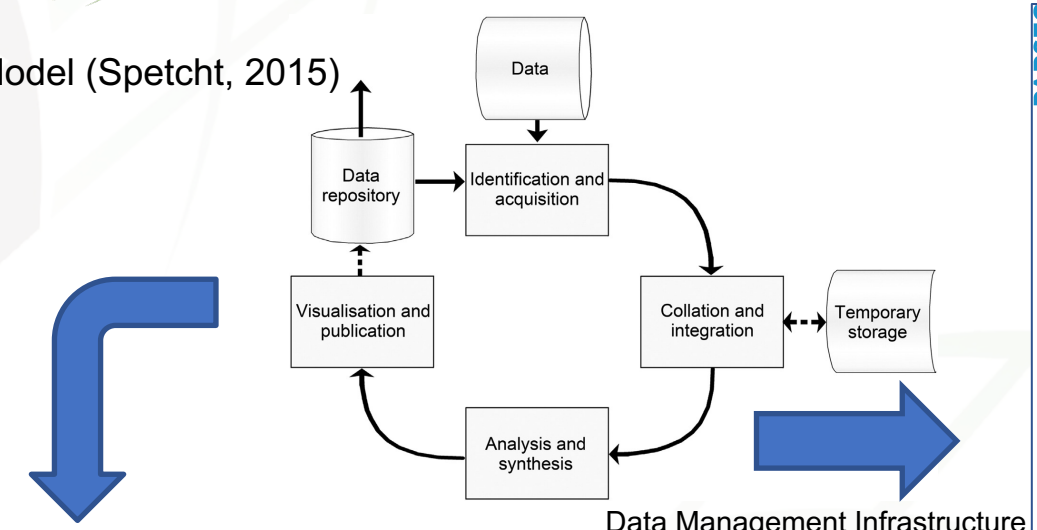
Associated organisations

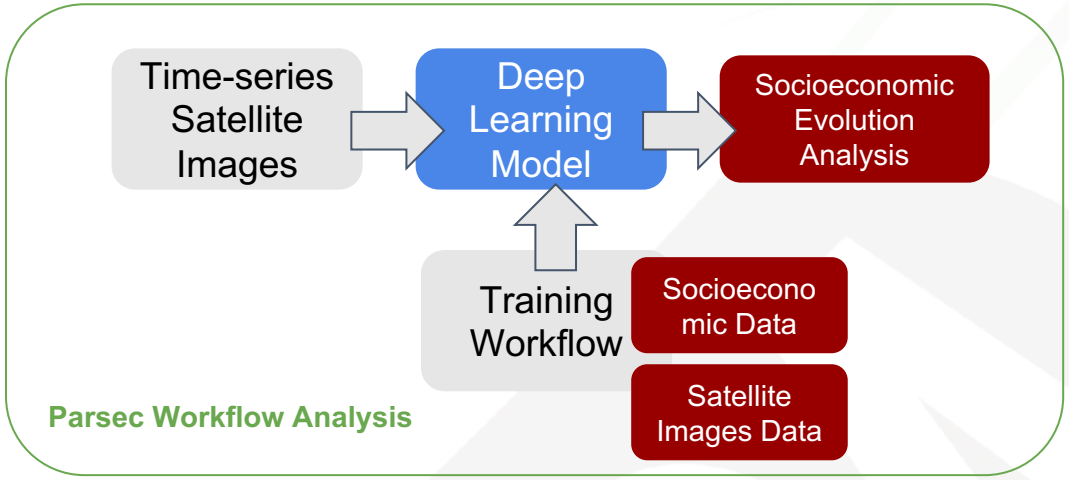
DataCite, ORCID, ESIP, RDA, EDI, WDS, AST, JWP, TNC

