**The rescue and synthesis of a national soil dataset across agricultural land in Mexico**

**lessons learned and best practices**

Ortega Mariana1, Todd-Brown Kathe1, Guevara Mario2

1Department example, University example, city, postal code, country

2Laboratory example, city, postal code, country

3Centro de Geociencias - Universidad Nacional Autónoma de México

Campus Juriquilla, Qro. MX.

*Correspondence to*:

**Important Links (and dataset accompanying files):**

**Dataset:** [**https://docs.google.com/spreadsheets/d/10AOAMb7r-fV1H1mDbv2ZRfuFsNSo1cGG1QbAIITJkWI/edit?usp=sharing**](https://docs.google.com/spreadsheets/d/10AOAMb7r-fV1H1mDbv2ZRfuFsNSo1cGG1QbAIITJkWI/edit?usp=sharing)

**Methodological document (in Spanish)**

[**https://docs.google.com/document/d/1lwtBJr2P4e-N52PVF5EYRlmdPWUqhUJAT0EvbZ16tAo/edit?usp=sharing**](https://docs.google.com/document/d/1lwtBJr2P4e-N52PVF5EYRlmdPWUqhUJAT0EvbZ16tAo/edit?usp=sharing)

**Project presentation (in Spanish)**

[**https://drive.google.com/file/d/1\_Hn17mv4bWJvakfqryqjZd\_0bCS9WwKN/view?usp=sharing**](https://drive.google.com/file/d/1_Hn17mv4bWJvakfqryqjZd_0bCS9WwKN/view?usp=sharing)

**Abstract.**

# 1 Introduction

* DATA ARCHIVING PRACTICES AND MIRRORING DATA IN MULTIPLE SERVERS

Data management practices in soil science are fundamental for ensuring the reproduction of scientific data and results.

* RELIABILITY OF DATA REPOSITORIES AND IMPORTANCE OF ARCHIVAL PRACTICES
* REPRODUCIBILITY
* ECONOMIC IMPLICATIONS
* FAO GROUPS
* There is a pressing need for soil data and soil information for a wide diversity of users to make informed decisions about land management across agricultural soil.
* In Mexico, increasing the quantity, quality and access to soil information is particularly relevant as soil degradation is present in >60 % of territory.
* While Mexico can be considered a rich data country, the large environmental heterogeneity makes the available information scarce and disperse
* We present the synthesis of a national dataset collected in 2012 by a coordinated effort between FAO and the Secretary of Agriculture of Mexico.
* This dataset was identified on the internet by the audience of the Global Symposium on Soil Organic Carbon organized by FAO in Rome in 2017
* The dataset was downloaded from the internet in 2017 and stored in a hard drive. After a while the downloading site was shut down.
* Our objective is to make available this information again and to provide a national synthesis of agricultural soil information as a demonstration of the usability and applicability of this national soil dataset.

## 2 Dataset description,

**The rescued dataset was generated by SAGARPA (Secretaria de Agricultura, Ganaderia, Desarrollo Rural, Pesca y Alimentación) with the purpose of contributing with a national baseline of a natural resources sustainability program (NRSP) supported by the FAO and focused on agricultural land (SAGARPA, 2012).** The objective of developing this dataset was to acquire contrafactual datasets to support the NRSP instrumentation. Thus, this dataset was planned to represent relevant soil quality indicators (physical, chemical and biological soil property values) to support informed decisions managing agricultural land.

Previous research shows that the rescued dataset has been potentially used by other scientists before the site was shut down as we have found citations of the supporting information in scientific papers (Azañero et al., 2020) and thesis (Rivera, 2018). The supporting information is currently available in the internet and it will appear immediately in some downloading sites after searching the name of the document (‘Subíndice de Uso Sustentable del Suelo – Metodología de Cálculo SAGARPA’) in any internet explorer. However there are no major insights about the availability of the actual dataset or nor official information associated with NRSP. While the team in charge of developing this dataset is known thanks to this supporting document, the work was limited to developing the dataset and not to maintain it in a stable repository site. This experience highlights the pressing need of adopting optimal data archiving and mirroring practices for securing large investments for collecting soil datasets.

Azañero Aquino, L., Ñique Alvarez, M., & Florida Rofner, N. (2020). Calidad del suelo en diferentes sistemas de uso en selva alta de Huánuco, Perú. *Revista Tayacaja*, *3*(1). <https://doi.org/10.46908/rict.v3i1.75>

Thesis of Rivera: <http://repositorio.unas.edu.pe/bitstream/handle/UNAS/1555/RCDA_2019.pdf?sequence=1&isAllowed=y>

2.1 Data dimensions and format

2.2 Variables included

2.3 Data collection methods and units

## 3 Dataset preparation and preprocessing

## 3.1 Database format, (e.g., translation)

3.2 Quality control of data values and spatial integrity (e.g., spatial coordinates verification)

3.3 Statistical description of data values

## 4 Practical application (Dataset analysis)

….

**5 General discussion**

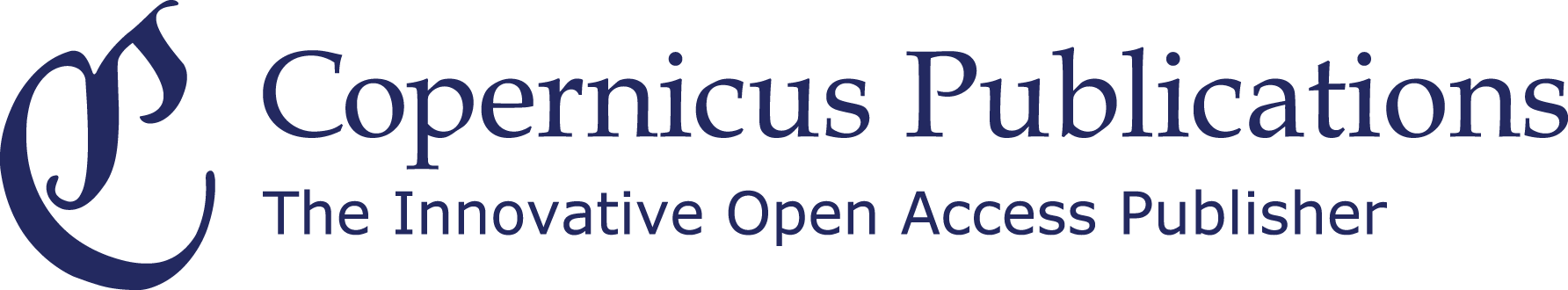
**6 Data availability (DOI from EDI)**

**7 Conclusions**

# References

Miller, B. B. and Carter, C.: The test article, J. Sci. Res., 12, 135–147, doi:10.1234/56789, 2015.

Smith, A. A., Carter, C., and Miller, B. B.: More test articles, J. Adv. Res., 35, 13–28, doi:10.2345/67890, 2014.



**Figure 1: The logo of Copernicus Publications.**