

AgroDataCube and AgInfra+

Operationalising Big Data for Agricultural Informatics

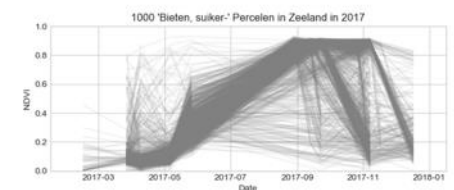
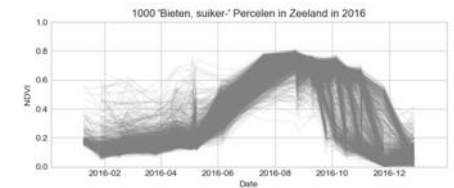
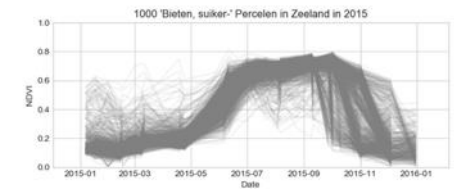
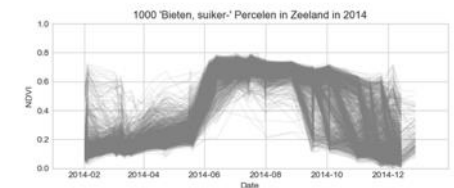
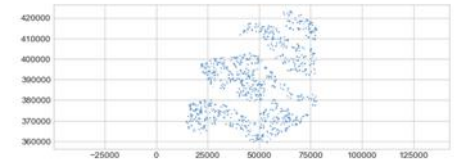
Rob Knapen, Rob Lokers, Yke van Randen, Sander Janssen, Henk Janssen

Wageningen Environmental Research, December 2018



Agricultural Informatics and Big Data

- Amount of available usable data increases rapidly
 - Remote Sensing – Drones
 - More satellite data – Sentinels
 - Precision agriculture – Sensors
- Technology to store and process the data is accessible
- Yet: Adoption is still in its infancies



(Our) barriers for adoption

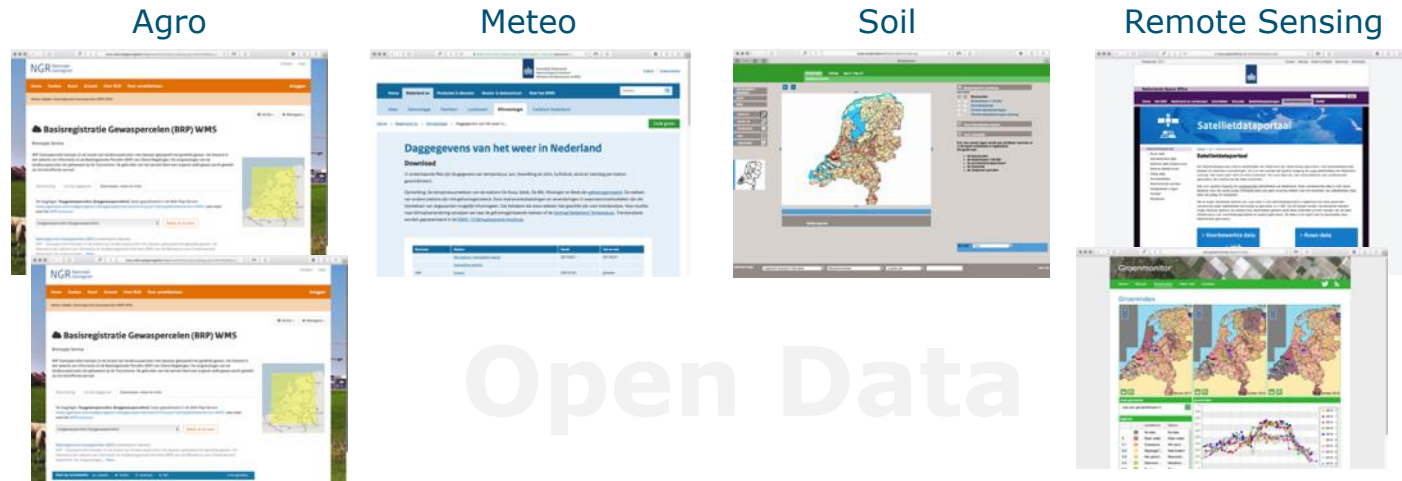
- Datasets live in isolation
- Datasets lack proper semantics
- Used to traditional, single node, data processing
- Invested in relational databases
- Historically GIS oriented

