



HAL
open science

Challenges in data and software management and sharing at the global level: the case of the PARSEC project.

Romain David, Corrêa Pedro Luiz Pizzigatti, Rorie Edmunds, Danton Ferreira Vellenich, Ana Heredia, Jeaneth Machicao, Laurence Mabile, Yasuhiro Murayama, Margaret O'Brien, Solange Santos, et al.

► To cite this version:

Romain David, Corrêa Pedro Luiz Pizzigatti, Rorie Edmunds, Danton Ferreira Vellenich, Ana Heredia, et al.. Challenges in data and software management and sharing at the global level: the case of the PARSEC project.. RDA's 16th Plenary - remotely, Costa Rica, "Knowledge Ecology", Nov 2020, San José, Costa Rica. , 2020, 10.5281/zenodo.4276640 . hal-03014752

HAL Id: hal-03014752

<https://hal.science/hal-03014752>

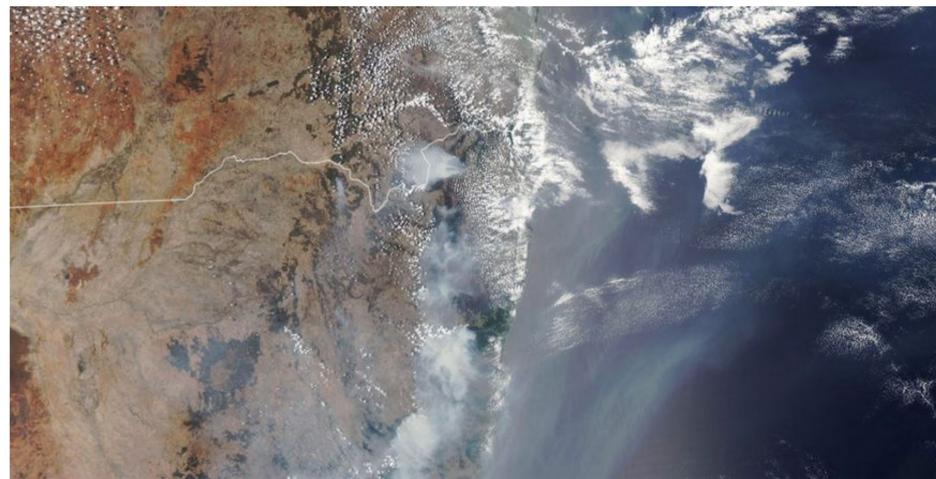
Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



PARSEC objectives:

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

The PARSEC project is designed to provide a unique opportunity for data and synthesis scientists to collaborate and exchange in real-time toward the goal of **improving research outcomes, data sharing, and data reuse**. It will also help pioneer new scientific and data science technologies aimed at improving both the **management and conservation of global biodiversity**.

We have two interactive teams in this project: a **Synthesis Science team** and **Data Science team**.

- The Synthesis Science team is employing **artificial intelligence techniques to analyse satellite images and socio-economic information** to better predict and mitigate the effect(s) of the creation of Protected Areas that may potentially threaten the livelihoods and health of local (indigenous) communities. Like most researchers who investigate complex environmental problems, the team *depends significantly on the availability of high quality, spatially dispersed, multidisciplinary, and time-series data*.
- The Data Science team of leading environmental data management professionals representative of data communities (RDA, ESIP), societies (AGU), and infrastructures for data attribution (e.g., DataCite and ORCID) are working to **develop leading practices on data citation, attribution, credit, and reuse**. Using an integrated work with the synthesis science team as a *case study of the wider scientific community*, the data-science team will provide a review of best practices for data management and stewardship *to optimise data access and reuse*. The team will also develop and implement a new tool to better track data usage and reuse.

Dealing with international complexity

PARSEC comprises over 30 team members from six countries, speaking 4 languages. There are 4 PIs, and a lead researcher representing each country. PARSEC team is disciplinarily diverse, with ecologists, statisticians and modellers, economists, spatial scientists, data scientists and experts in artificial intelligence. An overarching challenge for the project is sharing what we learn and develop with the broader research community. Engaging with working groups within the RDA is one way to get additional feedback.

