



# CocoaSoils Data Collection Tools & Ontology

Abidemi Elesho\*, Arun Pratihast, Ulan Turdukulov, Rob Knapen & Sander Janssen\*

## Open Data

Open and FAIR CocoaSoils data will have strong impact on cocoa production and research. The sharing of quality data requires the seamless integration of several working steps in a single workflow:

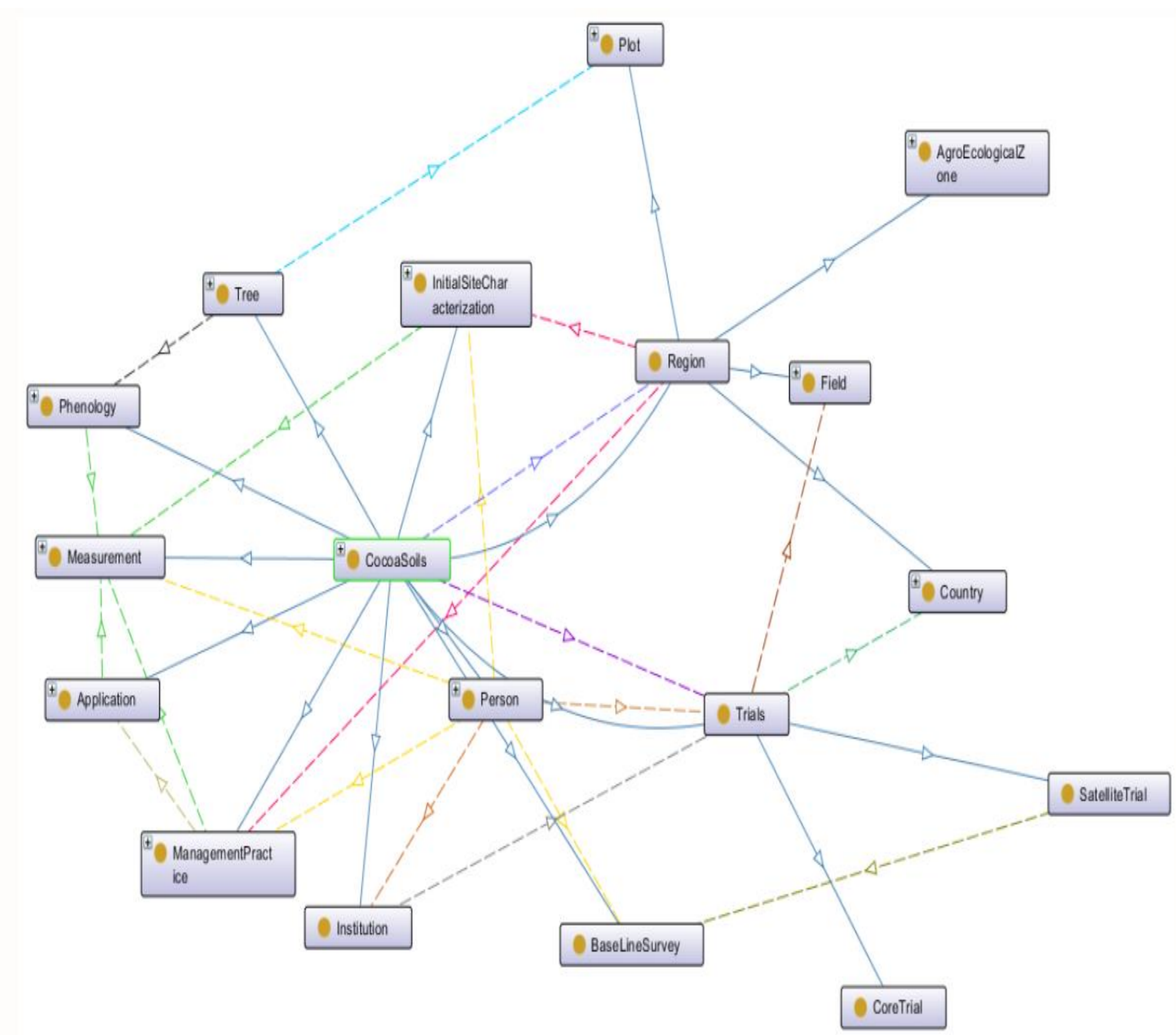
- Standardized and quality data collection
- Consistent storage and tailored analysis
- Easy data discovery and visualization

## CocoaSoils Ontology

- Refined and defined cocoa vocabulary
- Semantically enriched for easy discovery, analysis and visualization

## Common protocols and vocabularies

- Standardized naming of treatments and measurements makes the system robust, and will reduce data curation and processing efforts later on.
- We need to agree on uniform meaning of words in: treatments (& their units), measurements (& their units) in core and satellite trials.



