

Data Management

S5: Advanced data management (2)

Towards Data Science

Monday 11/12/2023 - 14:00-16:00 (CET)

Isabelle Alic, Farzaneh Kazemipour-Ricci - INRAE

Marie Weiss, Llorenç Cabrera Bosquet - INRAE

General objectives: Overview of data management for plant phenotyping - focus on FAIR data

Session 5

Advanced data management (2) - Towards Data Science

Overview

Towards Data Science

- Introduction & Review of last sessions
- Use case: from raw data to process analyses based on FAIR principles - 4P
- Data enrichment: Events, Annotation and Documents
- Use Case: Data analysis and integration
- Geospatial: Integration and a step towards decision support
- Conclusion & Perspectives

Introduction & Review of last sessions

Reminder - Data Management Best Practices

Step1 - Identification

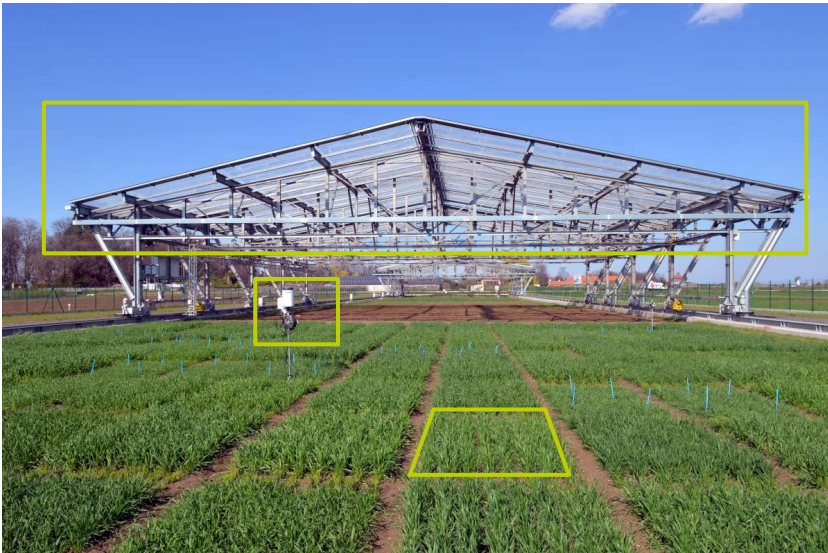
URI of Experiment: <emphasis:study56>

URI of Facility: <emphasis:gantry2>

URI of Device: <emphasis:st14>

URI of microplot: <emphasis:plot15879>

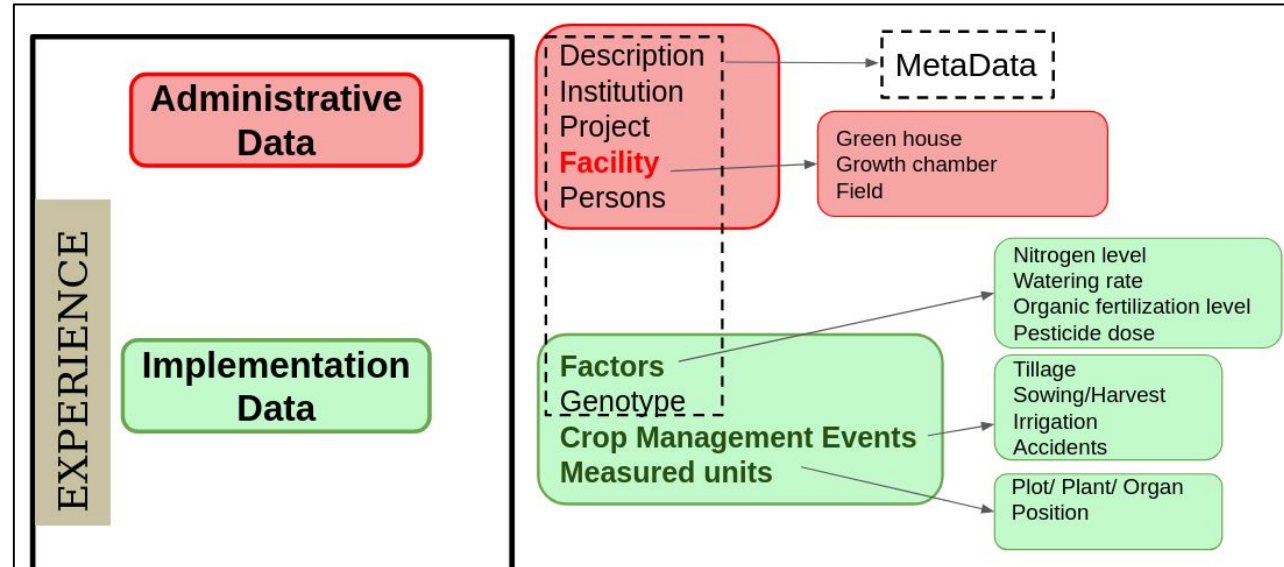
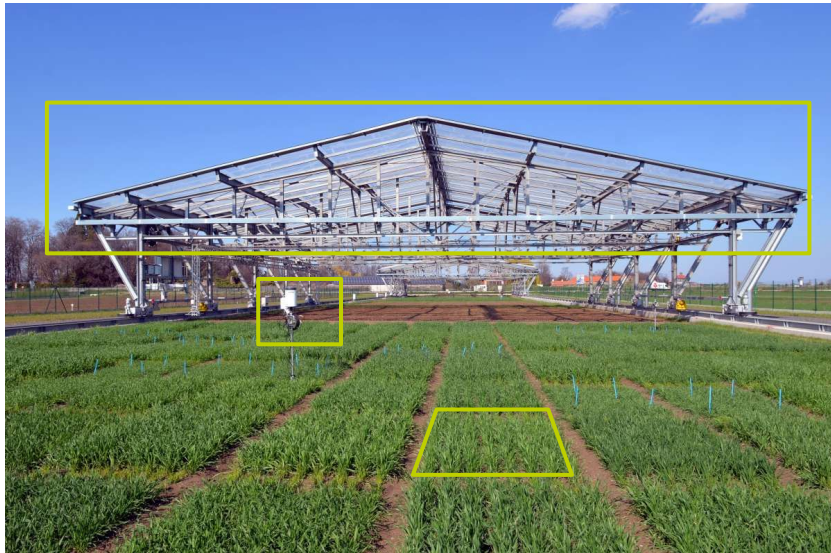
URI of germplasm:
<<https://www.geves.fr/catalogue/variete/1002491-aubusson>>



Introduction & Review of last sessions

Reminder - Data Management Best Practices

Step2 - Experiment Description



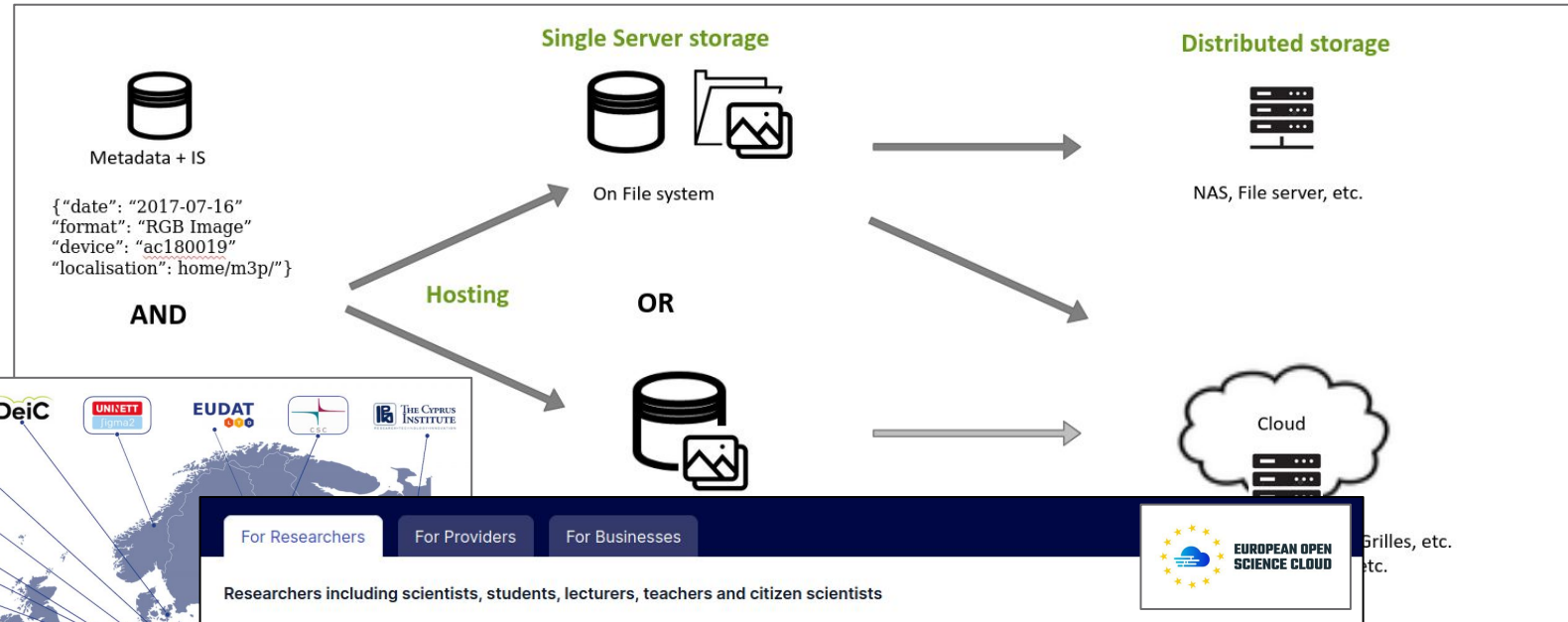
Introduction & Review of last sessions

Reminder - Data Management Best Practices



Step3 - Datafiles storage

Study the storage solution best suited to your needs and uses!



EUROPEAN OPEN SCIENCE CLOUD

For Researchers For Providers For Businesses

Researchers including scientists, students, lecturers, teachers and citizen scientists

Explore and Contribute

- Discover Research Outputs**
Find datasets, scientific publications and software for your research activities
- Publish Research Outputs**
Store, backup, archive your data, publications, software
- Find Funding Opportunities**
Learn about RDA/EOSC Future open calls

Tools

- Access Computing and Storage Resource**
Find HPC, IT centres for science, cloud computing, online storage
- Process and Analyse**
Verify, organise, transform and integrate data, then export it in the format you need
- Access Training Materials**
Find lessons, courses, videos

More

- [Research Data Management](#)
- [Research Infrastructures](#)
- [Instruments & Equipments](#)
- [Regional & Thematic Projects](#)

[Get Inspired](#)

Grilles, etc.
etc.

Introduction & Review of last sessions

Reminder - Data Management Best Practices

Step3 - Data Declaration

- Variable description
 - Using Model
 - With ID
- Provenance
 - Data Acquisition description
 - With ID



Trait

PLANT Entity + HEIGHT Characteristic + RULER Method + CM Scale (unit)

URI
http://phenome.inrae.fr/id/variable/canopyEar_Number_counting_PERm2

Entity *
canopyEar x

Entity of interest
Search and select an observation level

Characteristic *
number per area x

Species
barley x and 2 more

Method *
counting x

Unit/Scale *
per square metre x

Trait already existing in an ontology

PROVENANCE

- Description
- Start / End Date
- Vector
- Sensor
- Operator
- ...

Introduction & Review of last sessions

Reminder - Data Management Best Practices

Step3 - Data Declaration

Environmental Data

Provenance

Devices: emphasis:st14
Station: campbell-2018-dl53

Data

Variable: air_temperature_minimumDaily_degC
Date: 2023-07-01
Target: plot2023-BH53
Value: 12

Phenotypic Data

Provenance

Vector: fieldRobot1
Camera: campbell-2019-dl4
Software: 4P
ProvUsed: Datafile568

Data

Variable: Canopy_CoverFraction0deg_ImageSeg_Unitless
Date: 2023-07-01
Target: <emphasis:plot15879>
Value: 0.1



Introduction & Review of last sessions

Reminder - Data Management Best Practices



=> Having well-organized and well-described data will enable **data exploration and analysis**

e_minimumDaily_degC

ImageSeg_Unitless



Use case: from raw data to process analyses based on FAIR principles

4P: Plant Phenotyping Processing Platform - Marie Weiss



LIDAR



Multispectral

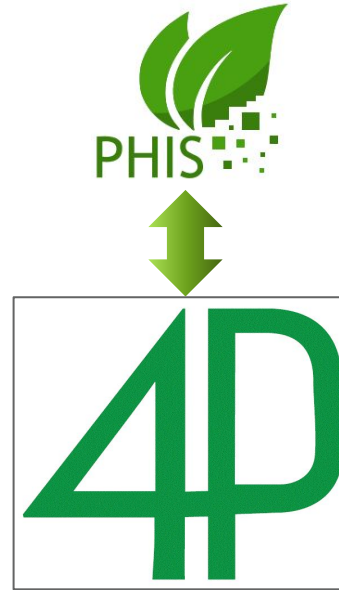


RGB

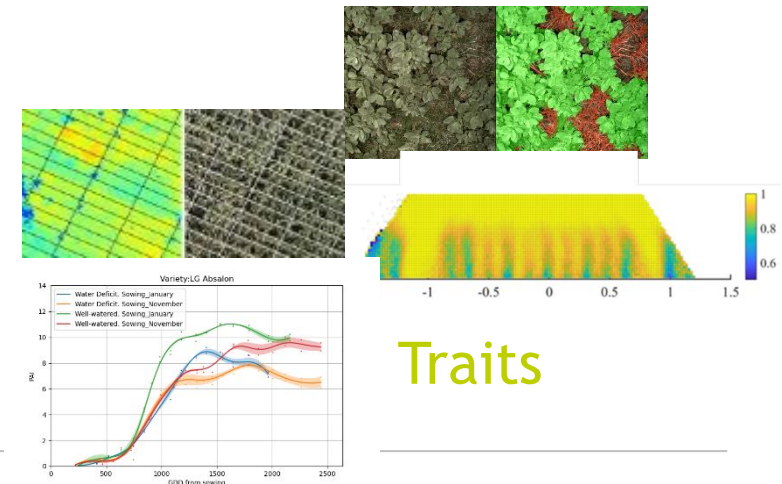
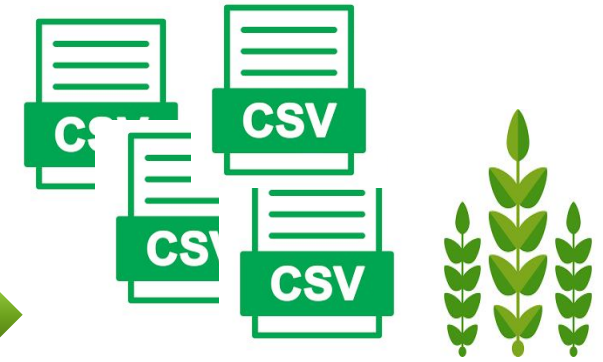


Stereovision

Field Acquisitions



Data Processing



Traits

Survey among phenotyping installations and users @ INRAE (2023)

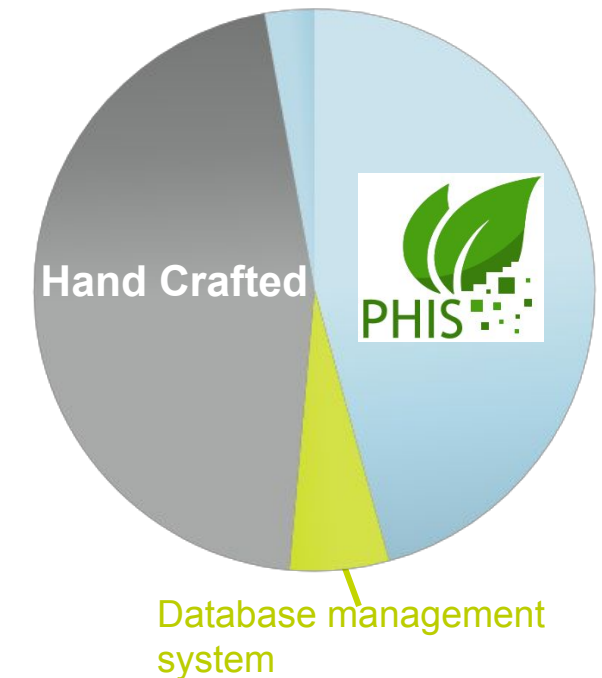
Prioritization of needs

1. Data traceability - FAIR : 28% Data only 71% Data+ Algorithms
2. Ergonomy / Easy to use
3. Quick Processing & computing facilities
4. Data Storage
5. Processing traceability

Capacities for processing

1. 57%: No in-house high-speed processing capacity
2. 21%: Processing facilities within the installation/lab
3. 16%: subcontracting specialized companies
4. 6% : Not yet defined

Current Data Management
500 Go to 10 To / campaign





Requirements

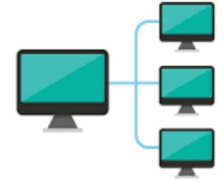
- Usable by all phenotyping installations @ INRAE
 - Compatible with all vectors/sensors
 - Processing pipelines easily shareable with PHENOME partners
 - Ergonomic: users are not developers
- Traceable
 - Storage of information related to sensor/vectors/data
 - Management and versioning of data processing chains
- Flexibility
 - Genericity : able to integrate processing chains developed in different languages, use commercial softwares (e.g. metashape)
 - Management and versioning of data processing chains
 - Scalability: update with new sensors/vectors/algo/traits