Using projects for improvement

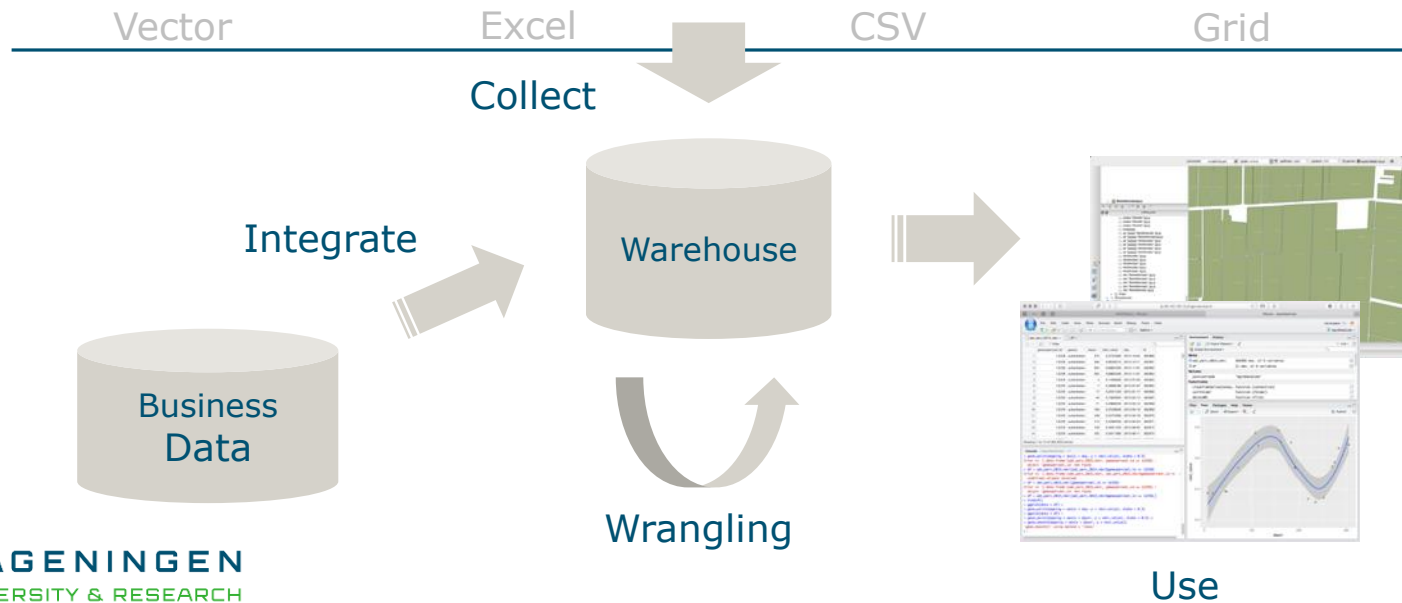
- AgroDataCube: agrodatacube.wur.nl
 - Building a Big Data platform for agriculture and food applications

- AgInfra+: plus.aginfra.eu
 - Exploit core EU e-Infrastructures for agriculture and food research communities

