

---

# CocoaSoils data interoperability vision

---

WENR, Earth Informatics group

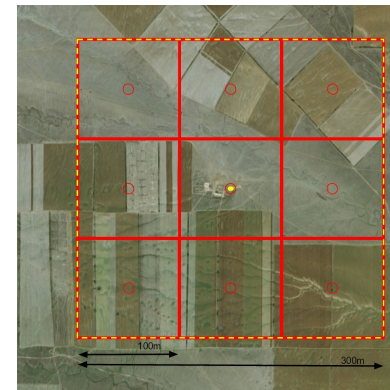
Ulan Turdukulov, Rob Knapen, Sander Janssen, Hendrik Boogaard,

Lotte Woittiez and Ken Giller

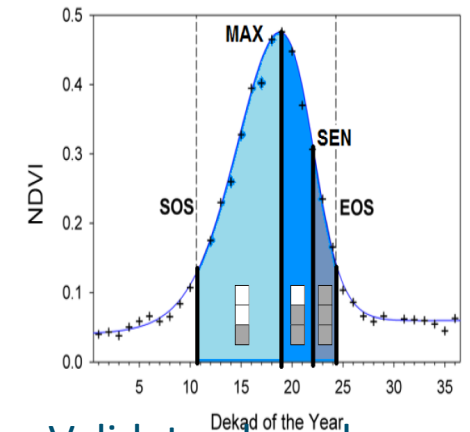


# Case for FAIR data

- Agronomy, food, plant sciences, and biodiversity can greatly benefit from the integration and re-use of the data that they produce.
- For instance, in-situ crop field data needed to improve accuracy and reliability of global and regional studies and monitoring systems through e.g. validation and/or calibration of crop growth models and satellite based products



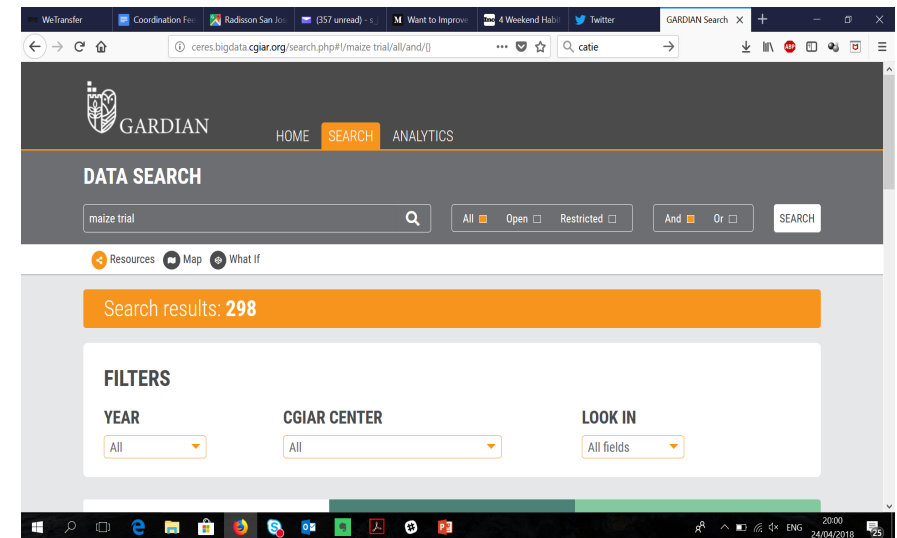
Validate crop map



Validate phenology metrics

# GARDIAN example

- Agricultural scientific community is fragmented in data management and lack commonly available reference data on (benchmarking of) agricultural production
- As results, published in-situ crop data is heterogeneous:
  - different attributes, locations, periods, lack of metadata and units,
  - requires extensive review of external sources such as paper, assumptions, ambiguities
- CGIAR are active in harvesting data ([GARDIAN](#)), published by researchers from across its 15 centres