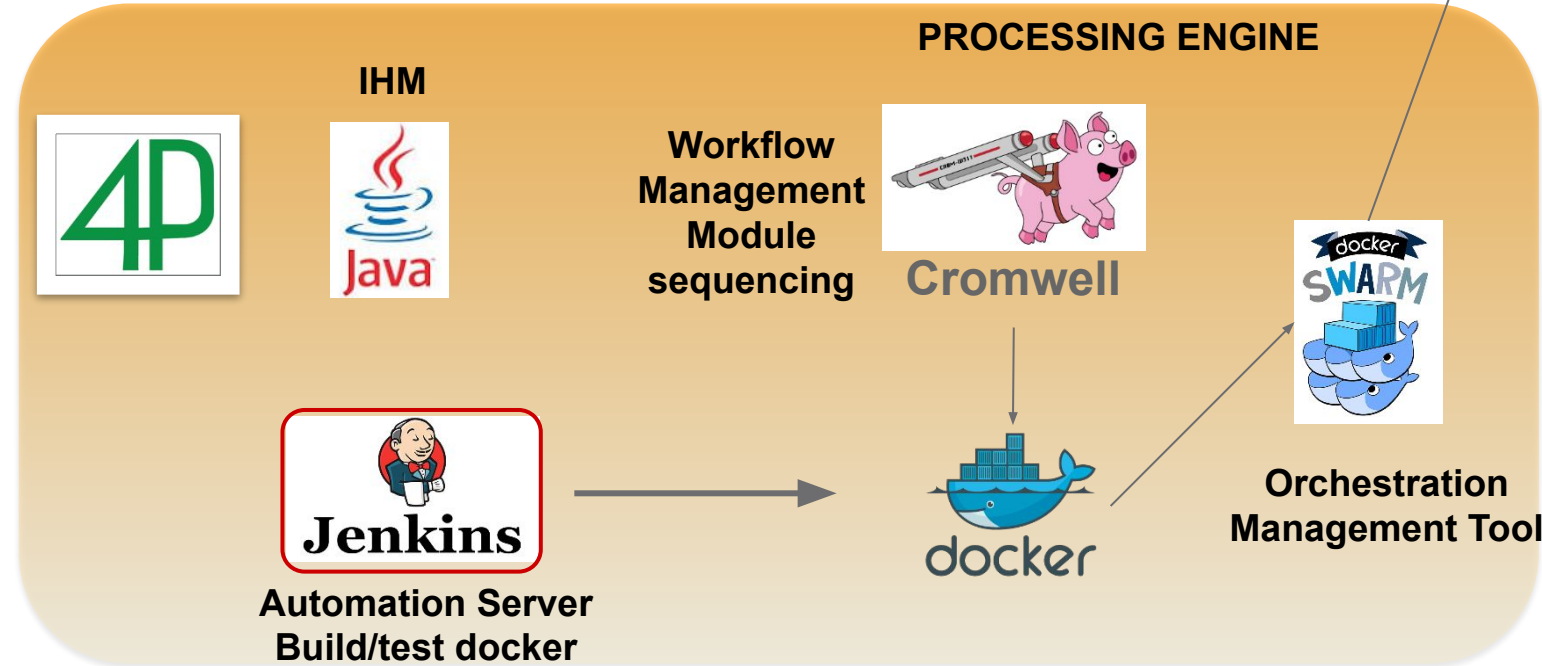
Proposed Solution



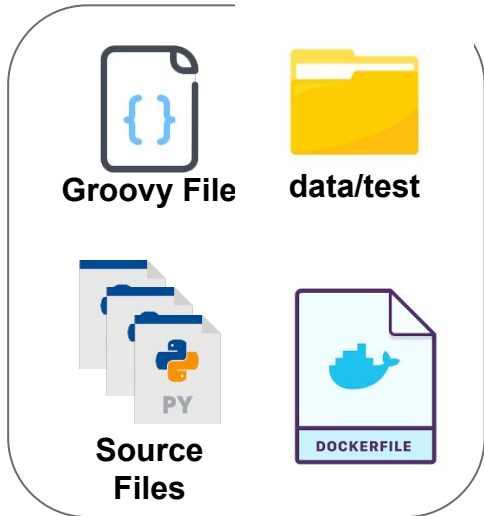
Virtual Machines



Import
Export



Modules (Source Code & Docker)



Dashboard

Data

Macros

Modules

Monitoring

Downloads

Data types

Catalog

Dashboard of Toulouse

Amount of data

11,4 TB

Upload raw data

Running processes

3

Run a process

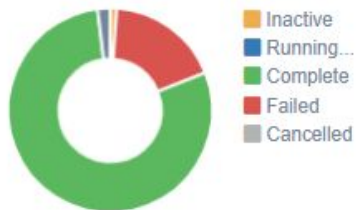
Disk space

35%

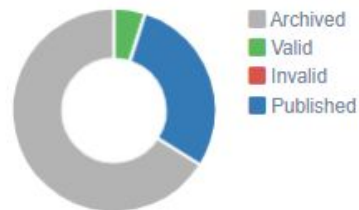
Running processes over the last 12 months



Processes status



Macros status



Datatypes



Home Page

Installation

Data Processing

status

Dashboard

Data

Macros

Modules

Monitoring

Downloads

Data types

Catalog

Data

Data Management

Load Data

Measuring session	Loading date	Experiment(s)	Vector	Data	PHIS	Size
DD/MM/YYYY - D	DD/MM/YYYY -	All	All	All	All	
30/03/2023 11:47	07/12/2023 12:08	23ZM8_FFAST	Phenomobile	HDF5 Files RGB Images Point Clouds		99.9 GB
17/11/2023 11:34	04/12/2023 13:04	24ZN8_FFAST	Phenomobile	HDF5 Files		50.2 GB
14/11/2023 12:34	27/11/2023 13:29	24ZN8_FFAST	Phenomobile	HDF5 Files RGB Images		78.9 GB
14/11/2023 12:34	22/11/2023 13:55	24ZN8_FFAST	Phenomobile	HDF5 Files		50.0 GB
26/07/2023 12:00	22/11/2023 09:46	23TE43_cimson	Drone	RGB Images		26.0 GB
16/06/2023 09:23	22/11/2023 08:23	23ZM15_GIP	Phenomobile	HDF5 Files VegBackg Seg RGB Images RGB Images Multispectral Images Cropped RGB Images GreenSen Seg RGB Images Point Clouds Gap Fraction Canopy GreenSen FVC Canopy CoverFraction Canopy Height		277.8 GB
20/07/2023 12:00	16/11/2023 14:03	23TE43_cimson	Drone	RGB Images		35.8 GB
14/11/2023 21:03	14/11/2023 20:05	24ZN8_FFAST	Phenomobile	HDF5 Files RGB Images Point Clouds		2.9 GB
12/07/2023 12:00	09/11/2023 14:38	23TE43_cimson	Drone	RGB Images		29.1 GB
03/07/2023 12:00	08/11/2023 14:50	23TE43_cimson	Drone	RGB Images		32.5 GB

Showing 1 to 10 of 368 entries

1 2 3 4 5 ... 37 Next

Dashboard

Data

Macros

Modules

Monitoring

Downloads

Data types

Catalog

Upload to PHIS Process dataset

Data Status

Description (ID 2780)

Experiment(s): 23ZM15_GIP
 Measuring session: 16/06/2023 09:23
 Loading date: 22/11/2023 08:23
 Size: 277.8 GB
 iRODS path:
 /FranceGrillesZone/home/4p/qualification/2780
 Vector: Phenomobile
 PHIS:

Data

Name	Version	State	PHIS	Size	Actions
HDF5 Files	v1	✓	☰	63.4 GB	☰
RGB Images	v1	✓	☰	40.5 GB	ⓘ
Multispectral Images	v1	✓	☰	15.9 GB	☰
Point Clouds	v1	✓	☰	121.7 GB	☰
Cropped RGB Images	v1	✓	☰	9.1 GB	☰
VegBackg Seg RGB Images	v1	✓	☰	9.7 GB	☰
GreenSen Seg RGB Images	v1	✓	☰	16.5 GB	☰
Gap Fraction	v1	✓	☰	1.0 GB	☰
Canopy Height	v1	✓	☰	194 KB	☰
Canopy CoverFraction	v1	✓	☰	327 KB	☰
Canopy GreenSen FVC	v1	✓	☰	419 KB	☰

Task history

Label	Input data	Output data	Author	Start date	Duration	Status / Progress	Actions
Macro: <u>Pheno_GreenSenFvc - RGB SegVegAll v1.0</u>	RGB Images v1 Cropped RGB Images v1 VegBackg Seg RGB Images v1	Canopy GreenSen FVC v1 GreenSen Seg RGB Images v1	Rémy Marandel	01/12/2023 14:36	02 h 33 min 43 s	Complete 3/3	☰
Macro: <u>Pheno_Canopy_CoverFraction - RGB 1.2</u>	RGB Images v1	Cropped RGB Images v1 VegBackg Seg RGB Images v1 Canopy CoverFraction v1	Rémy Marandel	01/12/2023 08:38	05 h 41 min 46 s	Complete 4/4	☰

Dashboard

Data

Macros

Modules

Monitoring

Downloads

Data types

Catalog

Macros / Pheno Canopy CoverFraction - RGB 1.2

Pheno Canopy CoverFraction - RGB 1.2 (ID 366)

Duplicate

Process dataset

Delete

Details

Workflow

Name: Pheno Canopy CoverFraction - RGB 1.2

Description:

Segmentation of Phenomobile RGB images into Background and vegetation, and calculation of the Coverfraction.

Compatible vectors: Phenomobile

Last use: 02/12/2023 10:59

Author: Ayoub Nachite (ayoub.nachite@inrae.fr)

Technical guide:

Input data: RGB Images

Output data: Cropped RGB Images VegBackg Seg RGB Images Canopy CoverFraction

Status: ✓ Published

Version: ✓ Modules last versions are used

Visibility: 🔒 Restricted to "PHIS" users only

Manage versions

Macro used 13 times

Date	Installation	Output data
02/12/2023 10:59	Toulouse	23ZM15 GIP Phenomobile 28/04/2023 09:06
01/12/2023 08:39	Toulouse	23ZM15 GIP Phenomobile 09/05/2023 09:15
01/12/2023 08:38	Toulouse	23ZM15 GIP Phenomobile 16/06/2023 09:23
20/11/2023 13:26	Toulouse	23ZM15 GIP Phenomobile 22/12/2022 12:51

Data processed with that pipeline

Dashboard

Data

Macros

Modules

Monitoring

Downloads

Data types

Catalog

Macros / Pheno Canopy CoverFraction - RGB 1.2 / Workflow

Pheno Canopy CoverFraction - RGB 1.2 (ID 366)

Duplicate Process dataset

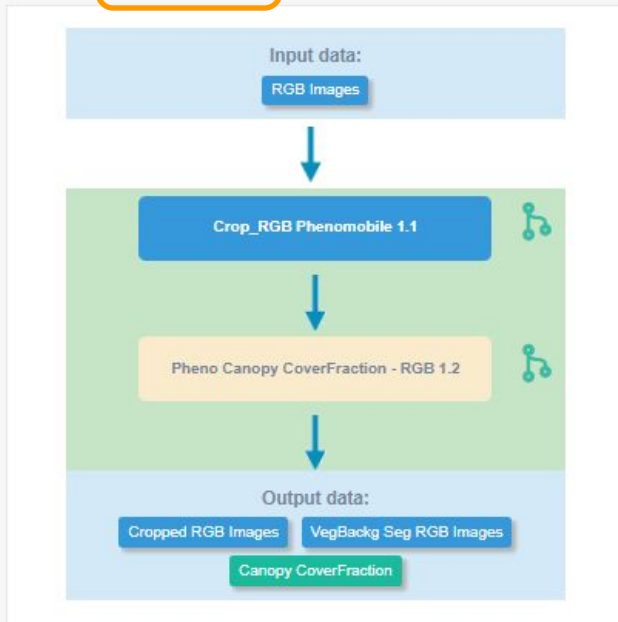
Display Processing chain

Delete

And each Module input/output

Details

Workflow



Module: [Crop_RGB Phenomobile 1.1](#)

Status: ✔ Published

Description:

module used to crop RGB Phenomobile images from the center to the cropAngle in degrees, before applying the Canopy Cover Fraction Segmentation module.

Input data: [RGB Images](#)

Inputs	Type	Source	Value
HDF5 Extracted Metadata ?	file	RGB Images -> HDF5 Extracted Metadata ▼	
HDF5 Extracted Positions ?	file	RGB Images -> HDF5 Extracted Positions ▼	
Height Crop angle ?	float	Fixed in the macro ▼	20
Rowsp ?	float	To be entered at the start of process ▼	
White Balanced RGB Images ?	list<file>	RGB Images -> White Balanced RGB Images ▼	
Width Crop angle ?	float	Fixed in the macro ▼	20

Export PHIS L2

Module Ex: Export 4P data to PHIS

Modify Duplicate

Delete

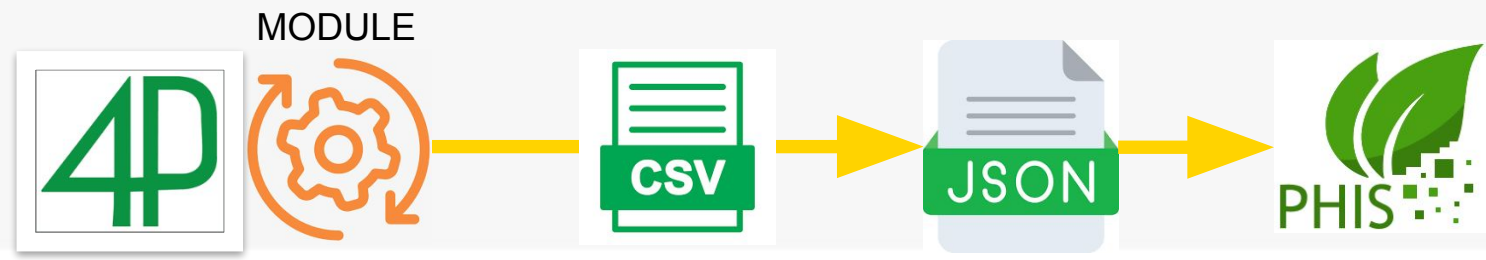
Description (ID 5093)

Module d'export des données L2 vers PHIS

Compatible species: All
 Type: Export module
 Author: Antony Tong (antony.tong@ephesia-consult.com)
 Compatible vectors:
 Technical guide: [Module - export L2 PHIS \(v1\)](#)
 Source repository : git@forgemia.inra.fr:4p/modules/phis-export-l2.git
 Docker image: phis-export-l2:1.0.0
 Status: ✓ Published from 03/01/2023 10:37
 Integration status: ✓ Integrated from 03/01/2023 10:39

Inputs Legend: Mandatory input Optional input

Prefix	Label	Type	Default value
-installation	Current installation	string	
-experiment	Experiment URIs	list<string>	
-json	Configuration File	file	
-provenanceL0	Provenance URI	string	
-sensorUri	Sensor URI	string	
-macro	Macro ID	integer	
-archiveUri	Archive URI	string	



All background tasks complete

Copyright © 2018 - 2023 INRAE All Rights Reserved - 4P - Version 2.8.1 of 30/11/2023

Catalog

[Load a document](#)

Name	Format	Size	Category	Vector	Organization	Status
<input type="text" value="Name"/>	<input type="text" value="All"/>		<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="Published"/>
Plant segmentation model configuration	JSON file	3 KB	Setting	Phenobile, Literal	Toulouse	✓ Published
Plant segmentation model weights	PTH file	40.7 MB	Deep Learning Model	Phenobile, Literal	Toulouse	✓ Published
Pheno eardensity model weights	PT file	175.1 MB	Deep Learning Model	Phenobile, Literal	Toulouse	✓ Published
GreenSen segmentation model		438 KB	Deep Learning Model	Phenobile, Literal	Toulouse	✓ Published
ksmodel	CSV file	14 KB	Deep Learning Model	Phenobile	Toulouse	✓ Published
Eardensity model weights	PT file	175.1 MB	Deep Learning Model	Phenobile	Toulouse	✓ Published
White balance factors file RGB	CSV file	113 B	Calibration	Phenobile	Toulouse	✓ Published
Soil height map	MAT file	1 KB	Measure	Drone	Toulouse	✓ Published
Transformation matrix 2 to 1	TXT file	408 B	Calibration	Phenobile	Toulouse	✓ Published
Transformation matrix 3 to 1	TXT file	410 B	Calibration	Phenobile	Toulouse	✓ Published

Showing 1 to 10 of 11 entries

1
2
Next



List of available traits

TRAIT	METHOD	SENSOR			VECTOR			REFERENCE
		RGB	Multispectral	LiDAR	UAV	Phenobile	LITERAL	
Plant height	Structure from motion/stereo		█		█			Madec et al., 2017 Weiss et al, 2017 Jay et al, en prep.
	Height Distribution			█		█		Madec et al., 2017
Vegetation Fraction (VF)	DL segmentation	█			█	█		Madec et al., 2022
	Height threshold			█		█		Lopez-Lozano et al., 2022
Green Fraction (GF)	VI Empirical		█		█			Jiang et al, 2018 Jay et al, 2019
	ML & DL segmentation	█			█	█		Serouart et al., 2022 Madec et al., 2022
	1D RTM inversion		█		█			Djamai et al, 2019 Camacho et al, 2021
Green Area Index (GAI)	VI Empirical		█		█			Jiang et al, 2018 Jay et al, 2019 Camacho, 2021
	1D RTM inversion		█		█			Djamai et al, 2019 Jay et al, 2019 Camacho et al, 2021
	3D RTM inversion	█		█		█		Liu et al., 2017 Jiang et al, 2019, 2020 Li et al, 2021 Lopez-Lozano et al, en prep
Plant Area Index (PAI)	1D Turbid	█		█		█		Lopez-Lozano et al, en prep