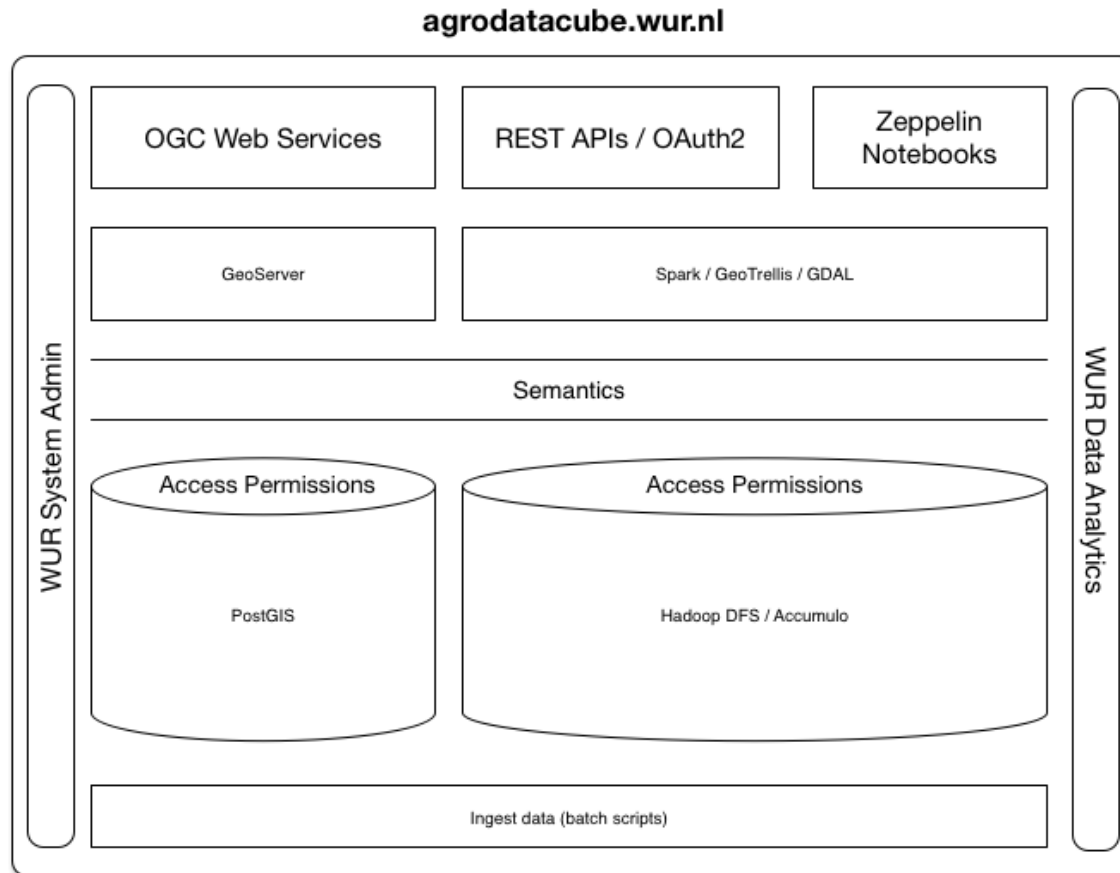
AgroDataCube



Open Data



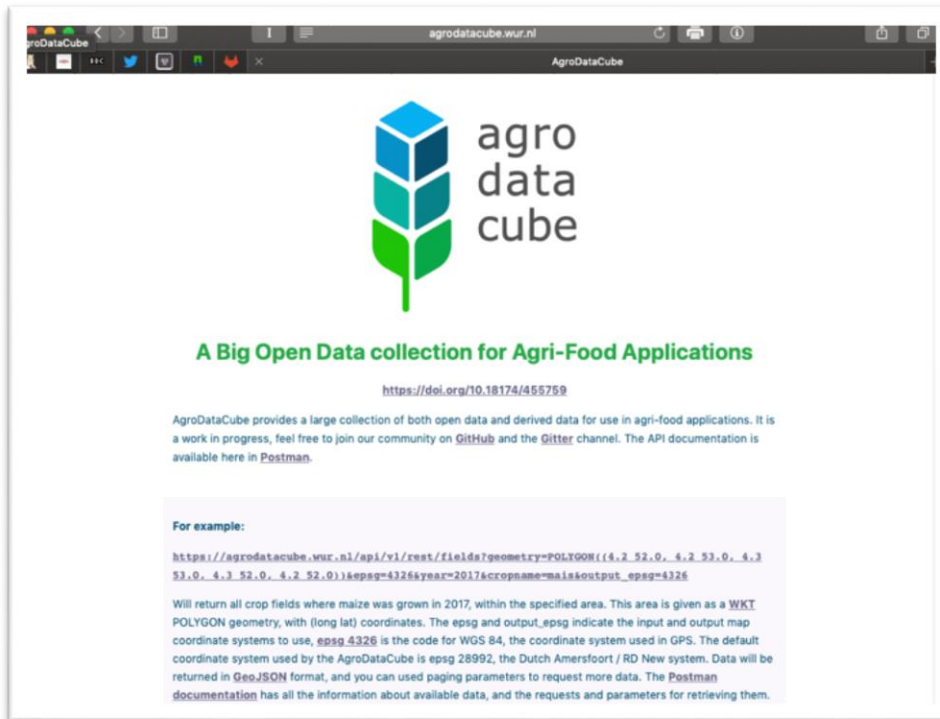