---

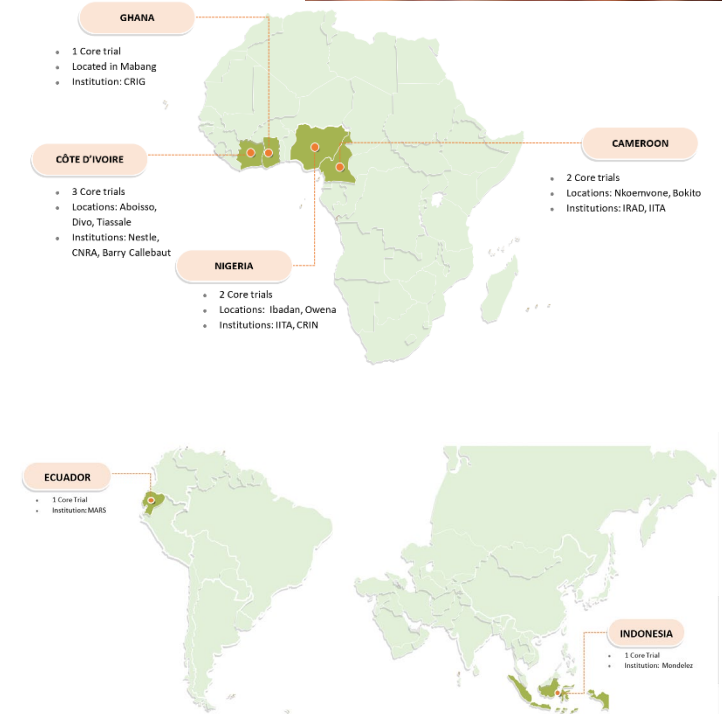
# Data curation (for AgroSTAC, modelling, etc)

---

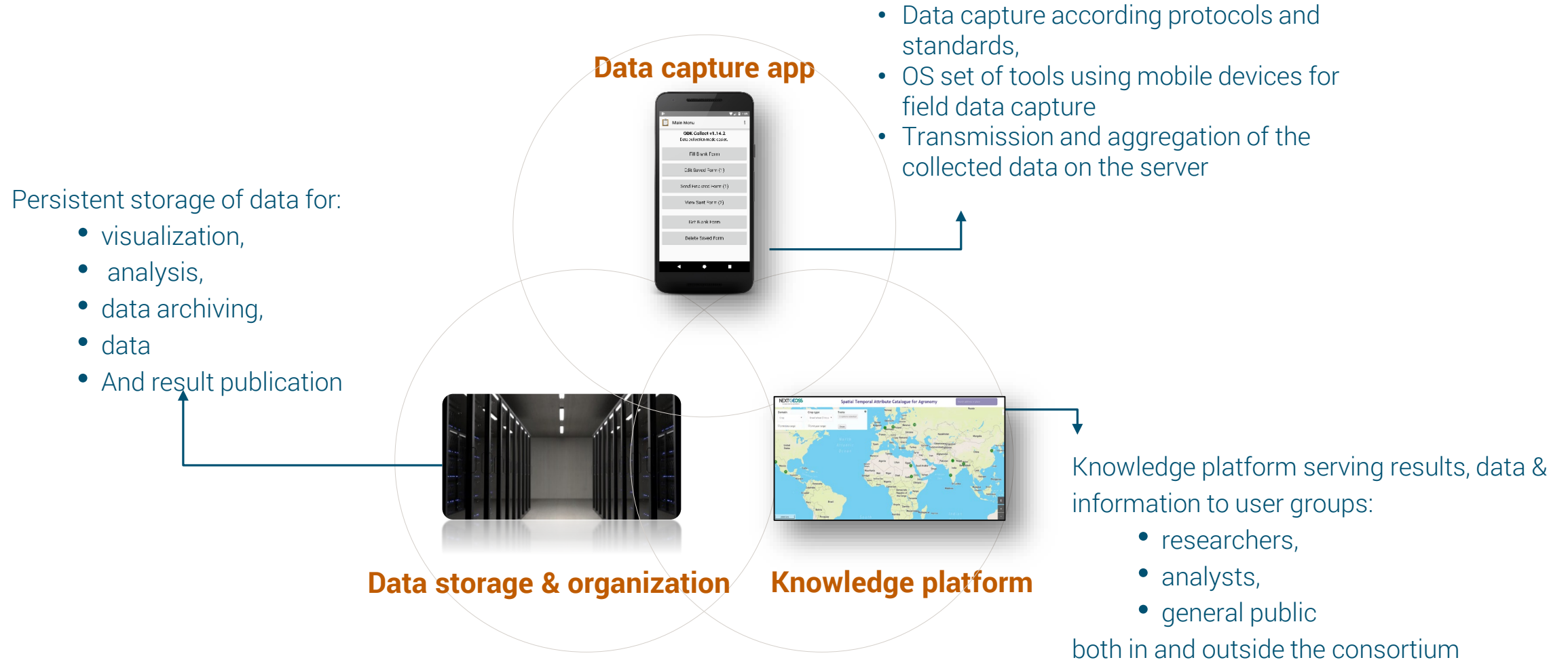
- Retrieve 42 from Gardian (Maize Trial, December 2017)
- Around 2/3 of the data sets were closed; contacted owners did not result in much
- Of the open data sets:
  - 11 were cleaned out and further made available
  - Many mapping problems to common vocabularies
  - Crucial meta-data often left out (for example, year of trial)
- Fieldbooks are crucial for 'easy' data curation
- ODJAR.org curated data sets were most easy for re-use