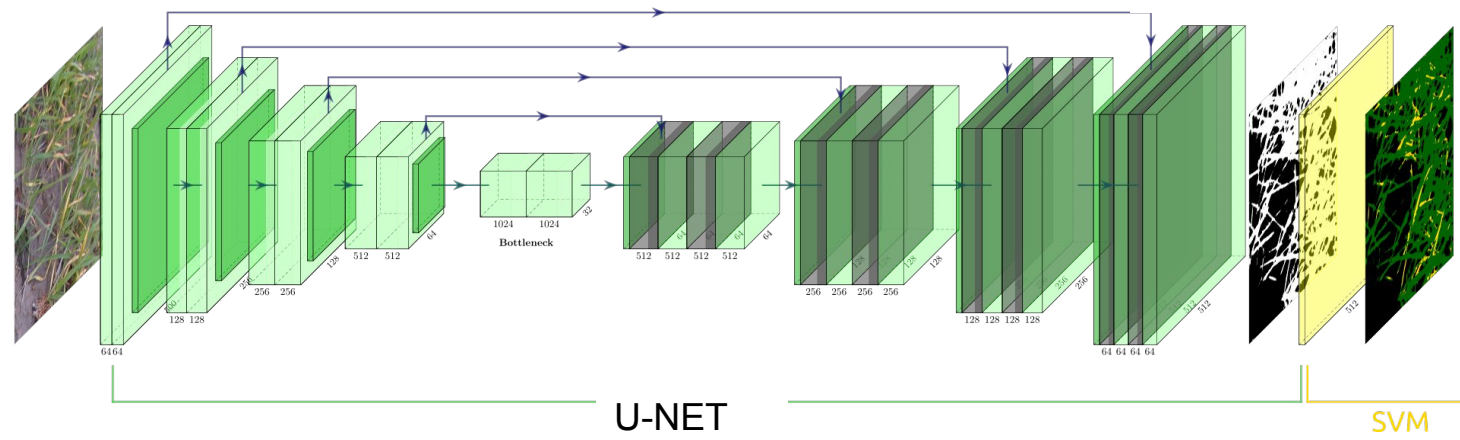
TRAIT	METHOD	SENSOR			VECTOR			REFERENCE
		RGB	Multispectral	LiDAR	UAV	Phenobile	LITERAL	
Fraction of Intercepted Radiation (FIPAR) & fAPAR	VI Empirical		█		█			Camacho et al, 2021
	1D RTM inversion		█		█			Jiang et al, 2017 Liu et al., 2019 Li et al, 2021 Camacho et al, 2021
	3D RTM inversion			█		█		Jiang et al., 2017
Average Inclination Angle (AIA)	1D RTM inversion		█		█			Liu et al., 2022 Lopez-Lozano et al., en prep Liu et al., 2019
	3D RTM inversion			█		█		Jiang et al., 2019
Canopy Chlorophyll Content (CCC)	1D RTM inversion		█		█			Delloye et al, 2018
	VI Empirical		█		█			Jay et al., 2019
3D Distribution of Leaf Area	1D Turbid			█		█		Liu et al., 2017 Lopez-Lozano et al, en prep
Plant density	DL	█			█	█		Jin et al., 2017 Velumani et al, 2021
Ear density	DL @ reprod, stage	█			█	█		Madec et al., 2019
Leaf Chlorophyll Content	1D RTM inversion		█		█			Jiang et al, 2018
	VI ML, Empirical	█	█		█			Jay et al., 2017, 2019 Jay et al, en prep



RGB Images/ Green Fraction

Serouart et al, 2022
Madec et al, 2023

1 – Vegetation/Background
Masks

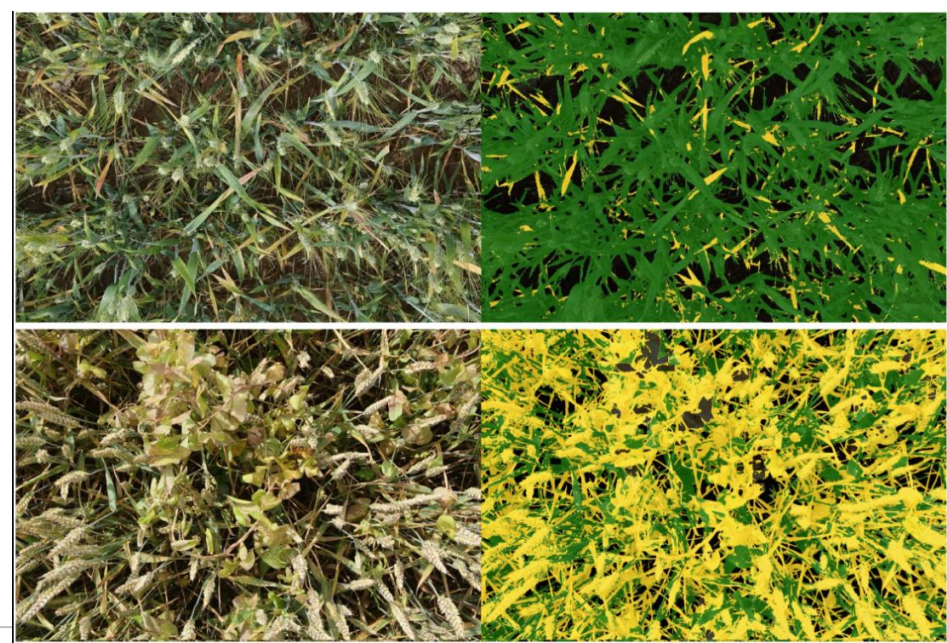


2- Green/Senescent vegetation

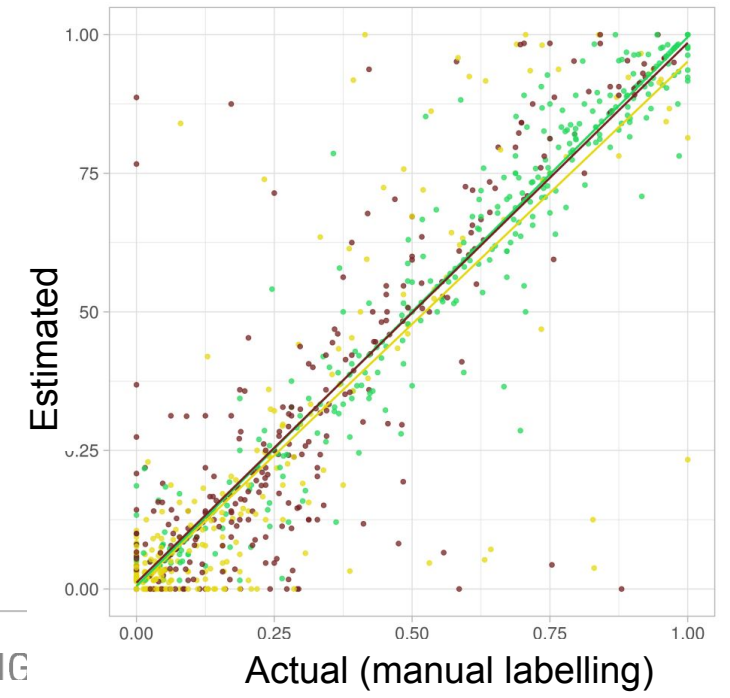
SVM

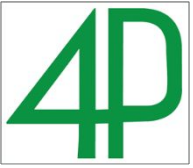
SVM

U-NET



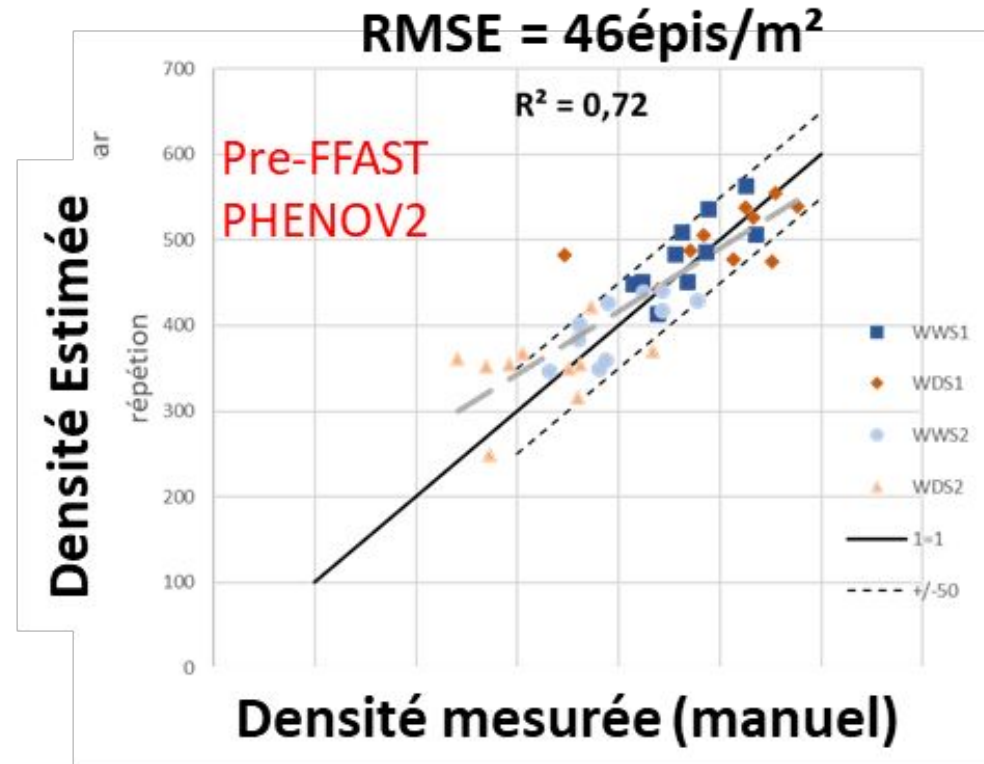
- Background
- Green Veg
- Senescent Veg





RGB Images+ LiDAR / Ear Density

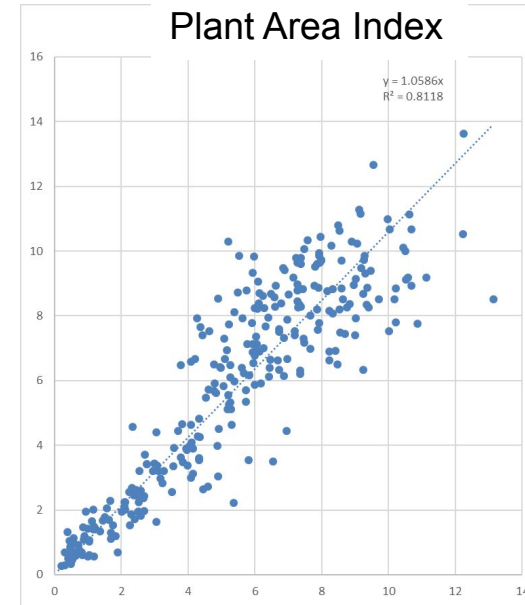
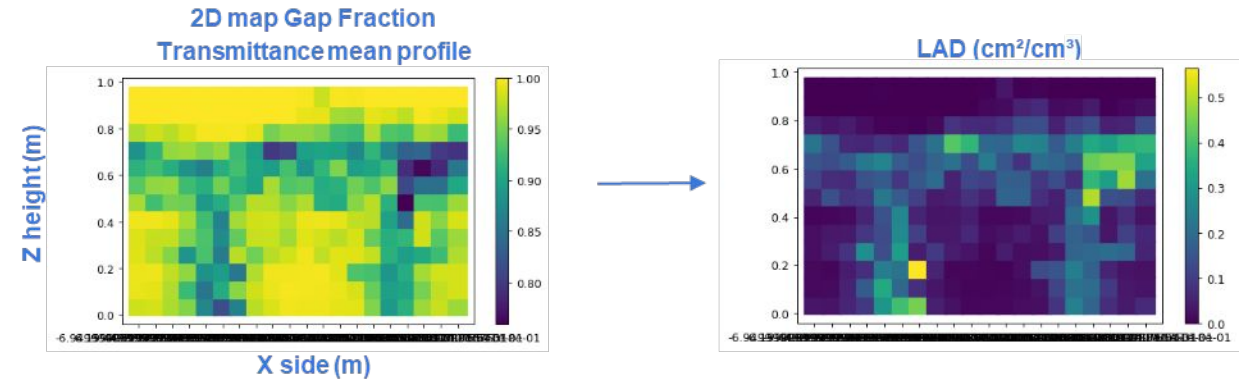
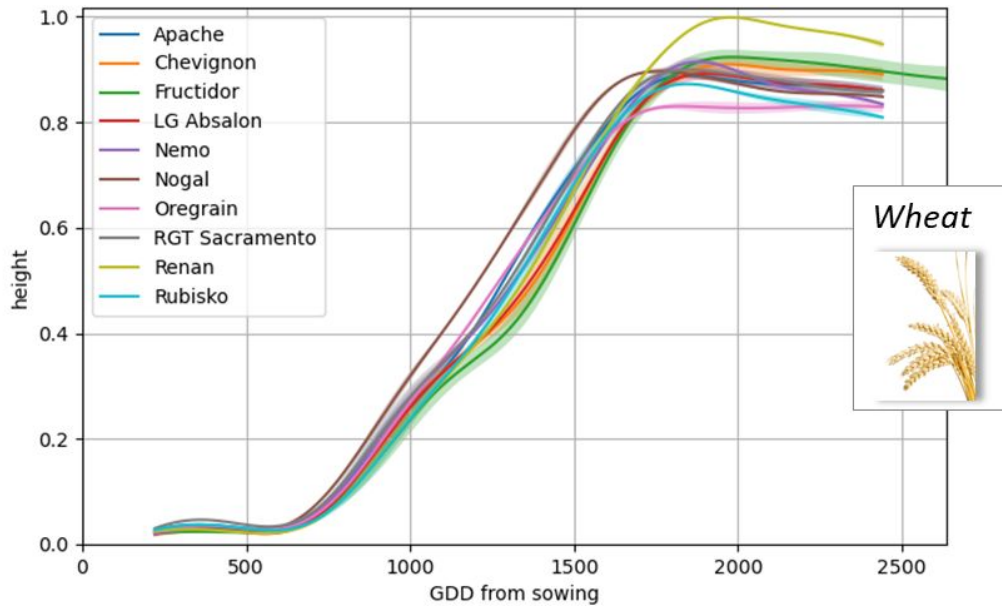
DL model from the global wheat head data challenge : David et al, 2021, 2023
Footprint of the RGB camera derived from camera height estimated from LidAR





LiDAR / Height / Gap Fraction / Plant Area Density

Lopez-Lozano et al, in prep



Pre-FFAST project (2021) @ DiaScope.
10 G x well-watered x november sowing x 2 replicates



THE TEAM



Platform
Development
Software
and functional
architecture



France Grille
Cloud computing and
storage



UMR EMMAH

Platform Management
Algorithms
Processing pipelines



UMR MISTEA

Information
System



Experimental

Units

UE GCA
UE PHACC
UE DIASCOPE

- Specifications
- Test
- Use



THANKS FOR YOUR ATTENTION

Data enrichment: Events, Annotation and Documents

Definitions

- **Events:** can be **processes, actions or facts** that occur in (or even precede) scientific experiments and have an influence on the experiment. Events are **identified and characterized**. Main categories are **controlled** (irrigation, fertilizer, maintenance, installation) or **uncontrolled** (hail, frost, pests, breakdown).

=> Treatment, Trouble, Pest attack, Move, etc.
- **Annotation:** A **short comment** added to an entity (data, object, experiment, device, etc.).

=> Motivations: Comment, Highlight, Classify, Moderate, Link, Tag, etc.
- **Documents:** A document refers to a **combination of a medium and information** (the content), the latter recorded in a persistent manner. It has explanatory, descriptive or evidential value.

=> Report, Publication, Technical document, Experimental protocol, Dataset, etc.

Data enrichment: Events, Annotation and Documents

Definitions

- **Events:** can be **processes, actions or facts** that occur in (or even precede) scientific experiments and have an influence on the experiment. Events are **identified and characterized**. Main categories are **controlled** (irrigation, fertilizer, maintenance, installation) or **uncontrolled** (hail, frost, pests, breakdown).

=> Treatment, Trouble, Pest attack, Move, etc.

=> MIAPPE, Ontology of Experimental Events (OEEV) - <https://agroportal.lirmm.fr/ontologies/OEEV>



- **Annotation:** A **short comment** added to an entity (data, object, experiment, device, etc.).

=> Motivations: Comment, Highlight, Classify, Moderate, Link, Tag, etc.

=> Web Annotation Ontology (OA) - <https://www.w3.org/ns/oa>



- **Documents:** A document refers to a **combination of a medium and information** (the content), the latter recorded in a persistent manner. It has explanatory, descriptive or evidential value.

=> Report, Publication, Technical document, Experimental protocol, Dataset, etc.

=> Dublin Core - <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>



Data enrichment: Events, Annotation and Documents

Definitions

- **Events:** can be processes, actions or facts that occur in (or even precede) scientific experiments and have an influence on the experiment. Events are identified and characterized. Main categories (e.g. controlled (treatment, fertilizer, maintenance, installation) or uncontrolled (hail, frost, pests, breakdown).

=> Treatment, Trouble, Pest attack, Move, etc.

=> MIAPPE, Ontology of Experimental Events

[http://www.miappe.org/ontologies/OEEV](#)



- **Annotation:** A short comment or note (e.g. instrument, device, etc.).

=> Motivation

Link, Tag, etc.

=> W3C

www.w3.org/ns/oa



- **Documents:** A combination of a medium and information (the content), the latter recorded in a persistent manner. It has scientific, descriptive or evidential value.