AgroDataCube



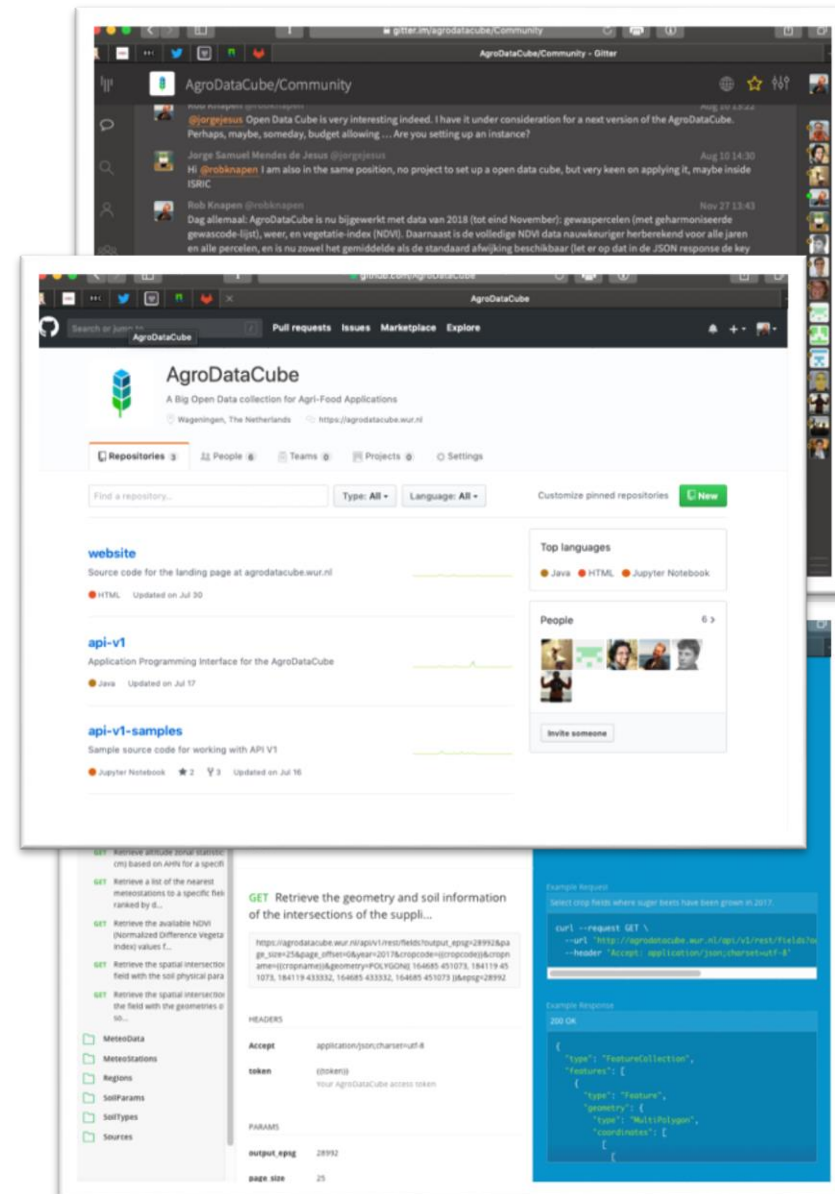
Open Data Sources (Excl. streaming data)



