OpenSoils: e-Science em Segurança de Solos

Sérgio Serra^{*}, M.B. Ceddia^{*}, R. J. Macário Costa, G. Rizzo, <u>Renan Carvalho Távora Miranda</u>, P.V. Cruz, E. Marinho, E. A Schitmz, S.S.C. de Oliveira, A. C. C. da Silva e F. Klinger



Agenda

- Introduction
- Soils Security
- Reasons to do data-centric research in soils
 - Open Data & Data Provenance
- OpenSoils Framework
- Use Cases
- Concluding Remarks





Introduction

- By 2050 world will have about 9 billion individuals.
 - Researches should grow more/better food without exacerbating environmental problems (climate changes and rational use of **soil resources**).
- To Koch et al. [1], the **soil crisis** that eclipses those of the past.
 - The soil degradation is a critical and growing problem.
 - The **pressures on soil are diverse**, the challenges demand new approaches to provide sufficient resources for the world's growing population.
 - Up to now, soils security investigations did not draw the same degree of attention from other data-intensive disciplines (e.g. bioinformatics and engineering).



Soils & Soils Data

- Soils are probably the most critical natural resource and biosystem that support the human and terrestrial life.
- Soils data deserves better efforts to improve its data collection, longterm preservation and data reuse [5].
- Nowadays, huge volumes of new soils data are being produced very quickly due to the automation of data collection (e.g., field experiments, UAV, sensors, satellites).

Soils & Soils Data

• Soil data are heterogeneous

- Stored as isolated data silos (e.g., articles, spreadsheets, text files)
- Poor semantics, lacking provenance and others metadata descriptors
 - Morphological data
 - Chemical Data
 - Physical Data
 - Hydrologic Data
 - Mineralogical Data
 - Sensors Data
 - Environmental Data



Soils Security

- Soils security is an evolving concept motivated to respond to the **problems that face the global soil stock**.
- Soil security is related with the maintenance and improvement of the global soil resource to produce food, fibers and fresh water, carbon sequestration, contribute to climate sustainability, and to maintain the biodiversity.

Reasons to do data-centric research in soils security

Problems	Consequences
1- Limitations of the Brazilian Soil Database (BDsolos)	No metadata descriptors, no public interfaces to allow the insertion of new data, poor interfaces to query data are hard to be used;
2- Disperse datasets organized in spreadsheets/or specific database with few soil attributes (data silos)	It hampers the global view of soil information and the relationships between attributes and soil types, as well as those with other areas of knowledge;
3 - In general, both the data provenance and the curation of the data is lacking	The users have to check not only wheter the data is repeated but also whether it is valid for use;
3- The data available is not open	The available data fails to fulfill the eight Open Data and also the FAIR principles
4- Lackness of an efficient system to collect new data in situ.	New data continues to be organized manually or through spreadsheets . There are no Apps for data collection in the field nor to integrate them with new soil proximal sensors.

Data Provenance in Soils Security

- Data provenance includes the data's origins, what happens to it and where it moves over time.
- Data provenance gives visibility while simplifying the ability to trace errors back to the root cause in a data analytics process.
- The scope of the data provenance determines the volume of metadata required to represent its lineage.
- Data provenance provides the audit trail of the data points at the highest granular level,
 - presentation of the lineage may be done at various zoom levels to simplify the vast information.

Open Data in Soils Security

- Open data is the idea that some **data should be freely available*** to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.
 - ✓ Complete
 ✓ Primary
 ✓ Timely
 ✓ Accessible

✓ Machine processable
 ✓ Non-discriminatory
 ✓ Non-proprietary
 ✓ Open reuse license

* Adapted from http://wiki.opengovdata.org/index.php?title=OpenDataPrinciples



OpenSoils Framework

- OpenSoils is an open, elastic, provenance-oriented and lightweight computational open science e-infrastructure which rely on four overarching layers.
- The database stores data from soils profiles and boreholes
- Each soil profile description recorded in the database has more than 40 entities, and 250 attributes to stores the soil properties and soil experiments (*e.g.*, mineralogical, morphological, chemical, physical, fertility, and environmental data, others).

OpenSoils Framework



Layer	Description
End-User Services	Hosts a web portal and web/mobile applications used to ingest and query soil data; Data collection in situ. Support soil specialist, researchers, farmers and decision makers.
Soils Services	Implements the scientific and business models of soils security in cloud-based environments; Implements as scientific workflows; API and Web services to transform and harmonize data; Add traceability of soil samples and provenance of the experiments in the e- infrastructure; Increase the reproducibility of Soil Security experiments;
Data Services	Stores data sets from soils profiles and boreholes and metadata descriptors. Uses relational databases to store operational data and governance descriptors [12, 13]. Uses of FAIR principles [16] and open repositories (e.g., CKAN, DSpace) to store curated data as open data sets. At this moment, the database schema and its standardization and harmonization aim to serve high quality-assessed, georeferenced data to the Brazilian community, however, it is being expanded to attend other communities.
Governance Services	Hosts analytical tools and visualization services that can be connected to other software (e.g., QGIS, R, Tableau or sci-kit-learn) to generate analytical reports, predictions and raster maps;

OpenSoils App (Layer 1)

Insert data



Old School

query data







Concluding Remarks

- The soil has an integral part to play in the global environmental sustainability challenges.
 - There is still a lot of computational work needed to be fully developed in soil sciences.
- The growth of curated open soils databases may aid scientist/farmers/decision makers to increase the reliability, robustness, and reproducibility of soils security experiments.
- OpenSoils is being developed & tested....
 - Layer 4 (we are seeking new students)
 - Petrobras (Digital Mapping project)



Acknowledgments





Fundo Nacional de Desenvolvimento da Educação



Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro

- Students (heavy work!)
 - OpenSoils Team
 - Elton Marinho, Carolina Cursky, (PhD Candidates)
 - Jessica Mendes (Master)
 - Gabriel Rizzo, Renan Miranda, (PET-SI/UFRRJ)
 - Ana Correia, Sabrina Oliveira, Felipe Klinger (PET-SI/UFRRJ)
 - Pedro Vieira, (Agronomy)
- Collaborators
 - Prof. Marcos Ceddia (UFRRJ) Soils Security
 - Prof. Macário Costa (UFRRJ) AI







OpenSoils: e-Science em Segurança de Solos

Sérgio Serra^{*}, M.B. Ceddia^{*}, R. J. Macário Costa, G. Rizzo, <u>Renan Carvalho Távora Miranda</u>, P.V. Cruz, E. Marinho, E. A Schitmz, S.S.C. de Oliveira, A. C. C. da Silva e F. Klinger