## Data Collection Application

### ODK

- ODK aggregate server has been setup on <https://cocoasoils.wur.nl/ODKAggregate/Aggregate.html>
- ODK app with forms based on the protocols and common vocabularies allows efficient data collection with fewer errors using scanners and data review shortly after collection.

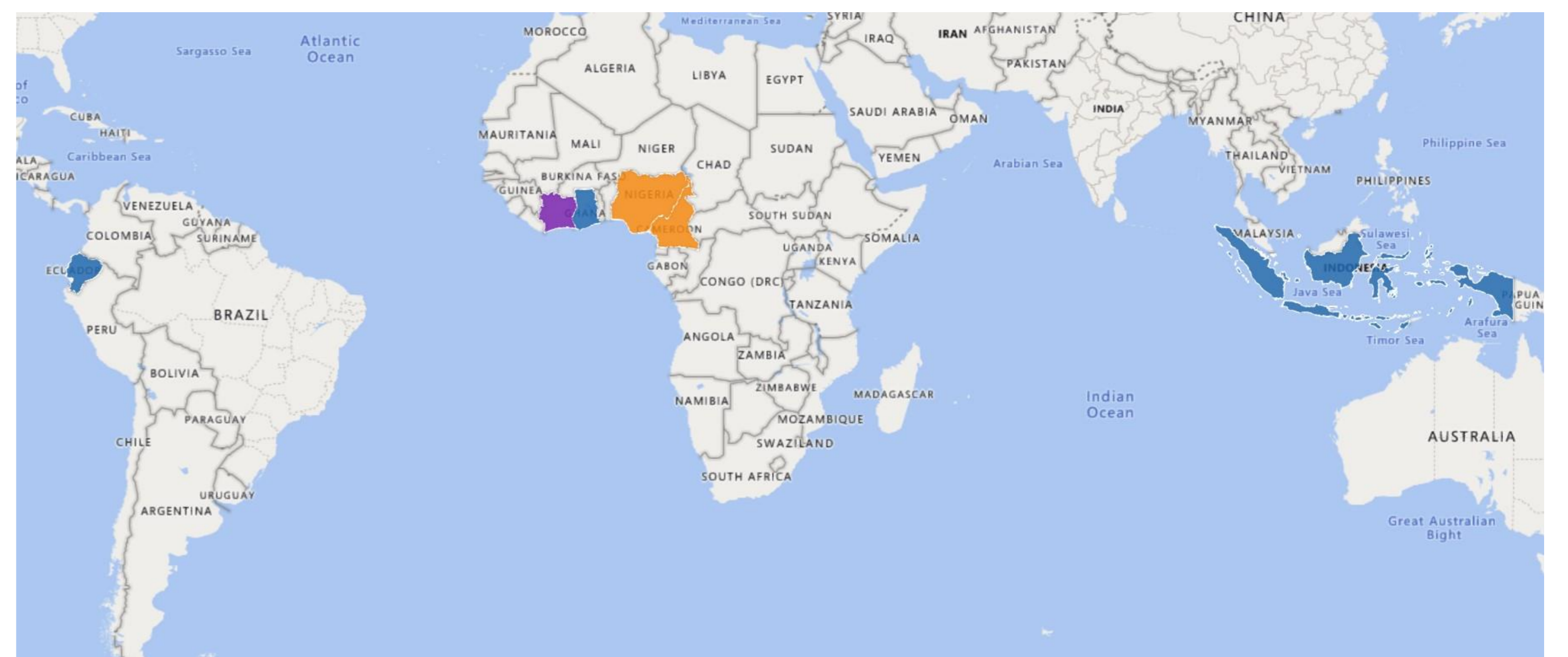
### Barcoding System

- Use of scanners results in fewer errors. Proposed labelling:
- Co => Core Trial
- CA => Country code
- 001 => Field ID
- 007 => Plot ID
- 01 => Tree ID