AgroDataCube – FAIR Data Science



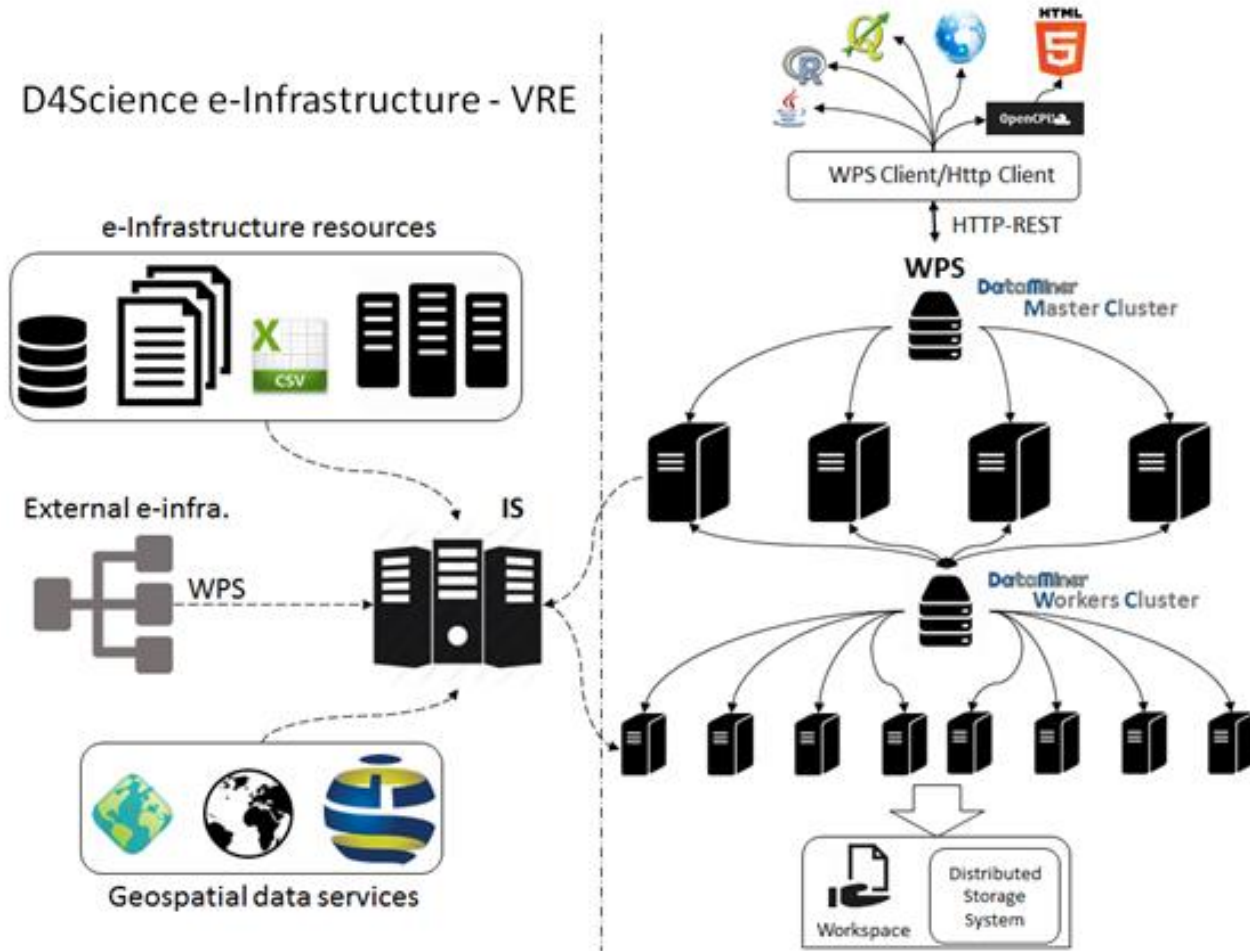
The screenshot shows the AgroDataCube website homepage. At the top, the logo consists of a stylized green and blue leaf icon next to the text "agro data cube". Below the logo is the heading "A Big Open Data collection for Agri-Food Applications" and a DOI link: <https://doi.org/10.18174/455759>. The main text states: "AgroDataCube provides a large collection of both open data and derived data for use in agri-food applications. It is a work in progress, feel free to join our community on GitHub and the Gitter channel. The API documentation is available here in Postman." Below this, there is a section titled "For example:" with a sample API request: `https://agrodatabcube.wur.nl/api/v1/rest/fields?geometry=POLYGON((4.2 52.0, 4.2 53.0, 4.3 53.0, 4.3 52.0, 4.2 52.0))&epsg=4326&year=2017&cropname=maiskoutput_epsg=4326`. A paragraph explains that this request will return all crop fields where maize was grown in 2017, within the specified area, and provides details about the WKT POLYGON geometry, coordinates, and coordinate systems (epsg 4326 for input, epsg 28992 for output).



This block contains three overlapping screenshots. The top screenshot is a Gitter chat window titled "AgroDataCube/Community - Gitter" showing a discussion about the project's future and data availability. The middle screenshot is a GitHub repository page for "AgroDataCube" with a description: "A Big Open Data collection for Agri-Food Applications" and navigation options for repositories, people, teams, projects, and settings. The bottom screenshot is a Postman API documentation page for "AgroDataCube" showing a "GET Retrieve the geometry and soil information of the intersections of the suppl..." endpoint. It includes a sample request, headers (Accept: application/json, charset=utf-8, token: [token]), and parameters (output_epsg: 28992, page_size: 25). A sample response is also shown, including metadata, regions, soil types, and sources.