Opportunities to work directly with research teams, such as those funded by our grantor, the Belmont Forum, will be an important first step. This poster **articulates the technological challenges prompted by the many aspects of team diversity and physical location as well as the cultural challenges**. We are seeking better understanding and we embrace the value of *different team dynamics, methods of communication, and preference to research approach*.

Planned outcomes

- To be able to determine the **influence of natural protected areas on consumption expenditure and asset health of rural communities** in a range of countries;
 - To have **refined recommendations for the research data workflow** and skills for research teams;
 - To have developed **tools for researchers** to view how the data they have deposited are used and cited. This should **improve the number of citations to data sets and better attribute them to the data creator**;
- To have contributed to the **promotion of FAIR, CARE and TRUST** through authentic examples from an active research project.



You reap what you sow!

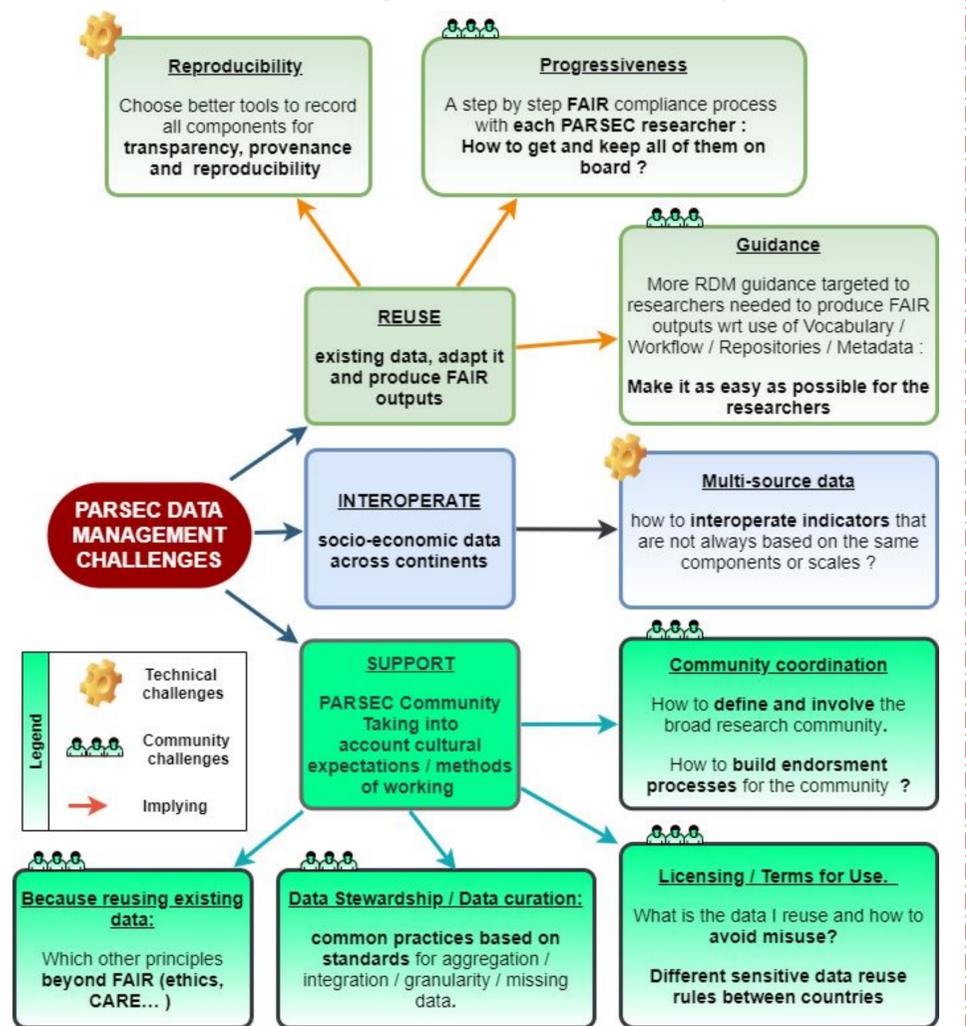
A garden that is planned, cultivated and managed well will yield more produce than one where you chose the wrong plants, forgot to mark what seeds were planted where, didn't fertilise or water regularly, and allowed the weeds to take over.