- Cocoa productivity in West Africa is declining
- The CocoaSoils project aims to improve the efficiency and sustainability of cocoa production by focusing on the Integrated Soil Fertility Management (ISFM).
- This knowledge is not yet available, and therefore the partners of the CocoaSoils project will set up number of trials: core, satellite

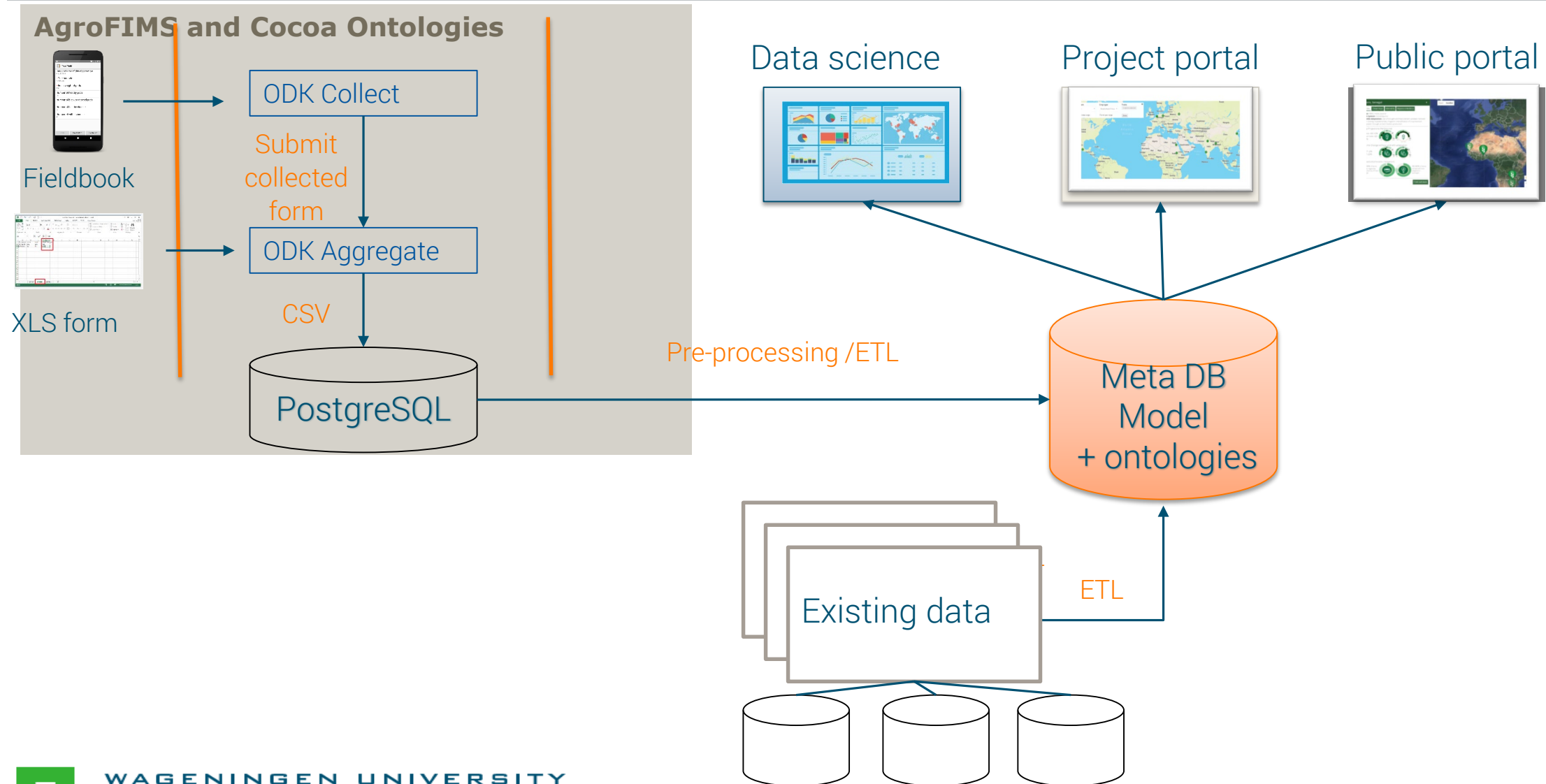


# WENR Earth Informatics role



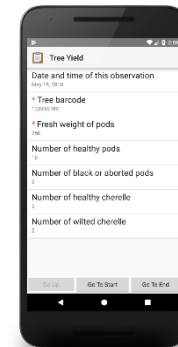
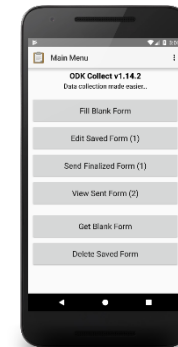


# Overall proposed architecture



# Quality data collection

- Incorporate previous experiences:
  - Peter Pypers (IITA) adopt data structure, architecture, data collection workflow in cassava
  - Chris Turnbull (University of Reading) – cocoa specific workflows
- Review ta collection tools (ODK, KDDart, FAIMS) and choose ODK (<https://opendatakit.org/> )