AgInfra+

D4Science e-Infrastructure - VRE



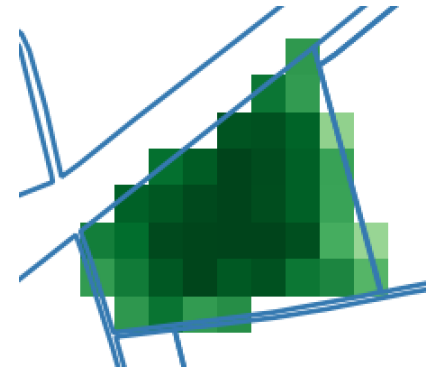
AgInfra+ - FAIR Data Science

The image displays four overlapping screenshots of the AgInfra+ FAIR Data Science ecosystem:

- Top-Left:** A screenshot of the **DataMiner** interface. It shows a list of simulation runs under the heading "Crop Model (3)". The runs include "Wofost" (Sample single run of the WOFOST model), "Wofost 20180424 1" (Wofost Crop Simulation Model), and "Wofost 20180430 1" (Wofost Crop Simulation Model). A "Data Access (1)" section shows the "Agrodatacube Reader 20180430 1" for retrieving AgroDataCube data through its REST API.
- Top-Right:** A screenshot of a **Jupyter EGI** environment. The browser address bar shows "aginfra.d4science.org/group/agroclimaticmodelling/jupyter-egi". The interface includes a navigation menu with "Agro-Climatic Modelling", "Members", "Analytics", "Visualization", "Administration", and "Activity Tracker". The main area shows a Jupyter notebook with a Python script for Wofost simulation, including comments and code for defining process IDs and inputs.
- Bottom-Left:** A screenshot of the **AGINFRA+ Catalogue** interface. The browser address bar shows "aginfra.d4science.org/group/aginfraplus/catalogue". The interface features a navigation menu with "AGINFRA+", "Members", "Project mgmt. tools", "Analytics", "Semantics services", "Catalogue", "Data discovery", and "Calendar". The main area shows a search for "crop simulation" with "3 items found". The search results are filtered by location and include tags like "crop simulation".
- Bottom-Right:** A screenshot showing a detailed view of a catalogue item titled "Wofost Inputs". The item is owned by "OWNER" and is a "binary" type. The description states: "This folder contains sample input files for use with the WOFOST DataMiner algorithm. Each .ser file is a binary archive with all the required input parameters (crop...". Other items listed include "Wofost crop simulation model - single predefined run" (Method) and "Crop Parameters" (ResearchObject).

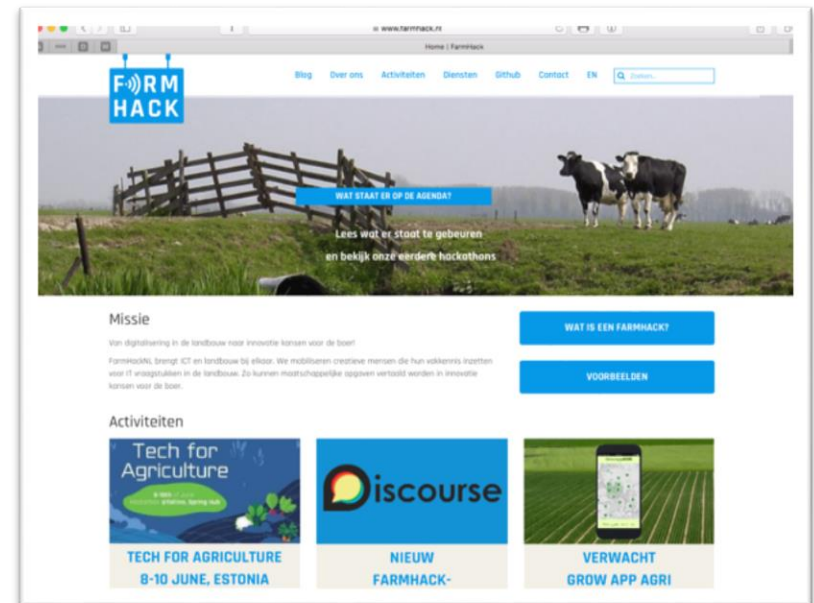
Findings

- Changing old habits and traditions is time consuming
- People need time and interest to learn (lots of) new technologies
- Big Data technology is being researched, it takes some perseverance to use it
- Hard to get 'easy' results, have to invest in the technology first