Like watering and weeding, a little, *consistent work throughout the research life cycle* will result in data that are of value to others and able to be shared as part of any research output. In PARSEC, we value scientific data as an important contribution to scholarly work. We are taking careful and measured steps towards ensuring our data are **well managed, FAIR and capable of having an enduring role in the scholarly ecosystem beyond the lifetime of our project**.

Data Strand in PARSEC

In the PARSEC project, data science experts are partnering with a multi-country synthesis science research team to **build relevant tools and processes for better data and software management** that integrate into the research lifecycle and are to be shared with the wider research community. In recent months, we have seen important papers and books published that support this larger goal of helping researchers incorporate data and software management during their whole lifecycle, but there is still much work to be done to **develop the methods and foster the cultural change** needed to implement them in practice across scientific communities. The main challenges identified so far can be grouped into (a) **technological ones** and (b) **cultural ones**. We would like to share these considerations and work collaboratively on incremental improvements that would be beneficial for all.

PARSEC challenges to meet data strand objectives



PARSEC roadmap proposal to meet data strand objectives

- Step 1: **Inventory** success stories / landscape / good practices / tools / RDA or other fora recommendations
- Step 2: **Collect the needs** of researchers / how to **support** them
- Step 3: Propose practices to researchers / validation by the whole community / training material development
- Step 4: **Train and support** research teams, **adapt best tools** to researchers needs
- Step 5: *Asses efficiency* to resolve issues and ensure sustainability of applied practices

Participating countries

- BRAZIL:** University of São Paulo – The São Paulo Research Foundation, **FAPESP** (P. Pizzigatti Corrêa) plus postdoc and technical support (FAPESP)
- FRANCE:** Foundation for Research on Biodiversity (FRB), University of Toulouse – INSERM, FR), <https://orcid.org/0000-0002-7724-1721>, Jeaneth Machicao (University of São Paulo, BR), <https://orcid.org/0000-0002-1202-0194>, Yasuhiro Murayama (National Institute of Information and Communications Technology, JA), <https://orcid.org/0000-0003-1129-334X>, Margaret O'Brien (University of California Santa Barbara, USA), <https://orcid.org/0000-0002-1693-8322>, Solange Santos (Scientific Electronic Library Online), <https://orcid.org/0000-0002-2713-6202>, Alison Specht (The University of Queensland, AU), <https://orcid.org/0000-0002-2623-0854>, Shelley Stall, American Geophysical Union, USA), <https://orcid.org/0000-0003-2926-8353>, Danton Ferreira Vellenich (University of São Paulo, BR), <https://orcid.org/0000-0002-3223-6996>, Lesley Wyborn (Australian National University, AU), <https://orcid.org/0000-0001-5976-4943>.
- JAPAN:** National Institute of Information & Communications Technology (NICT), Research Institute for Humanity and Nature (RIHN) – JST (Y. Murayama)
- USA:** American Geophysical Union (AGU) – NSF (S. Stall)

* First data strand PARSEC references:

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A. and Research Data Alliance – SHaring Reward and Credit (SHARC) Interest Group, T.R.D., 2020. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. Data Science Journal, 19(1), p.32. DOI: <http://doi.org/10.5334/dsj-2020-032>

Acknowledgements:
 PARSEC is funded by the Belmont Forum through the National Science Foundation (NSF), The São Paulo Research Foundation (FAPESP), the French National Research Agency (ANR), and the Japan Science and Technology Agency (JST).