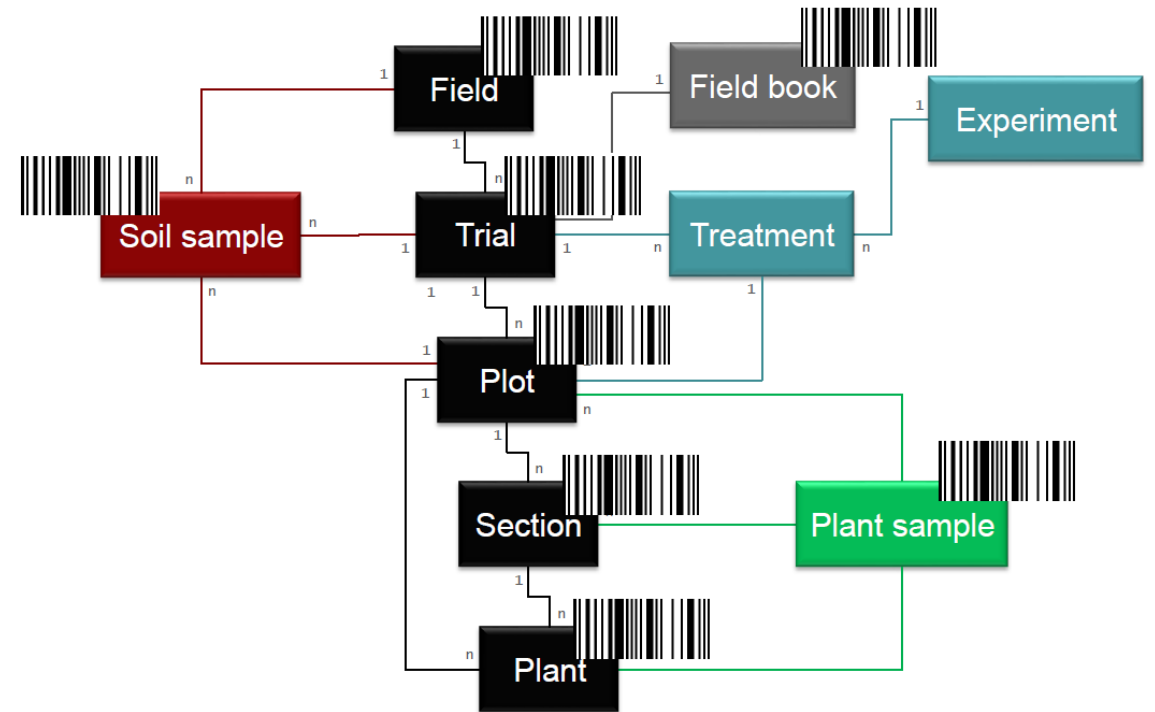


# Field data collection and organisation



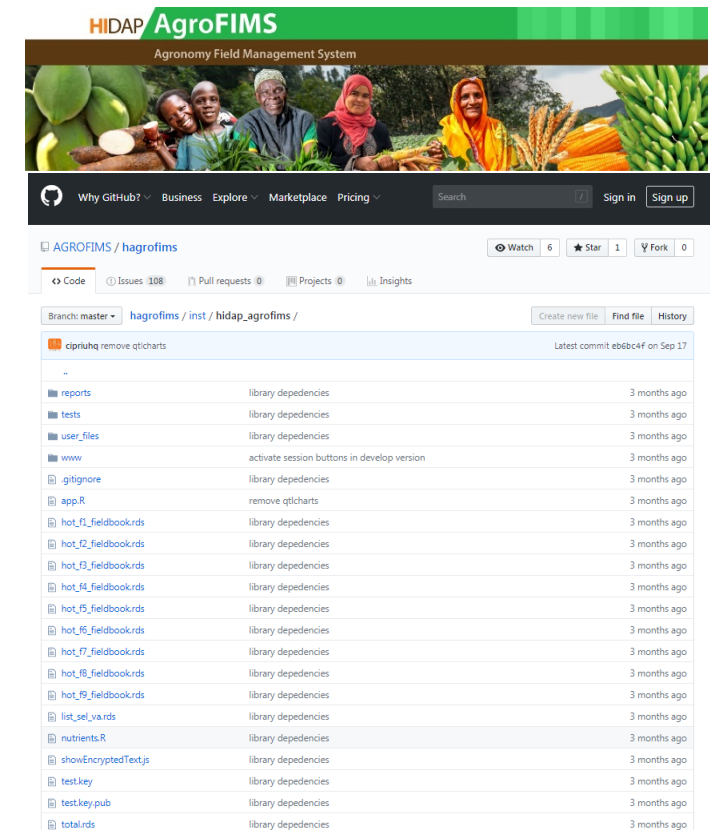
Project: AC = ACAI  
 Entity: TL = Trial  
 County: NG = Nigeria  
 seqNr: unique nr ( $10^6$ )

Abbr.	Entity
FD	Field
FB	FieldBook
PA	Plant
PS	PlantSample
PO	Plot
SE	Section
SS	SoilSample
TT	Treatment
TL	Trial