(Our) main approach

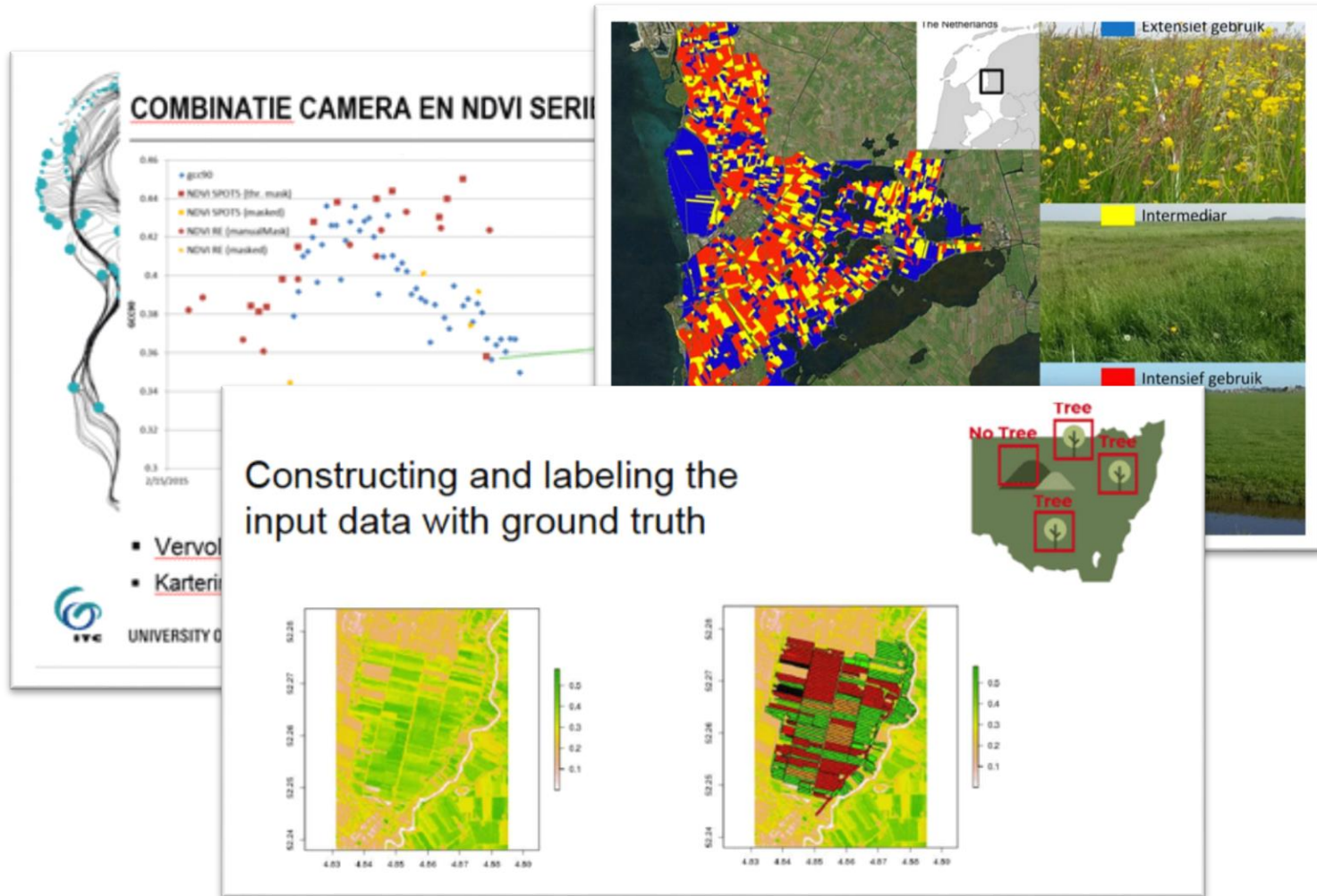
- “If you build it, they will come”. However:
- Use iterations and get feedback fast and often
 - e.g. hackathons
- And: focus on cloud services and container platforms



Some progress

- Increasing interest and awareness
- “WUR is serious about data”
- Wageningen Data Competence Centre (WDCC)
- IT department working on Azure cloud services and Docker / Kubernetes platforms