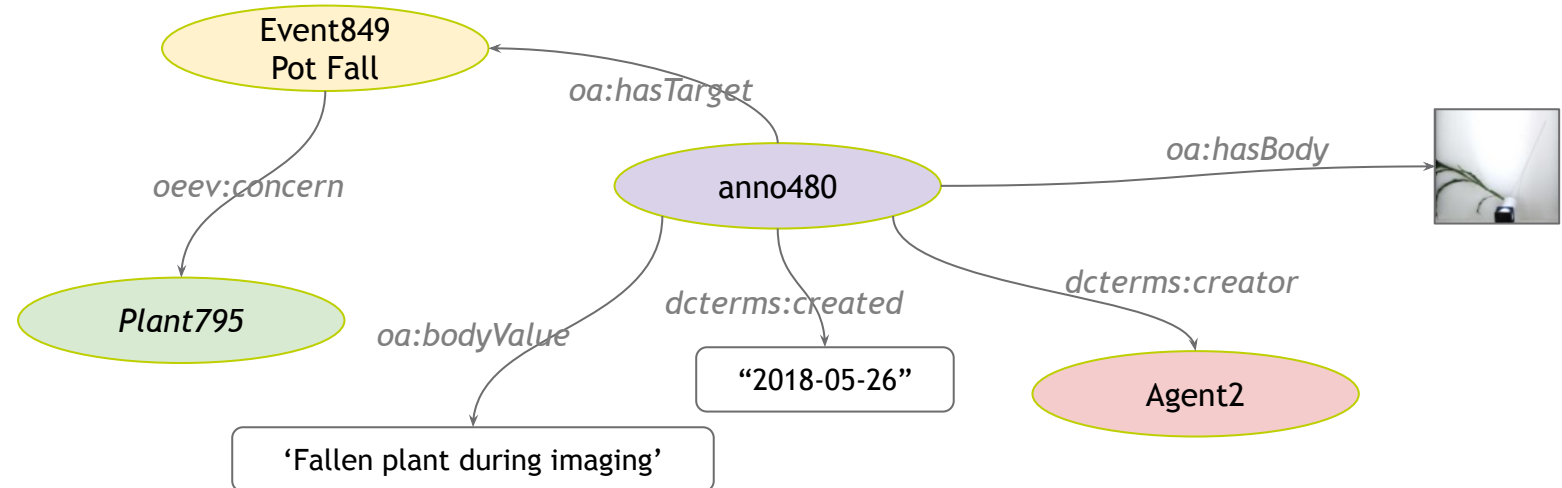
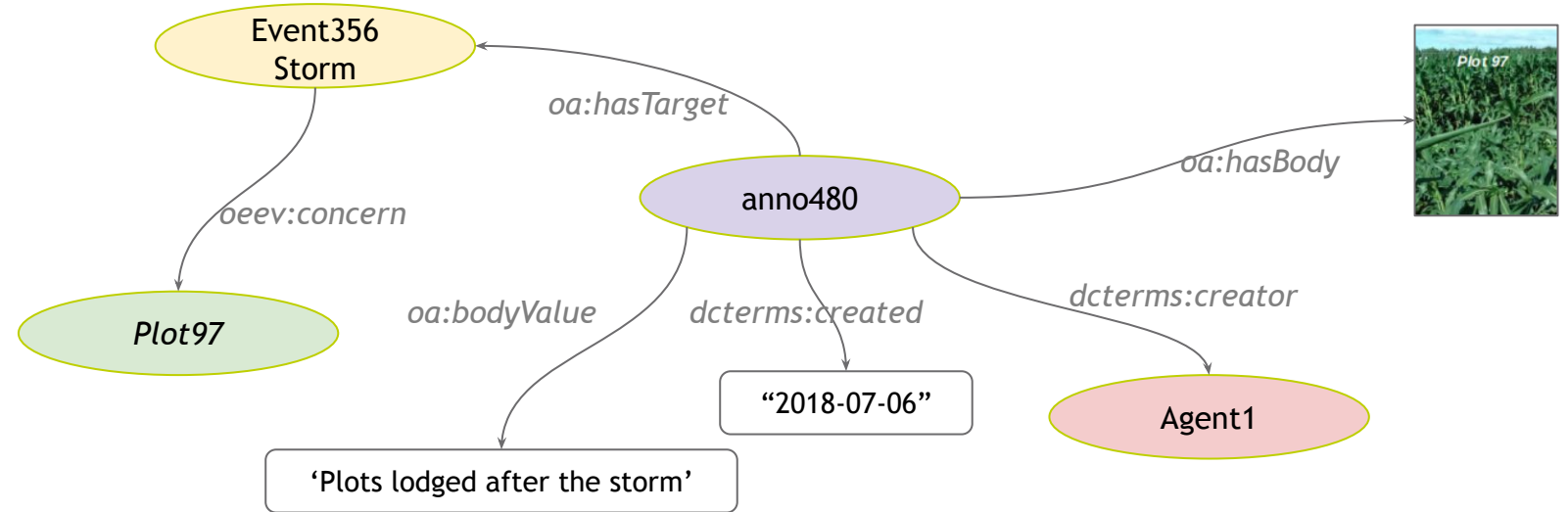
=> Report, Publication, Technical document, Experimental protocol, Dataset, etc.

=> Dublin Core - <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>



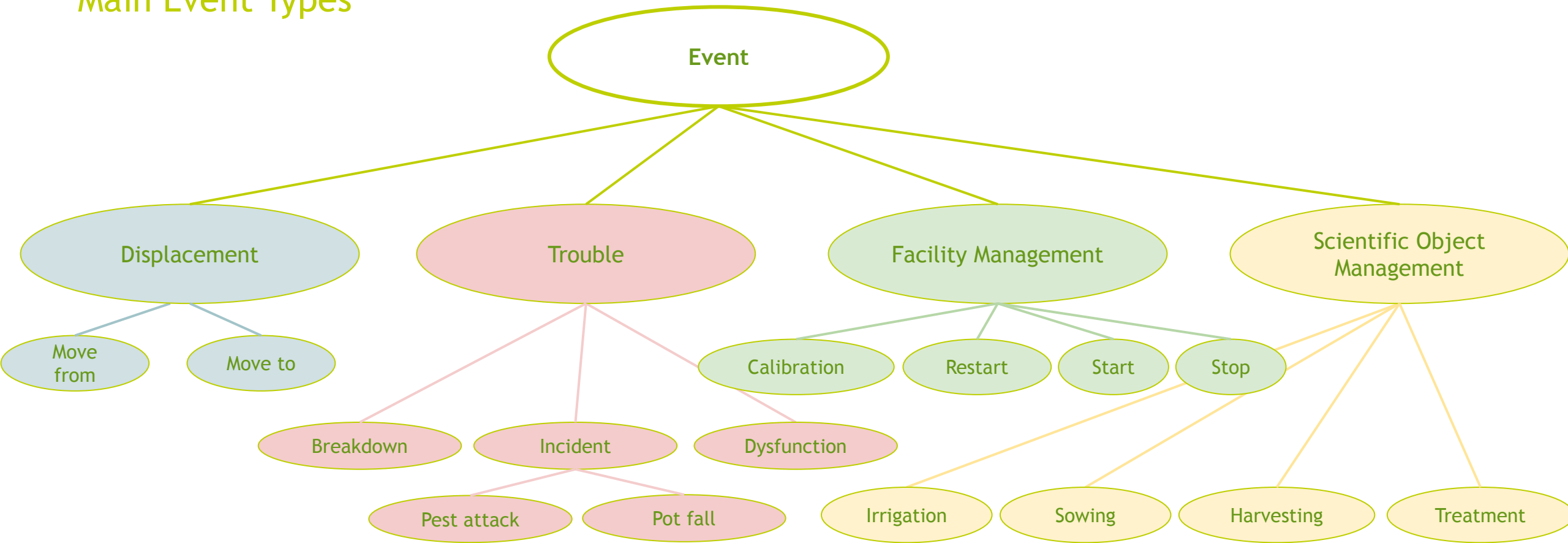
Data enrichment: Events, Annotation and Documents

Illustration



Events

Main Event Types



Events - MIAPPE important Section

Event

- Discrete, dated or time-stamped occurrence
- Natural (e.g. rain, pathogen attack)
- Cultural practice (e.g. sowing, irrigation)
- Applied to the whole study or by observation unit
- **This is not the factor**, but additional information.
- Event can be used to achieve a factor.

Event
Event type
Event accession number
Event description
Event date

Metadata : name, description and time/date

Example: POPLAR [2]

- the field establishment date, 2003
- the orchard was subjected to 15mm of rain on March 15, 2012 (fiction)



Study	Event		
	Name	Description	Date
Monclus <i>et al.</i> , 2012	Rain	15mm of rain on the orchard	2012-03-15

Events

Event Types & Properties

- Event ID
- Event Type
- Event Description
- Event Date / Period
- Target

Events

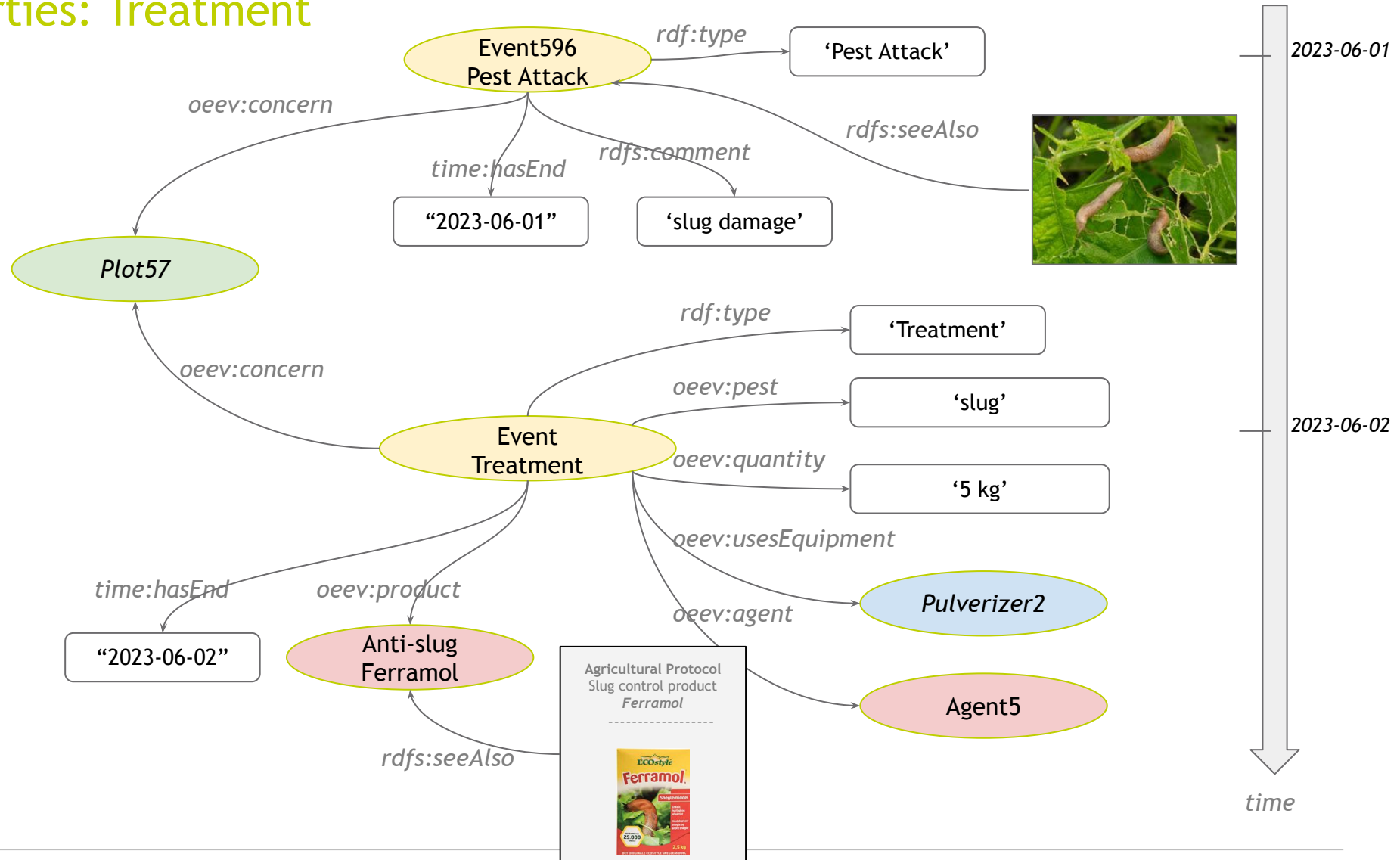
Event Types & Properties: Move

- Event ID
- Event Type
- Event Description
- Event Date / Period
- Target
- From
- To
- Coordinates
- X, Y, Z Position
- Textual position

Events

Event Types & Properties: Treatment

- Event ID
- Event Type
- Event Description
- Event Date / Period
- Target
- Product
- Quantity
- Pest
- Equipment used
- Agent



Events

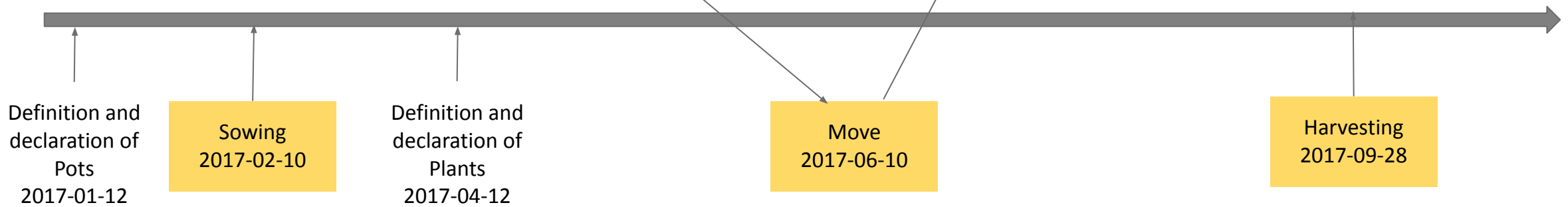
Describe a scientific object life cycle



Pot (*MAC5*)

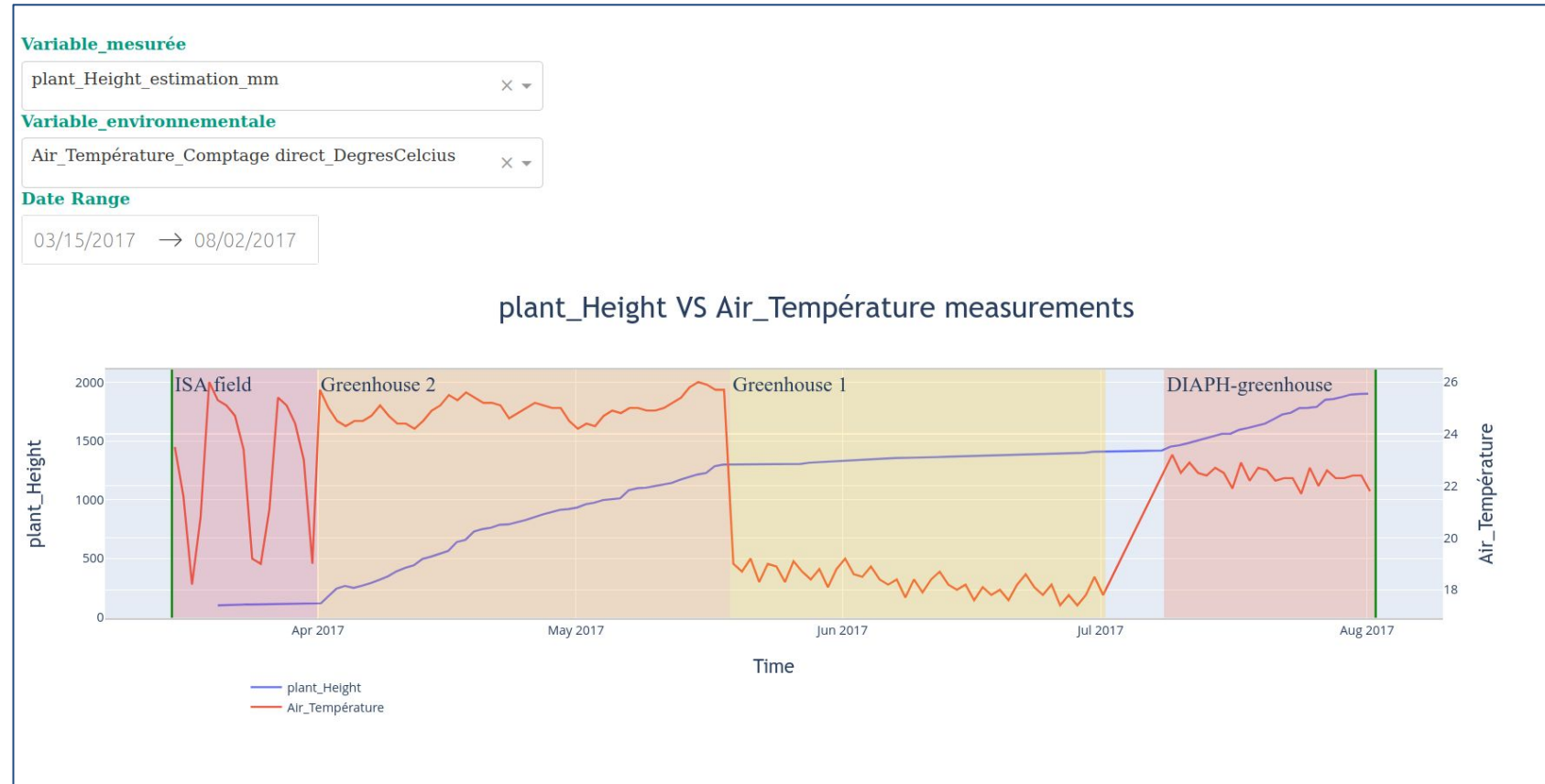


Greenhouse (*GH1*)



Data Exploration: environmental and phenotypic data

Visualising and linking environmental and phenotypic data



Definition and declaration of Plots
2017-01-12

Sowing
2017-02-10

Definition and declaration of Plants
2017-04-12

Move
2017-06-10

Harvesting
2017-09-28

Data Integration Architecture

Unified View Model



Query as if you had a single source !

Unified Schema

SPARQL EndPoint

RDF connector

Data virtual
integration

NoSQL connector



Use Case: Data analysis and integration

How FAIR data may help in managing and analyzing data in HTPP experiments? - Llorenç Cabrera Bosquet



Data analysis and integration

How FAIR data may help in managing and analyzing data in HTPP experiments?