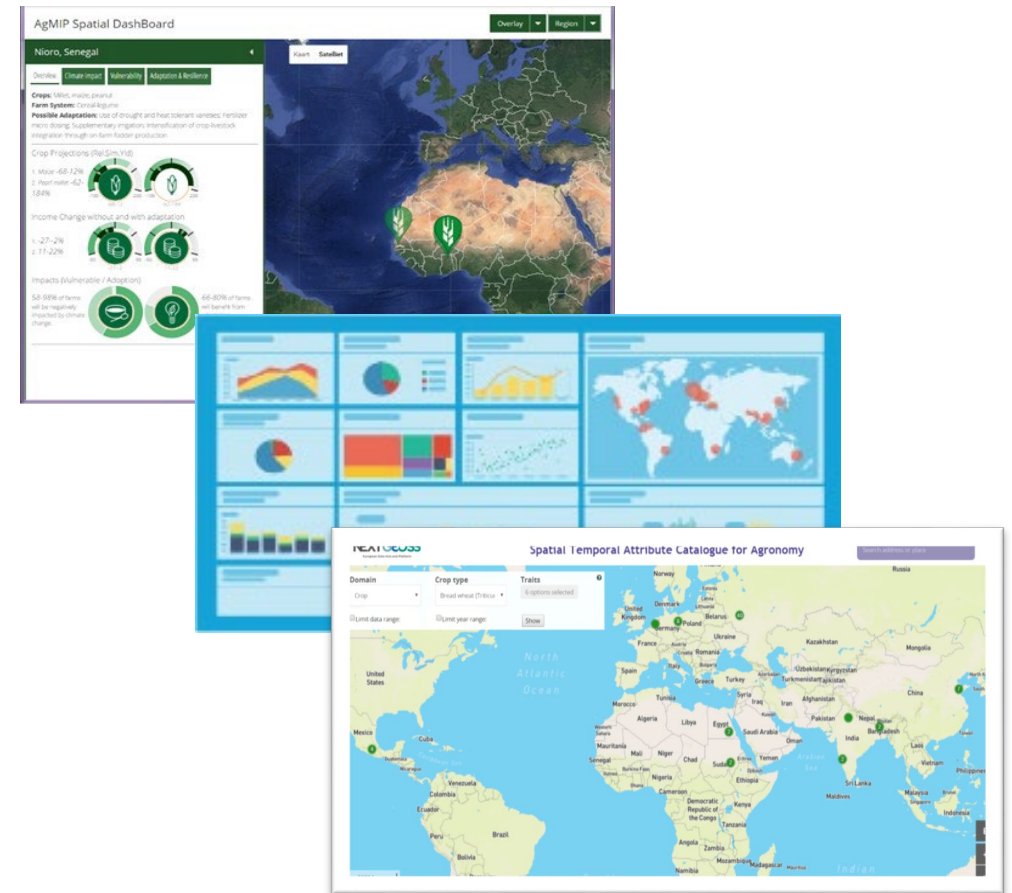
# Semantic enrichment

- AgroFIMS for Agronomy Ontologies
  - Relies on Environmental ontologies, Minimum Information About a Plant Phenotyping Experiment (MIAPPE), Breeding API (BrAPI)
- No cocoa specific ontologies – perhaps time to form Cocoa Data Interoperability group (similar to wheat) to make CocoaSoils datasets interoperable in order to enable value addition and data-driven innovation in cocoa research.



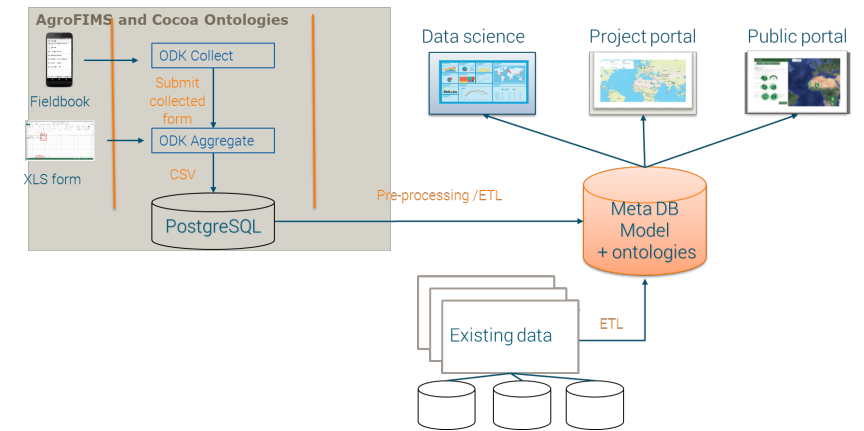
# Knowledge sharing portal

- Public facing portal
  - Accessible to all -first draft version is online
- Scientific analysis facility:
  - set of automated analytical procedures, data import, QC, export, analysis
- 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



# Concluding remarks

- Agronomy data is heterogeneous
- CocoaSoils is an opportunity to build data curation workflow: from quality data collection to semantic interoperability
- It is an opportunity to be involved in development of agronomy ontologies (AgroFIMS) and cocoa specific ontologies
- Chance to integrate cocoasoils data in wider research domains (i.e., crop growth modeling) through common set of vocabularies and ontologies



---

Feedback:

[ulan.turdukulov@wur.nl](mailto:ulan.turdukulov@wur.nl)

---

Thank you for your attention !