AgroDataCube for Machine Learning



<https://www.farmhack.nl/resultaten-rewarding-nature-hack/>

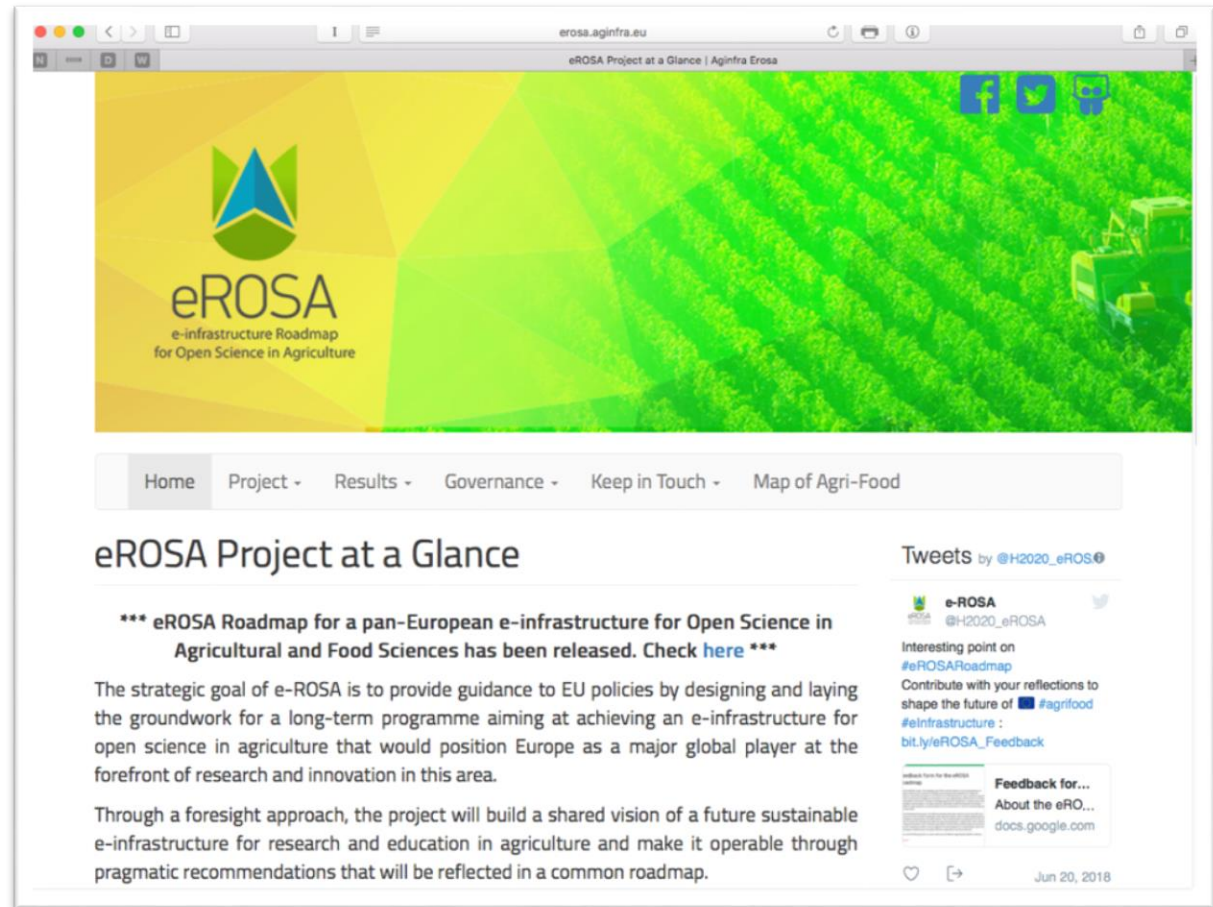
eROSA

■ Roadmap

- Past
- Present
- Future

■ Focus

- Open
- Sharing
- Collaboration



The screenshot shows a web browser window displaying the eROSA Project website. The browser's address bar shows the URL `erosa.aginfra.eu`. The page title is "eROSA Project at a Glance | Aginfra Erosa". The main content area features a large banner with the eROSA logo (a stylized blue and green triangle) and the text "eROSA e-infrastructure Roadmap for Open Science in Agriculture". Below the banner is a navigation menu with links: "Home", "Project", "Results", "Governance", "Keep in Touch", and "Map of Agri-Food". The main heading is "eROSA Project at a Glance". Below this, there is a tweet from @H2020_eROSA dated Jun 20, 2018. The tweet text reads: "*** eROSA Roadmap for a pan-European e-infrastructure for Open Science in Agricultural and Food Sciences has been released. Check here ***". The tweet also includes a link to the roadmap and a call to action: "Contribute with your reflections to shape the future of #agrifood #einfrastructure : bit.ly/eROSA_Feedback".

Thank You!

rob.knapen@wur.nl