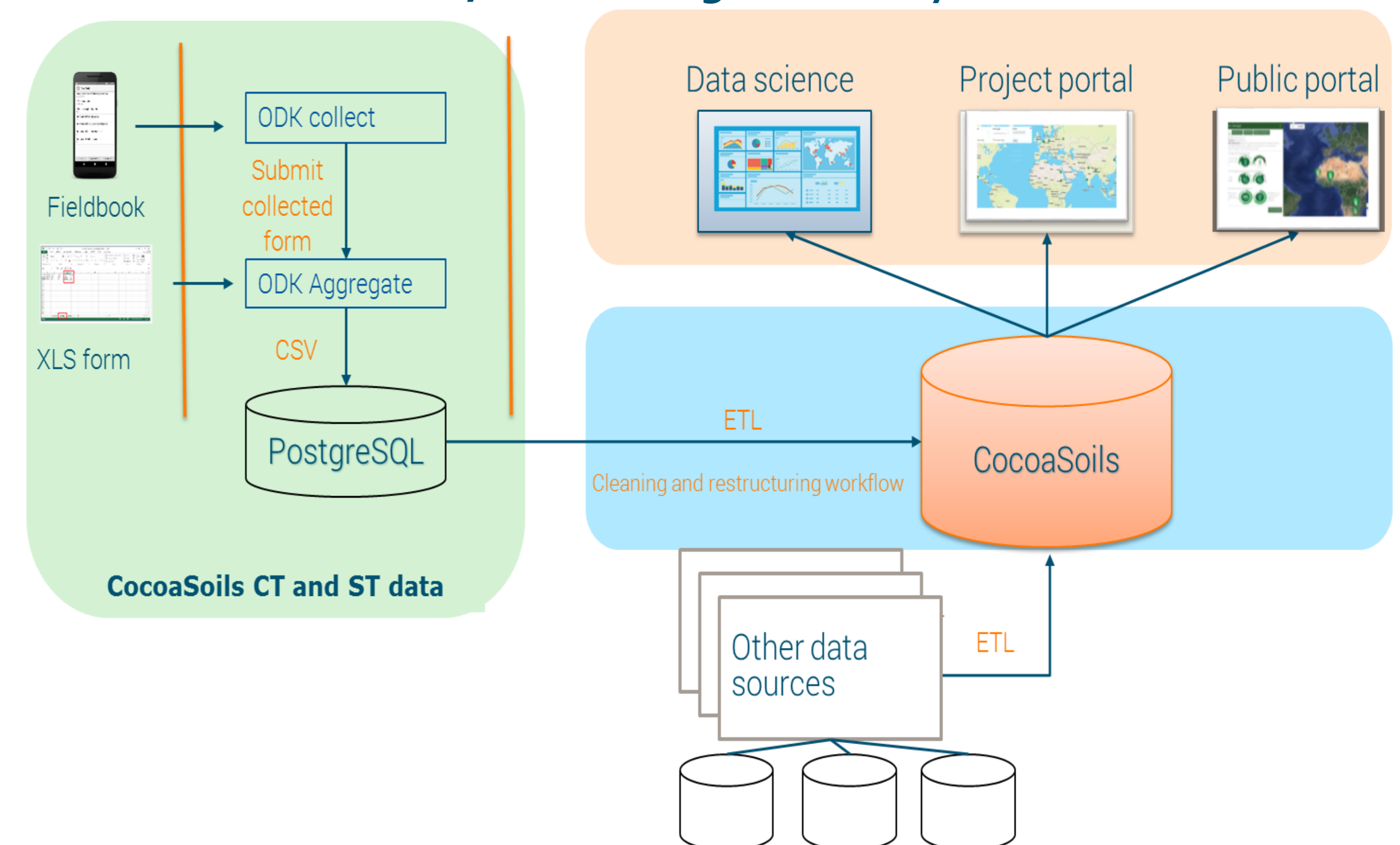


## Project Locations

The trial sites are hosted across three different continent (South America, Africa & Asia) and six countries (Nigeria, Ghana, Cote d'ivoire, Cameroun, Ecuador & Indonesia).



## Database Architecture, Data storage and analysis



## CocoaSoils Data portal

- Public facing portal.
- Scientific analysis facility: set of automated analytical and visualization procedures for CocoaSoils data (aka Dashboards).
- Project data management and sharing facility
- Data visualization platform - a facility to visualize in an attractive way the data for experts working in cocoa research and industry.
- Data discovery functionality – search through internal and external datasets using common metadata fields and descriptions.