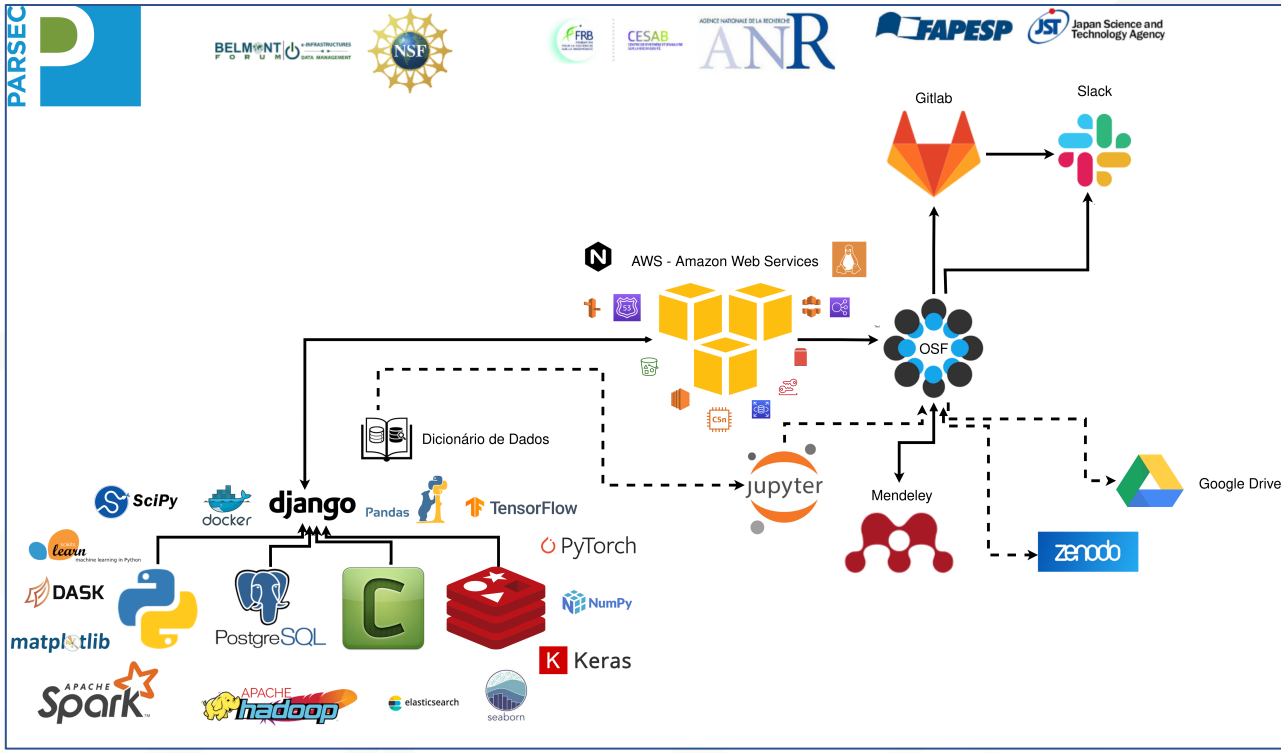
https://www.rd-alliance.org/plenaries/rda-17th-plenary-meeting-edinburgh-virtual/making-global-connections-earth-and

PARSEC - Computational Infrastructure

Synthesis Model (Spetcht, 2015)





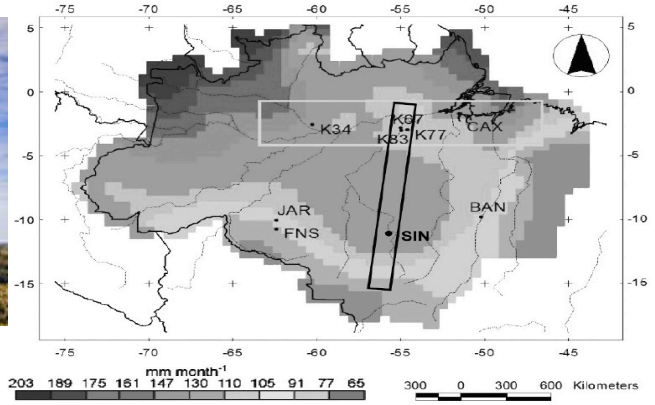


Further details can be found on the process and methods used for PARSEC:

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

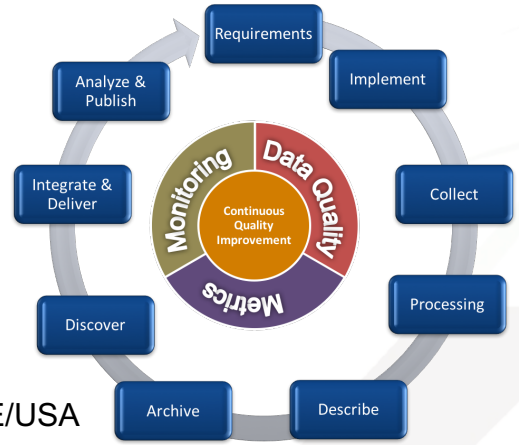
Greenhouse Gas Control - Synthesis

- Objective: Data Science Information System and Support Services for Modeling and Analysis of Greenhouse Gas Control Processes in the Amazon.
- FAPESP/SHELL Project (2022 – 2025)
- Data sources :
  - 10 flux towers data
    - Available at https://daac.ornl.gov
    - Amazon Tall Tower Observatory
      - (Max-Planck-Gesellschaft)
    - Data available in LBA infrastructure (ftp☺ - 2016 – now)
  - NOAA-USA, AERONET-USA
- Challenge: Create a Computational Infrastructure based on service(cloud) computing to remote access repositories to process data using Machine Learning Algorithmics

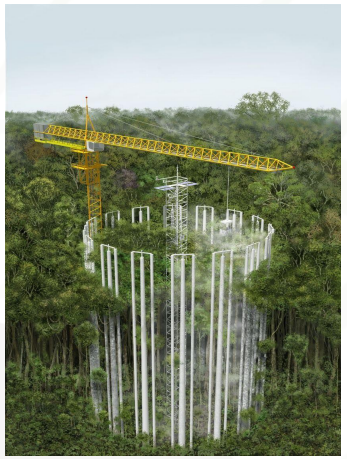


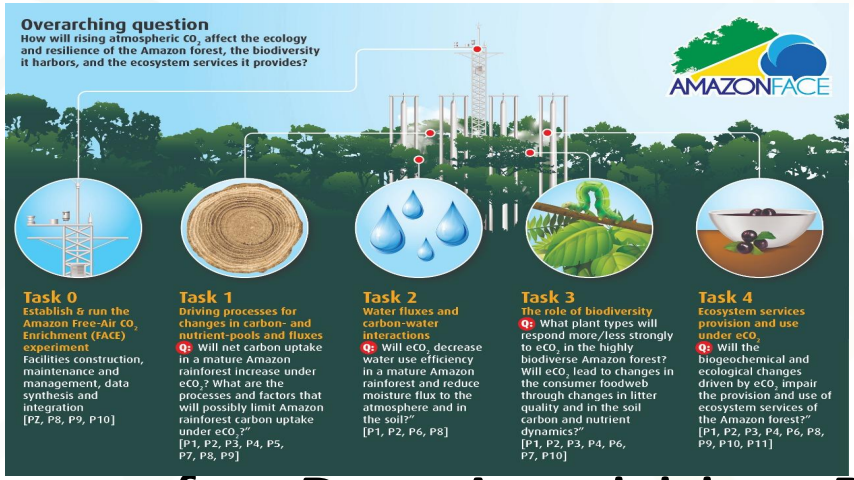
AmazonFace – Data Acquisition

- Objective: understand the future of the forest based on climate change, deforestation, forest degradation and fires.
- Joint Project: Brazil, Germain and USA. Start in 2022.



ARM/DoE/USA





- Main focus on develop an infrastructure for: Data Acquisition, Data Quality, Data Publish and Data Publish

General Recomendations to improve FAIR in Brazil

- Be transparent in methods, platforms and infrastructure - Clear Data Policy ...
- International collaborations, USGS, ORNL, NCI, (Workshops, visiting researchers), active participation in international projects, international forums and Communities (RDA, CODATA, ESIP, AGU)
- Be prepared for the transition to the next generations of hardware
- Continuous understanding of nature in our business model and value chain;
- Multidisciplinary teams, also involving areas of knowledge in Computing and Information Science
- International collaboration in research, software development, training and qualification of people - Engaged more people!
- People don’t scale, systems do

Making Global Connections in Earth and Environmental Science Data Infrastructures and Repositories

April, 21st – 2021

Earth and Environmental Science Data Infrastructures in Brazil

Prof. 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**

https://www.rd-alliance.org/plenaries/rda-17th-plenary-meeting-edinburgh-virtual/making-global-connections-earth-and