First ->Some prerequisites

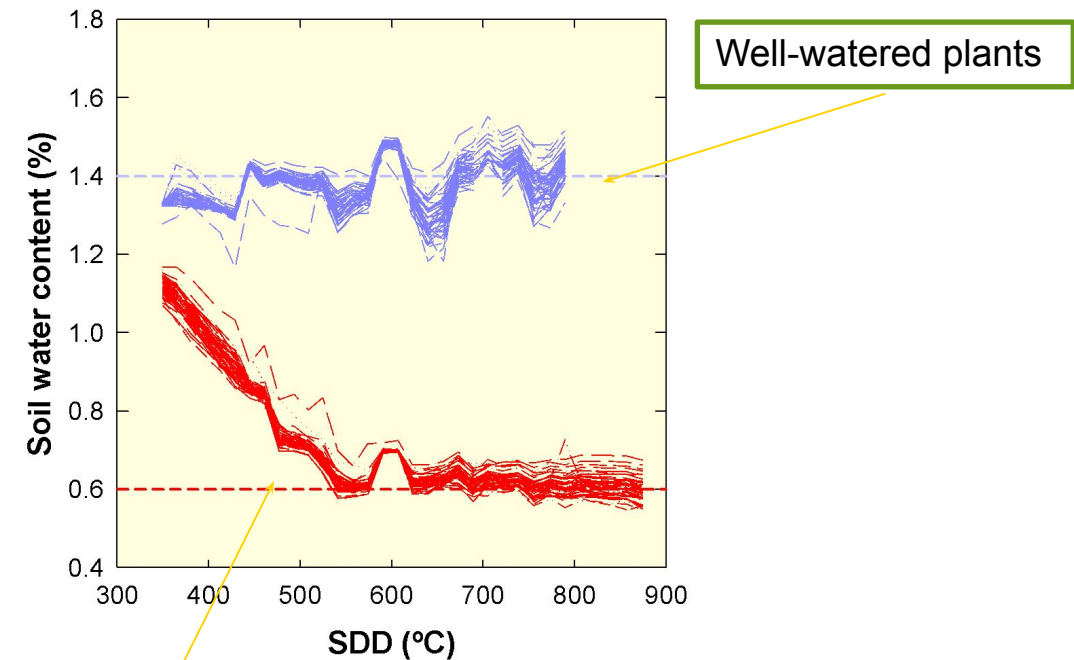
- Data are organised and managed following FAIR principles
- All data includes phenotypic, environmental and metadata measurements
- Ideally, data is managed in ‘real time‘ (allows decision-taking)

Exemple 1: Piloting and monitoring water deficit in thousands of plants

We need to continuously and individually monitor and manage water scenarios in most experiments



PHENOARCH platform Montpellier, INRAE



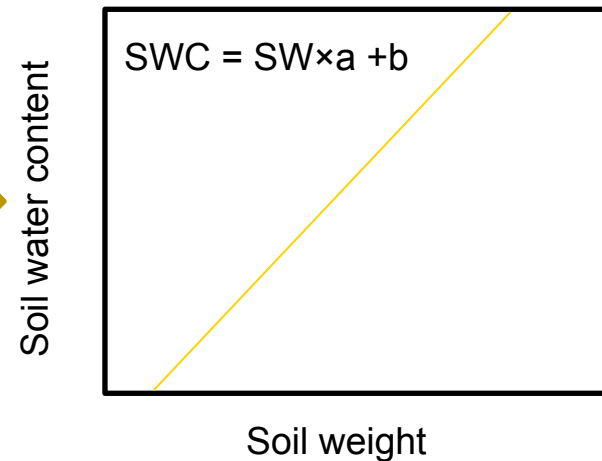
Imposition of water deficit (ca. -4 bar)

How soil water content (SWC) and SWP is measured?

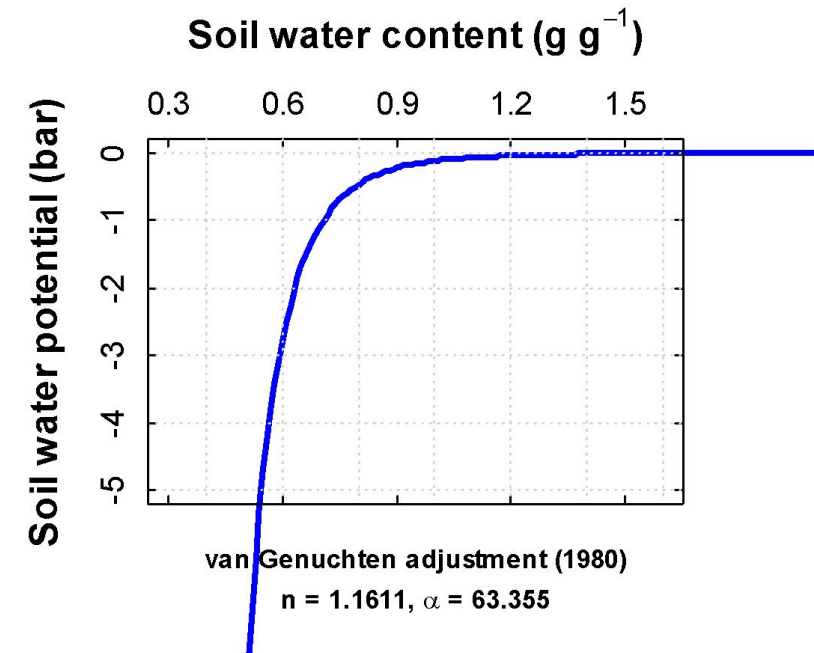
1. Plant are weighed individually on a daily basis (or more)



2. Soil water content (g H₂O per g of dry soil) is deduced using a formula



3. Soil water potential (bar) is calculated using a soil retention curve (Ψ vs Φ)



This is not so easy...

1. Platform weighs the whole 'system', not only soil weight



This is not so easy...

1. Platform weighs the whole 'system', not only soil weight

WE NEED



Proper identification of individual components

URI of plant

`<m3p:arch/2017/c17000118>`

URI of pot:

`<m3p:arch/2013/pc13001542>`

URI of cart:

`<m3p:arch/2013/ct1300123>`

This is not so easy...

1. Platform weighs the whole 'system', not only soil weight

WE NEED



Proper identification of individual components

URI of plant

`<m3p:arch/2017/c17000118>`

URI of pot:

`<m3p:arch/2013/pc13001542>`

URI of cart:

`<m3p:arch/2013/ct1300123>`

Weight of individual components + metadata

- Weight of pots, carts, soil, etc.
- Soil used and its characteristics
- Initial soil water content at potting
- Events that may have affect the weight (e.g. fertilisation, adding of sticks, etc.)

This is not so easy...

2. Plants grow over time



This may have a HUGE impact on soil water content estimation if plant biomass is not taken into account

This is not so easy...

2. Plants grow over time

WE NEED



Identification of relevant information

URI of plant

`<m3p:arch/2017/c17000118>`

URI of cabin:

`<m3p:arch/2018/ac180015>`

URI of camera:

`<m3p:arch/2018/ac180019>`

URI of image:

`<m3p:arch/2017/ic17002295855>`

This may have a HUGE impact on soil water content estimation if not taken into account

This is not so easy...

2. Plants grow over time

WE NEED



Identification of relevant information

URI of plant

`<m3p:arch/2017/c17000118>`

URI of cabin:

`<m3p:arch/2018/ac180015>`

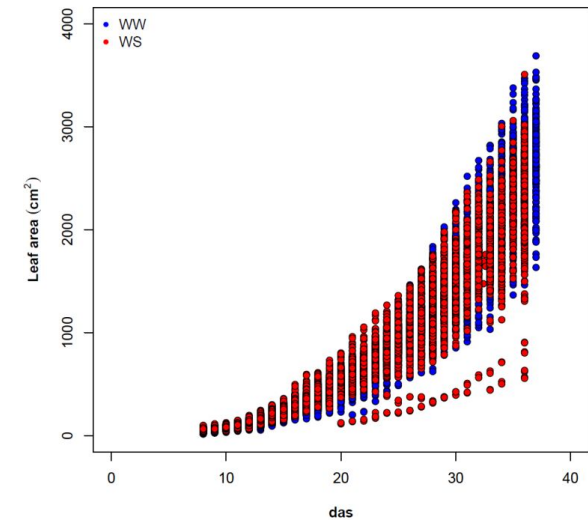
URI of camera:

`<m3p:arch/2018/ac180019>`

URI of image:

`<m3p:arch/2017/ic17002295855>`

Plant biomass estimation over time



Requires real time monitoring and analysis of data

This is not so easy...

2. Plants grow over time



WE NEED

Identification of relevant information

URI of plant

`<m3p:arch/2017/c17000118>`

URI of cabin:

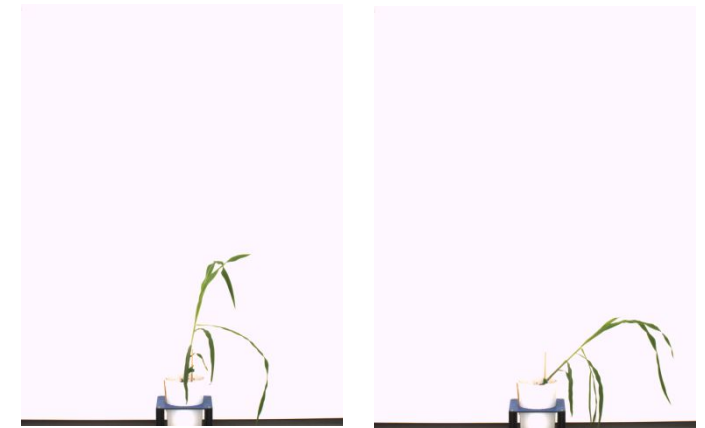
`<m3p:arch/2018/ac180015>`

URI of camera:

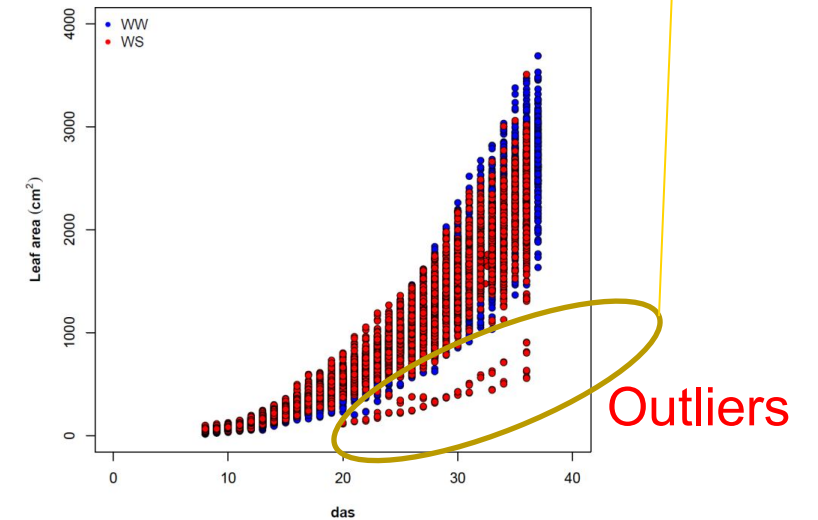
`<m3p:arch/2018/ac180019>`

URI of image:

`<m3p:arch/2017/ic17002295855>`

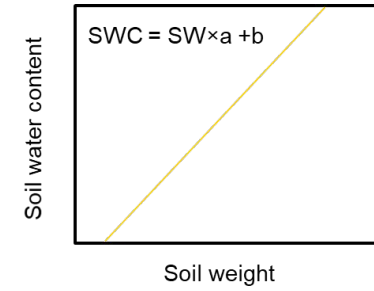


Plant biomass estimation over time



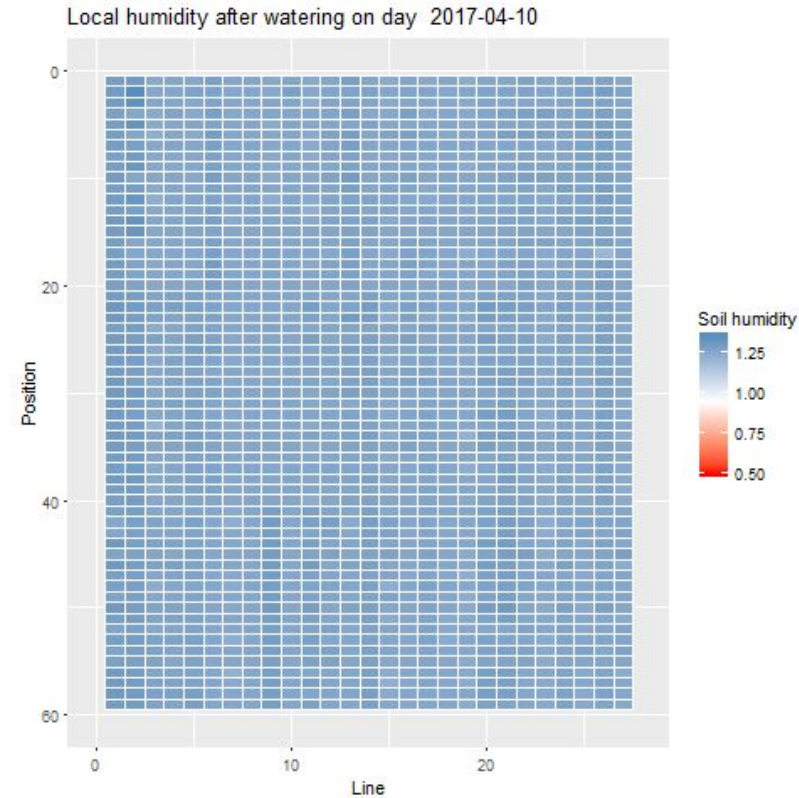
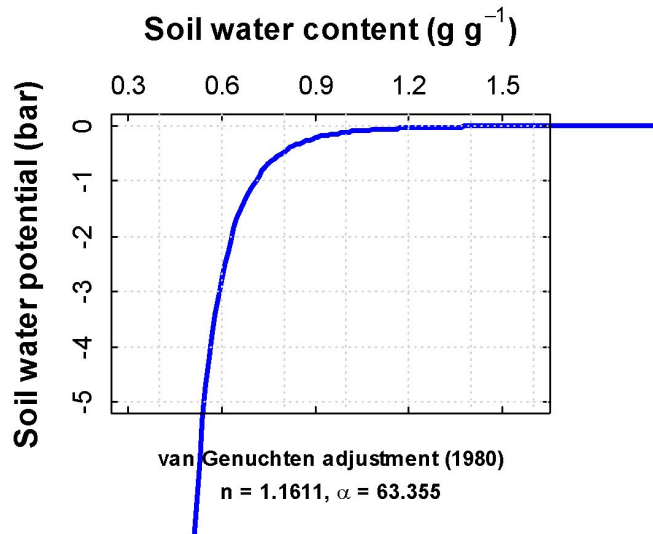
Requires real time monitoring and analysis of data

Finally SWC and SWP are calculated



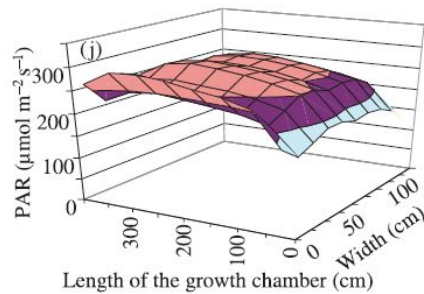
SWC = Weight of 'whole system' - Plant biomass - Weight of Pot - Weight of Cart - b / a

$$SWP = ((SWC/2)^{-n/(n-1)} - 1)^{1/n} / (-\alpha) / 10$$



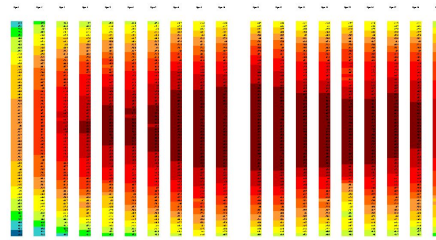
Envirotyping: Spatial and temporal variability and consequences for phenotyping

- Climate information is an essential **co-variate for monitoring growth** as drives most growth processes.
- Most platforms display a large temporal and spatial variability of environmental conditions

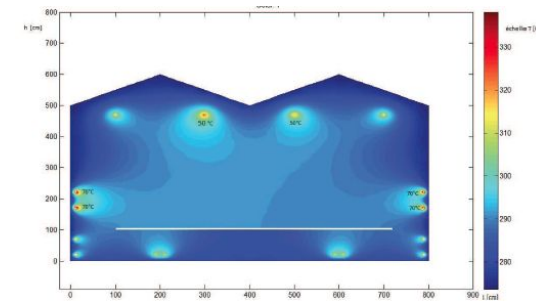


Granier et al 2006

PAR distribution
(lights on)



4PMI Dijon
Julien Martinet,
Christophe Salon



RootPhair,
UCL
Xavier Draye



There are some consequences for phenotyping...

IGNORING IT => depreciates the value of phenotypic data.

MEASURING IT enables

=> The combination of data from different experiments, different platforms or from field and platforms.

=> Taking advantage of (undesired) climate variability to determine environmental response curves during an experiment.

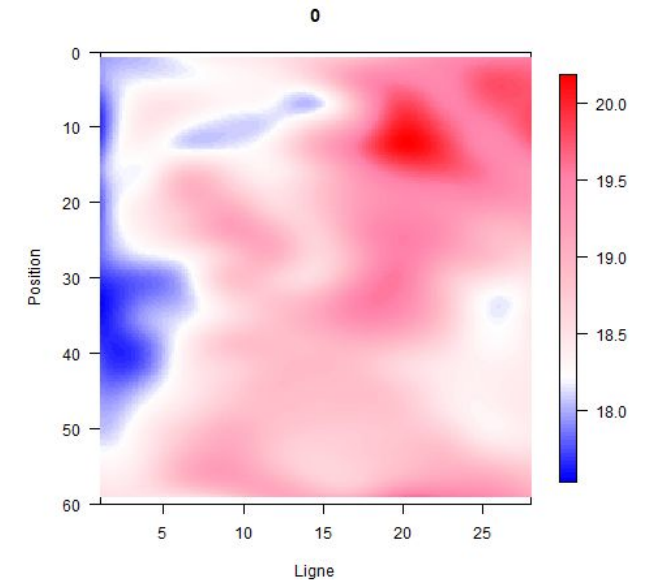
Exemple 2: Environmental quantification

AIM : Calculate thermal time from sowing for each plant in an experiment

$$ThermalTime = \sum_i^n T_{air} - T_b$$

Apparently easy but...

- Plants sown and harvested in batches at different dates
- Plants belonging to different crops
- Plants located in different positions of the greenhouse
- Sensors at different positions displaying significant variation in temperature



Exemple 2: Environmental quantification

AIM : Calculate thermal time from sowing for each plant in an experiment

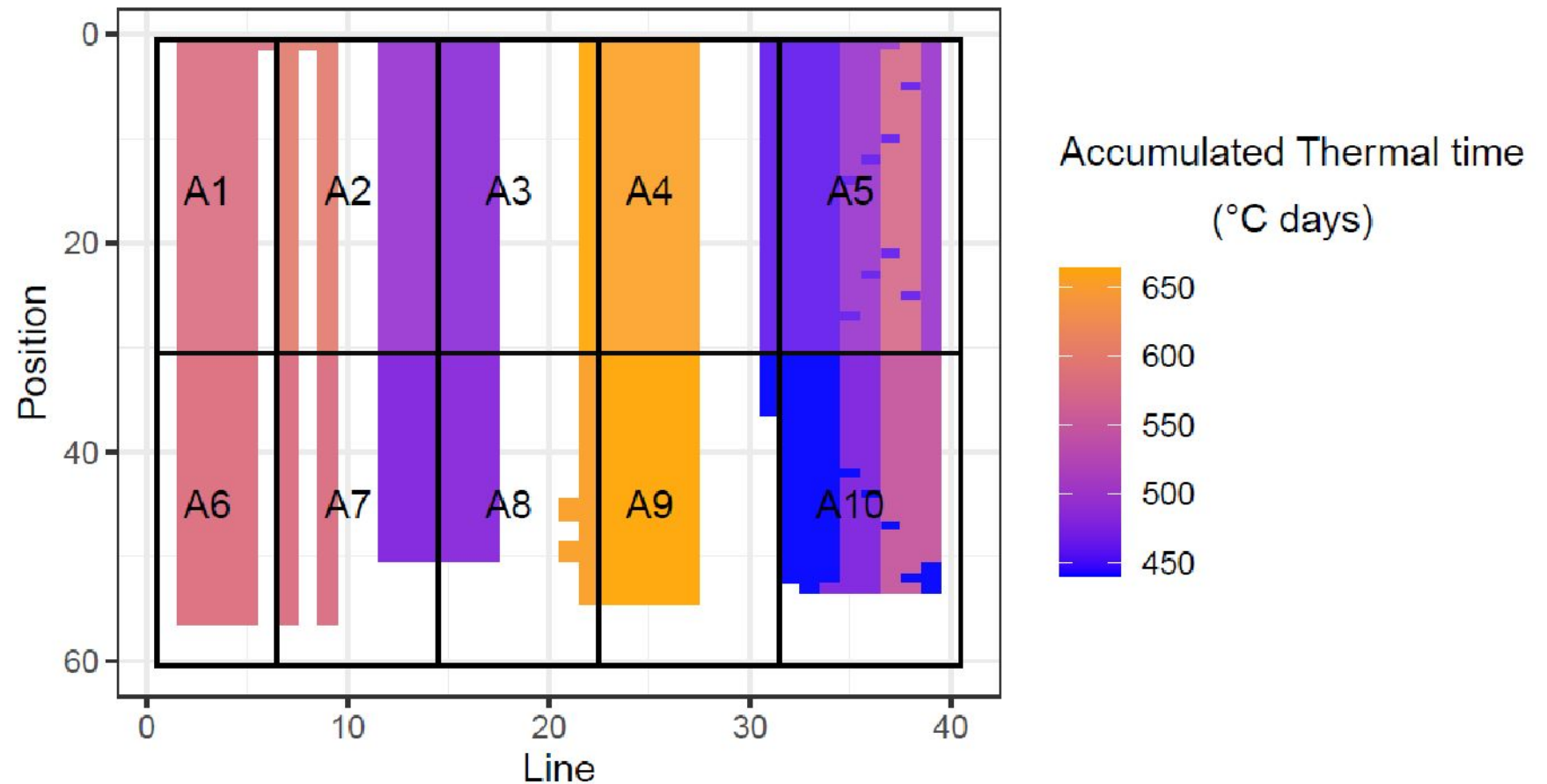
$$\textit{ThermalTime} = \sum_i^n T_{air} - T_b$$

WE NEED

- Plants sown in batches at different dates => Sowing and harvest dates for each plant (events)
- Plants belonging to different crops => T_b specific for each crop (ex. 8°C for maize, 0°C for wheat)
- Plants located in different positions of the greenhouse => tracking of plants' position
- Sensors displaying significant variation in temperature => Sensor position, variable measured and units

Exemple 2: Environmental quantification

Taking into account all this annotated date allows to calculate thermal time for each plant in an experiment



Statistical methods

Incorporation of statistical methods for platform management, data 'cleaning,' and analysis

1. **Detection of outlier plants** (bad germination, disease, seed error)

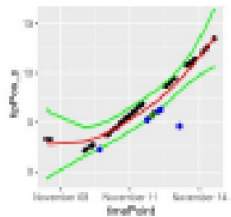
Alvarez Prado *et al.* 2019

2. **Design Generator:** User friendly online service aimed at construction and generation of experimental designs <https://eppn2020design.com/>

3. **statgenHTP R package** (outlier detection, correction of spatial trends, parameter extraction from time series)

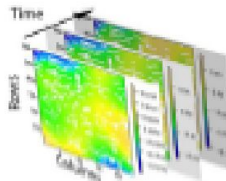


Detect outlying single observation



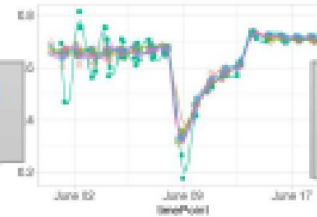
Raw clean data

Correct for spatial trend



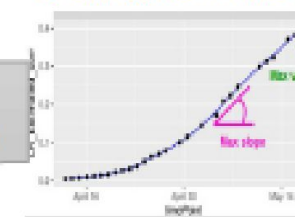
Corrected data

Detect outlying time series

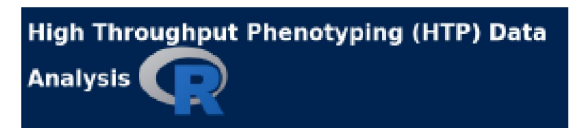


Corrected clean data

Extract parameters from time course



<https://cran.r-project.org/web/packages/statgenHTP/>
https://biometris.github.io/statgenHTP/articles/vignettesSite/Intro_HTP.html



R package: statgenHTP

WAGeningen
UNIVERSITY & RESEARCH
Emilie Millet
Bart-Jan van Rossum
Martin Boer
Fred van Eeuwijk

INRAE
Nadine Hilgert
Isabelle Sanchez

(b)cam
bioinformatics center for applied mathematics
Diana Perez
Coté Rodriguez



Spatio-temporal analysis of HTTP data

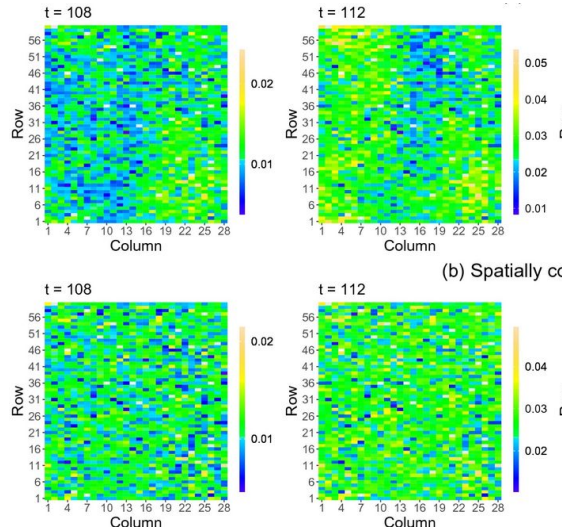
scientific reports

Article | Open Access | Published: 24 February 2022

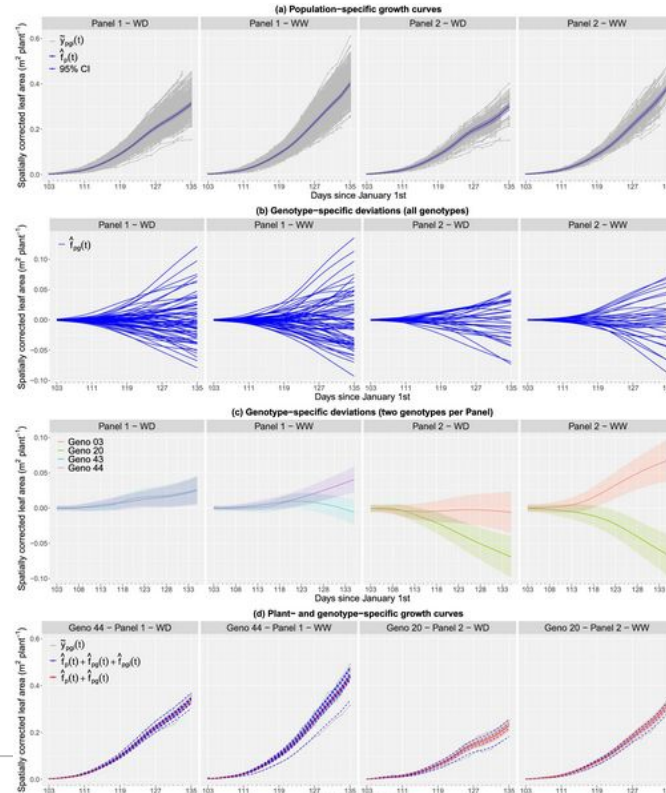
A two-stage approach for the spatio-temporal analysis of high-throughput phenotyping data

Diana M. Pérez-Valencia, María Xosé Rodríguez-Álvarez, Martin P. Boer, Lukas Kronenberg, Andreas Hund, Llorenç Cabrera-Bosquet, Emilie J. Millet & Fred A. van Feuwijk

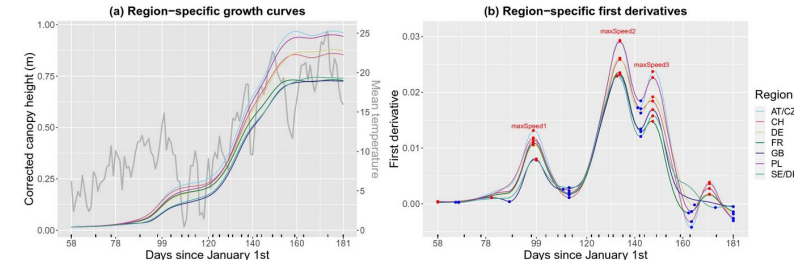
1. Design features and spatial trends correction per time point.



2. Longitudinal modelling of the spatially corrected data, thereby taking advantage of shared longitudinal features between genotypes and plants within genotypes



3. Extracting time-independent attributes to characterise genotypes



Code available



PHENOARCH - INRAE



FIP - ETH Zürich

Geospatial Events

Event with an Area as a Target

- Event ID
- Event Type
- Event Description
- Event Date / Period
- Target = Area
 - Type
 - Name
 - Description
 - **Geometry**

Geospatial Events

Event with an Area as a Target

- Event ID
- Event Type
- Event Description
- Event Date / Period
- Target = Area
 - Type
 - Name
 - Description
 - **Geometry**

Events and also...

- Any observed object
- Any device
- Any facility
- Adresse (site, institution, ...)
- ...

IS vs GIS: How to deal with spatial data?

Vector

Import

Export

Generation

Update

Attribute table

Raster

Import

Export

Generation

Processing

Georeferencing
Projection sys

Set up

Up date

Change

Search &
Geolocalisation

Maps

Generation

Symbology

Atlas

Layout

Analysis

Filter

Query

Measurements

Geostatistics

IS vs GIS: How to deal with spatial data?

Do we need a GIS to manage properly and FAIRly Geospatial phenotyping data?

What we need as minimum

- GIS: What are the limits?
 - File sizes, Transformation & Projection \Rightarrow access limits
 - DBMS : mostly relational
 - Limited performance and capacity in Big Data applications and MetaData management
 - Limited flexibility to handle complex data types and relationships

What we need as minimum

- GIS: What are the limits?
 - File sizes, Transformation & Projection \Rightarrow access limits
 - DBMS : mostly relational
 - Limited performance and capacity in Big Data applications and MetaData management
 - Limited flexibility to handle complex data types and relationships



Make a high-performance I.S. interoperable with a GIS

What we need as minimum

- GIS: What are the limits?
 - File sizes, Transformation & Projection \Rightarrow access limits
 - DBMS : mostly relational
 - Limited performance and capacity in Big Data applications and MetaData management
 - Limited flexibility to handle complex data types and relationships



Make a high-performance I.S. interoperable with a GIS



- Query & Analyse
- Exporting different formats (shapefiles, geojson, CSV,...)
- Spatial functions and indexation
- DB: PostGreSQL and extension PostGIS, Oracle Spatial, MongoDB, MySQL Spatial

What we need as minimum

Vector

Import

Export

Generation

Update

Attribute table

Raster

Import

Export

Generation

Processing

Georeferencing
Projection sys

Set up

Up date

Change

Search &
Geolocalisation

Maps

Generation

Symbology

Atlas

Layout

Analysis

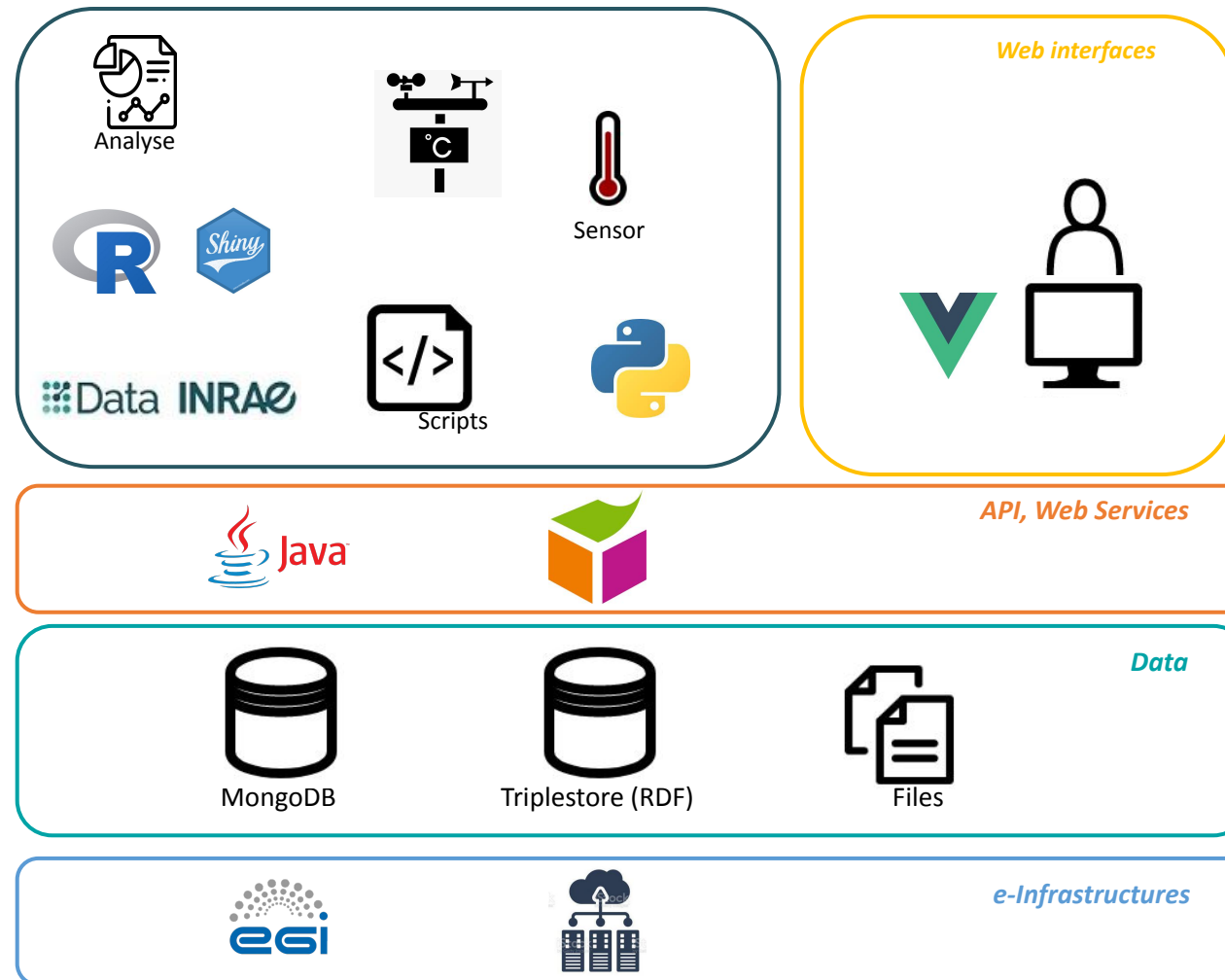
Filter

Query

Measurements

Geostatistics

How to integrate spatial data (PHIS example)



How to integrate spatial data (PHIS example)

- Implementation examples :

Spatial data storage in *MongoDB* using *SOSA* ontology

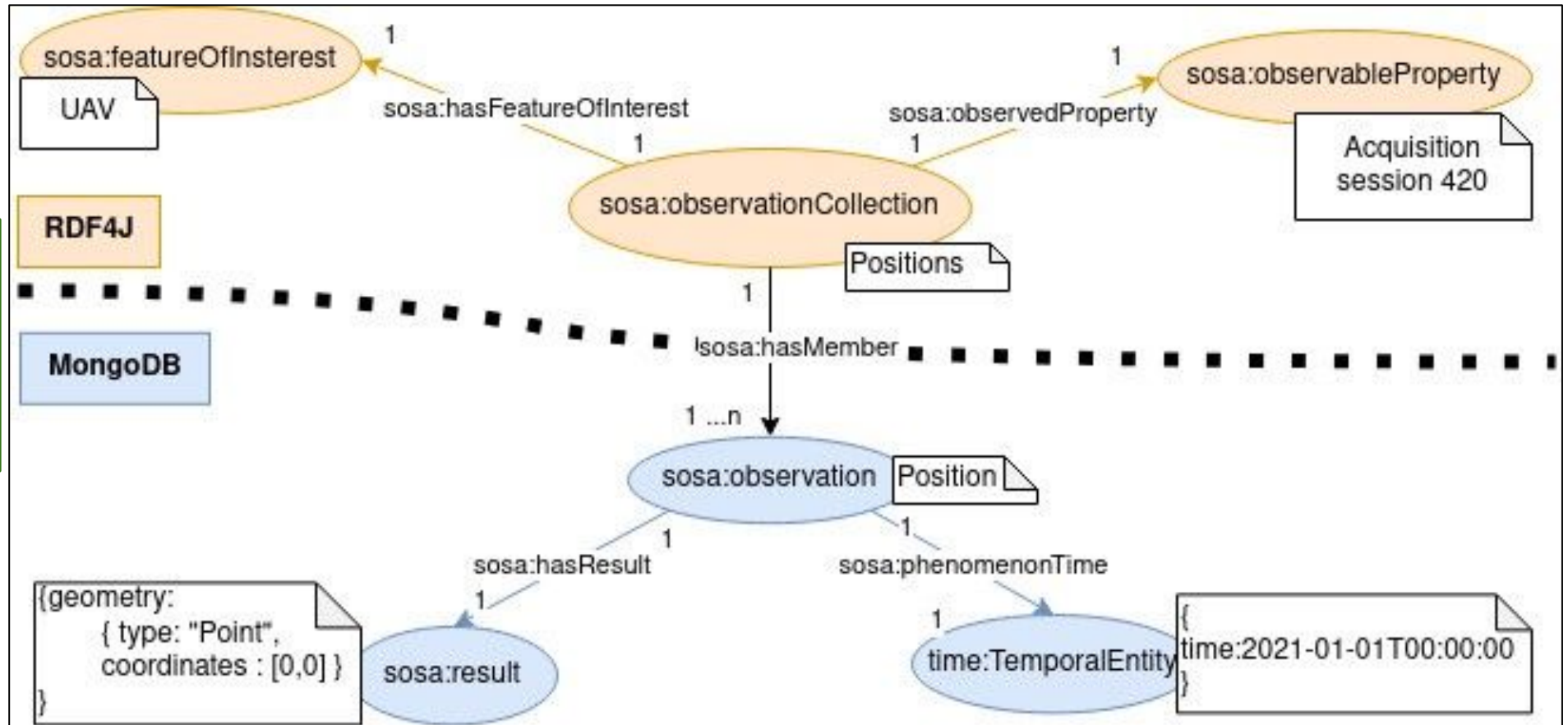
MongoDB:

- Document oriented DBMS (Object = document, organized in collections)
 - NoSQL
- Data format : BSON (binary JSON)
- Spatial data format : GeoJson (WGS84)

SOSA Ontology (Semantic Sensor Network Ontology) :

- Formal but lightweight general specification
- Modeling the interaction between entities
 - Sensor
 - Observation
 - Sample
 - Actuator
- Use the observation collection of SOSA :
 - 1 spatial data of 1 object = 1 observation
 - 1 liste of spatial data of 1 object = 1 observation collection

How to integrate spatial data (PHIS example)



How to integrate spatial data (PHIS example)

- Spatial Data format
 - Projection system: mostly WGS84 (World Geodetic System 1984 - EPSG:4326) : longitude/latitude
 - GeoJson, WKT, KML,
 - Visualization libraries: OpenLayers, MapBox, LeafLet



```
{  
  "type": "Feature",  
  "geometry": {  
    "type": "Point",  
    "coordinates": [125.6, 10.1]  
  },  
  "properties": {  
    "name": "Dinagat Islands"  
  }  
}
```

GeoJSON

```
<?xml version="1.0" encoding="UTF-8"?>  
<kml xmlns="http://www.opengis.net/kml/2.2">  
  <Document>  
    <Placemark>  
      <name>New York City</name>  
      <description>New York City</description>  
      <Point>  
        <coordinates>-74.006393,40.714172,0</coordinates>  
      </Point>  
    </Placemark>  
  </Document>  
</kml>
```

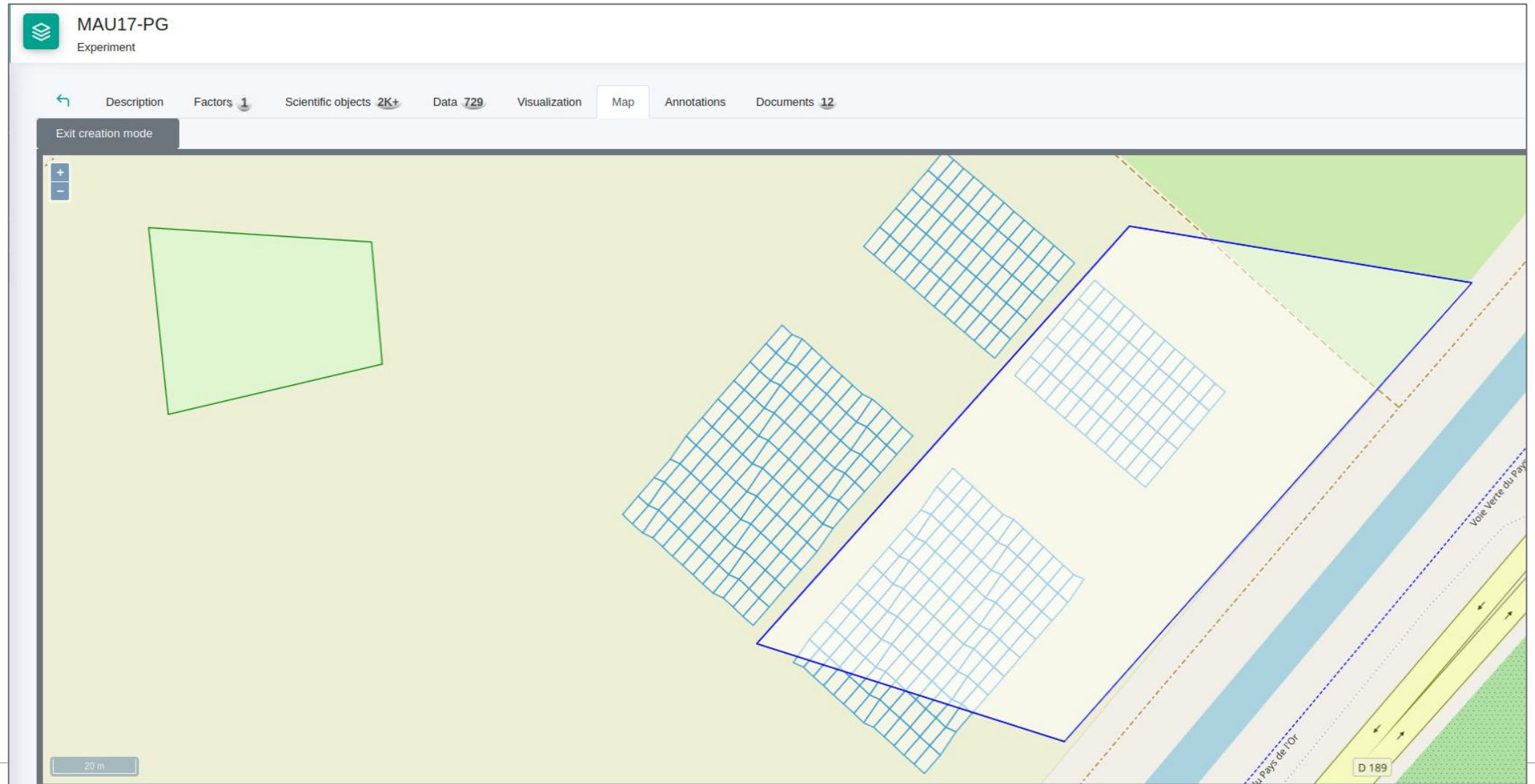
KML

```
POINT(6 10)  
LINESTRING(3 4,10 50,20 25)  
POLYGON((1 1,5 1,5 1,5,1 1))
```

WKT

Geospatial targets (Events: PHIS)

Example: Irrigation location - Draw the area



Geospatial targets (Events: PHIS)

Example: Irrigation location - Define the area

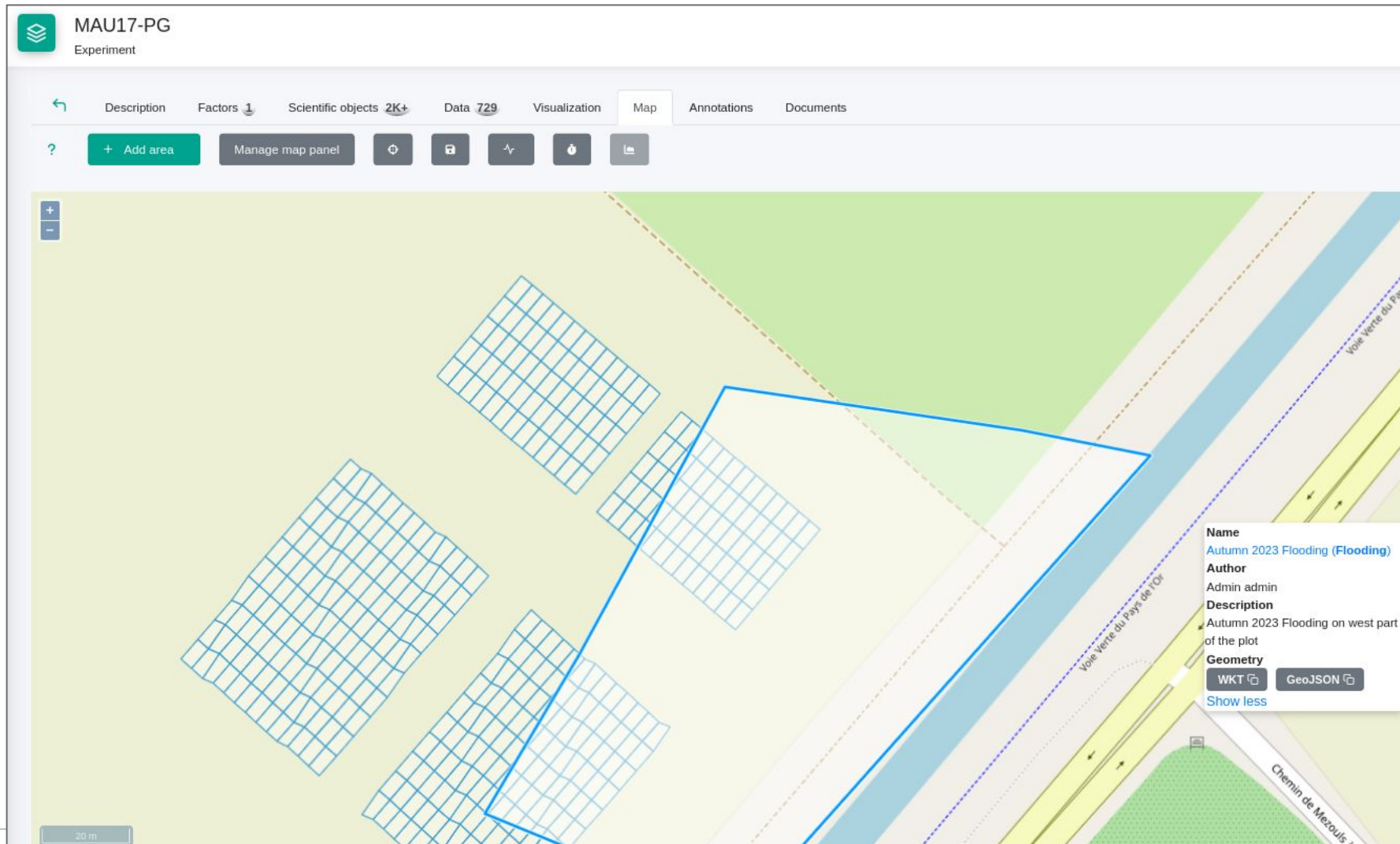
The screenshot displays a software interface for defining a geospatial area. On the left, a 'MAP PANEL' sidebar lists categories: Scientific Objects (390), Plot (390), Areas (1), Structural area (1), Temporal area (0), Devices (0), and Filters (0). A '+ Create filter' button is visible. The main map area shows a green polygon on a plot. A 'Description of the area' dialog box is open, containing the following fields:

- Area URI:** autogenerated URI
- Name:** Autumn 2023 Flooding
- Type of area:** Structural area, Temporal area
- Type:** Flooding
- Instantaneous event:**
- Begin:** 10/20/2023 17:07
- End:** 10/21/2023 14:00
- Description:** Autumn 2023 Flooding on west part of the plot

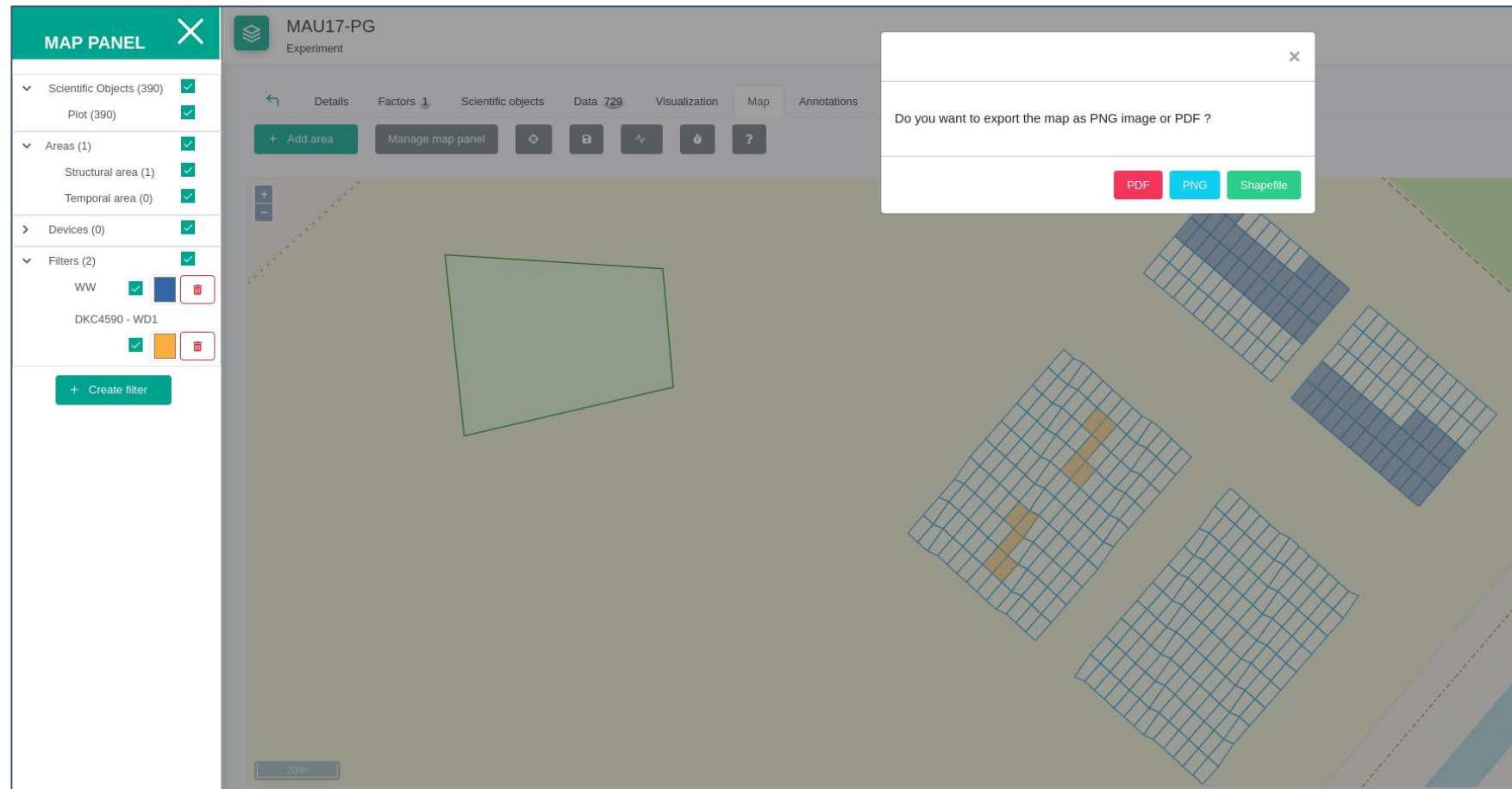
Buttons for 'Cancel' and 'OK' are located at the bottom right of the dialog box.

Geospatial targets (Events: PHIS)

Example: Irrigation location - Visualize the area



Geospatial targets (Objects: PHIS)



PHIS to QGIS (objects, devices, events)

The image displays two overlapping screenshots of the QGIS software interface. The top screenshot shows a file explorer window titled "Ouvrir des jeux de données vectorielles gérés par GDAL" with a table of files:

Nom	Taille	Type	Modifié
OS.shp			22 mai
Areas			22 mai
Devices			22 mai

The bottom screenshot shows the main QGIS workspace with a map view. The left sidebar contains the "Explorateur" (Project Tree) and "Couches" (Layers) panels. The "Couches" panel lists the following layers:

- POINT
- POLYGON
- OS_Point_2023-05-22
- OS_LineString_2023-05-22
- OS_Polygon_2023-05-22

The map view shows a terrain map with several green grid-like polygons and red star-shaped points overlaid on it. A blue arrow points from the "OS.shp" file in the top screenshot to the "OS_Polygon_2023-05-22" layer in the bottom screenshot.

And then?

- I.S. managing all data including spatial data
 - Import and export
 - Query
 - Visualization
 - Interoperable with GIS

⇒ **Meta analysis & Geostatistic**

⇒ **Historical data as well as data for the current year**

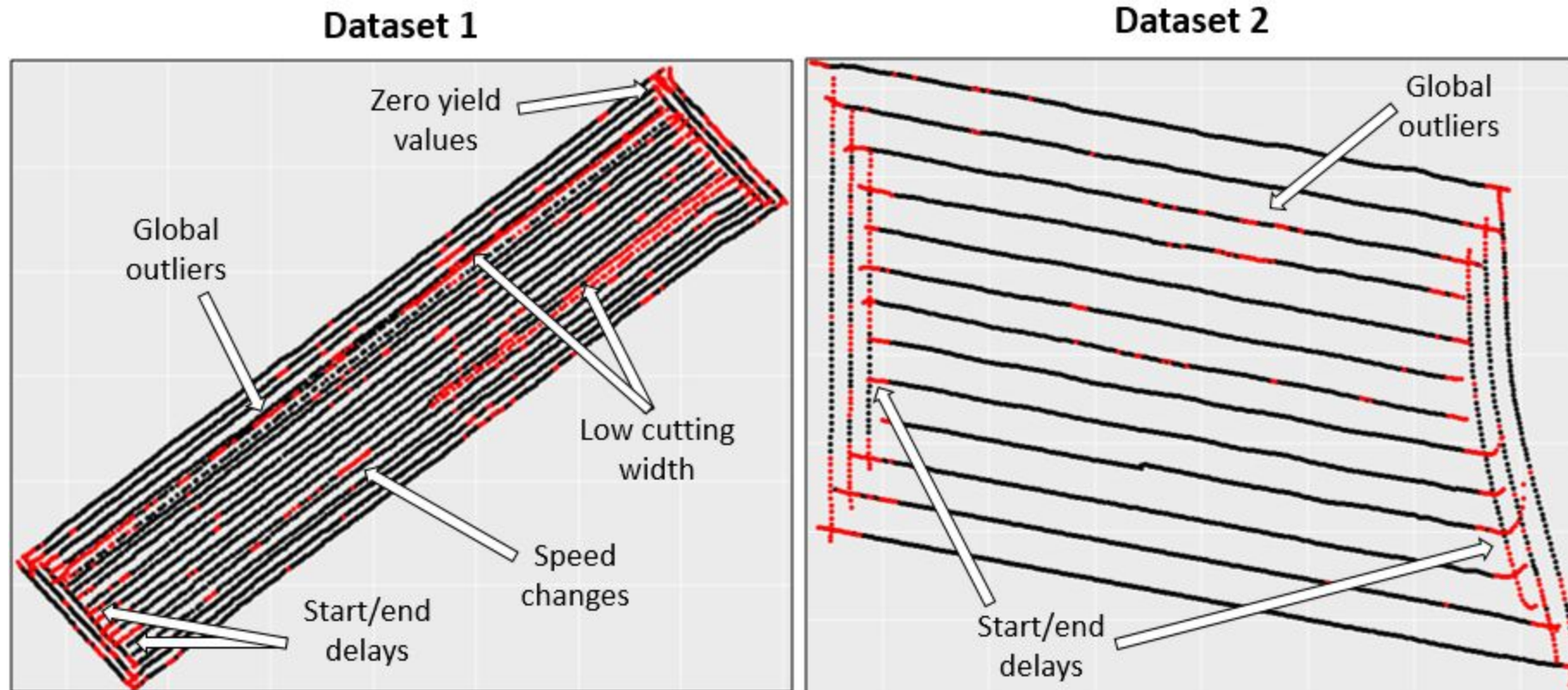
Geostatistics & Spatial tools: Zoning & beyond

Mostly used for yield data ... and other spatial data!

- More and more agronomic variables
- Possibility of adaptive management based on intra-field variability
 - How and where implement an experiment?
 - How to filter and clean spatial data?
 - How to validate the relevance of the outlier detection and filtering?
 - How to integrate expertise (agricultures, advisors, ...)?
 - How to simulate a spatial data set based on historical and current data?
 - How to generate recommendation maps (Decision-Support Tools)

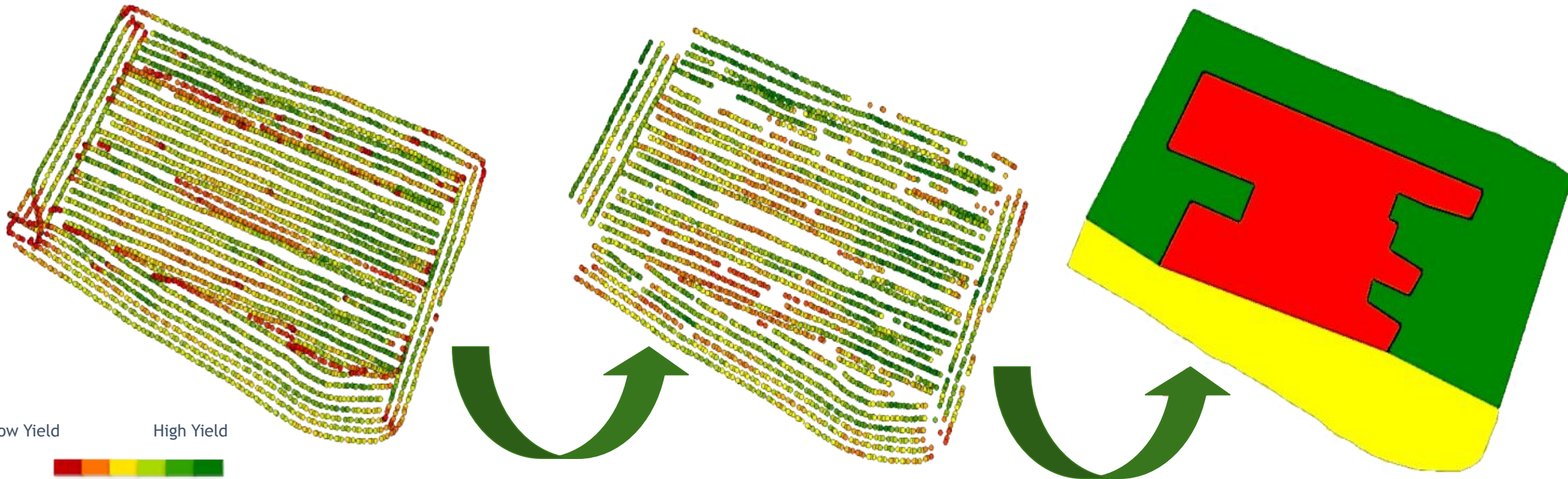
Geostatistics & Spatial tools: Outlier detection & filtering

Mostly used for yield data ... and other spatial data! (C. Leroux & H. Jones)



Outlier detection, filtering & zoning (fertilization)

Mostly used for yield data ... and other spatial data! (C. Leroux & H. Jones)

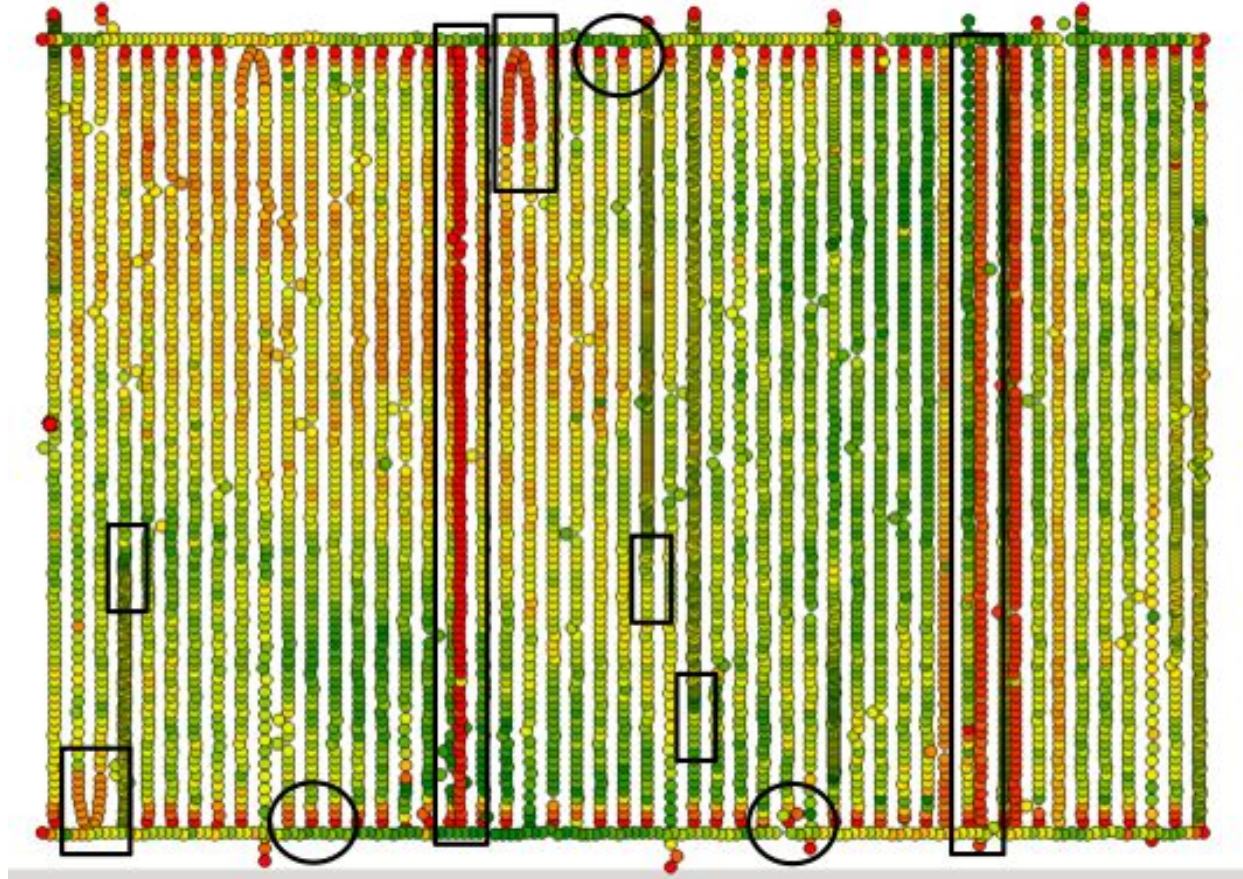


- Outlier deleting
- Reassessment of yield characteristics

- Operational zoning for yield data (clustering)
- Fertilization application

Geostatistics & Spatial tools: Yield map simulation

Mostly used for yield data ... and other spatial data! (C. Leroux & H. Jones)



Geostatistics & Spatial tools

Zoning: Historical phenotyping data for experiment implementation



- Maximizing:
 - Inter zones heterogeneity
 - Intra zone homogeneity
- Shape constraints

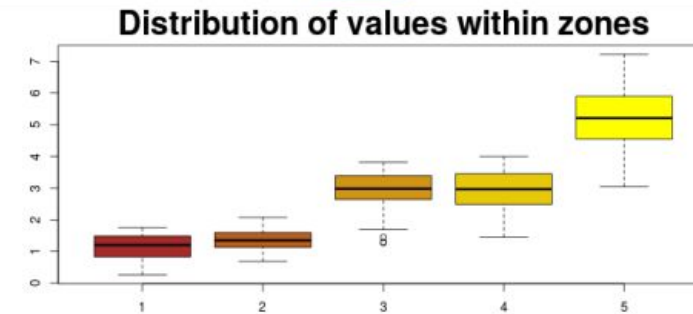
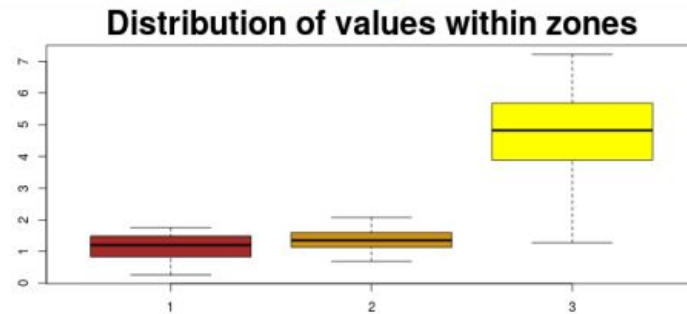
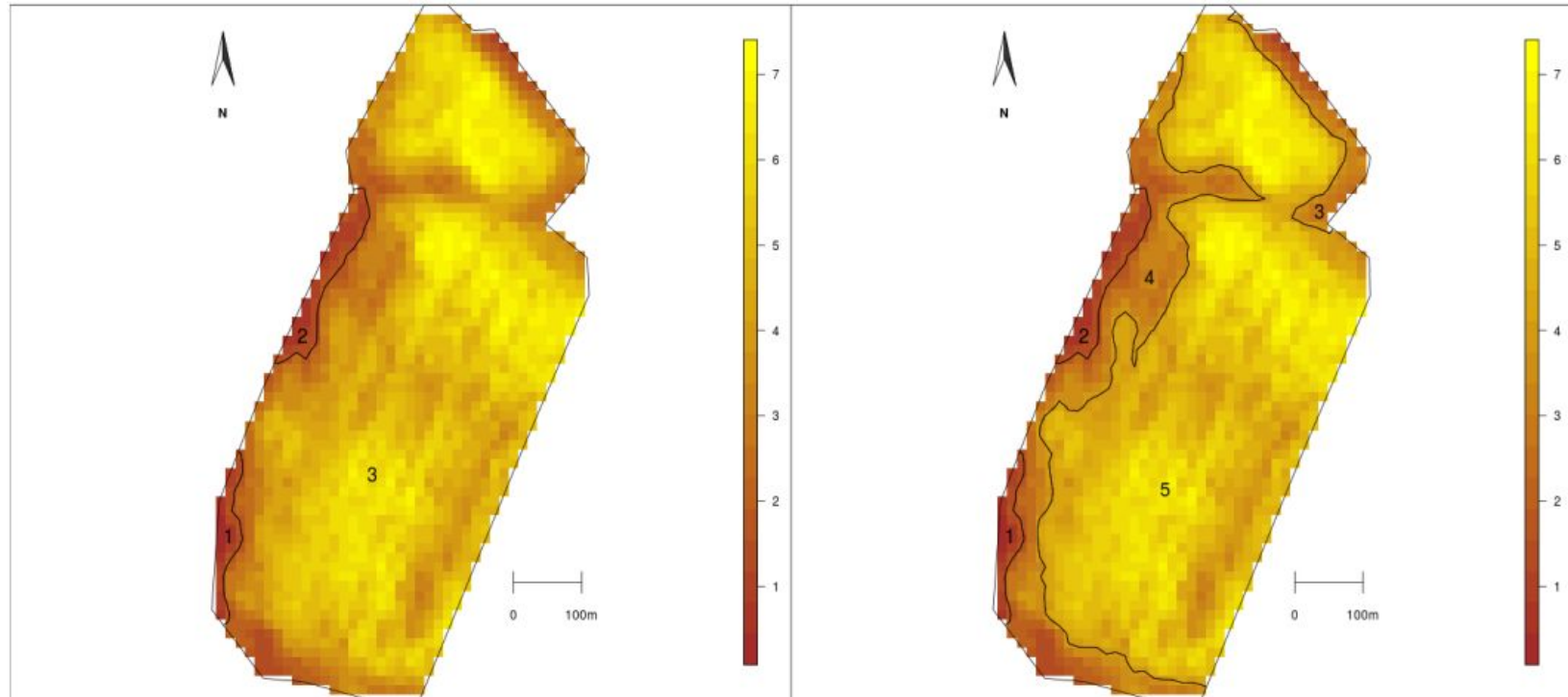
Variable: Yield, Soil parameters, Biomass, ...

Geostatistics & Spatial tools: Zoning

Zoning: Historical phenotyping data for experiment implementation

Value: Yield (N-1)

(C. Leroux & H. Jones)

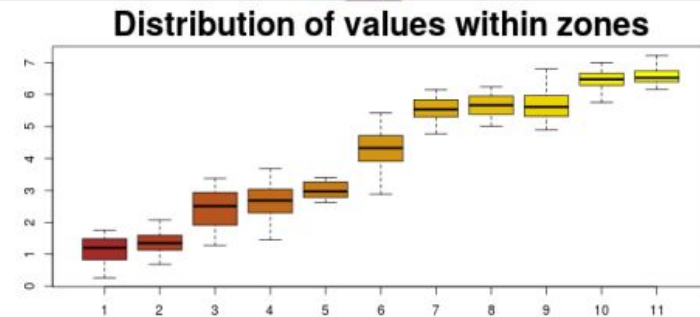
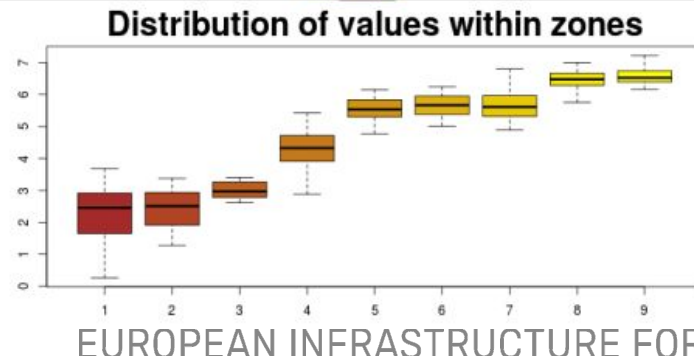
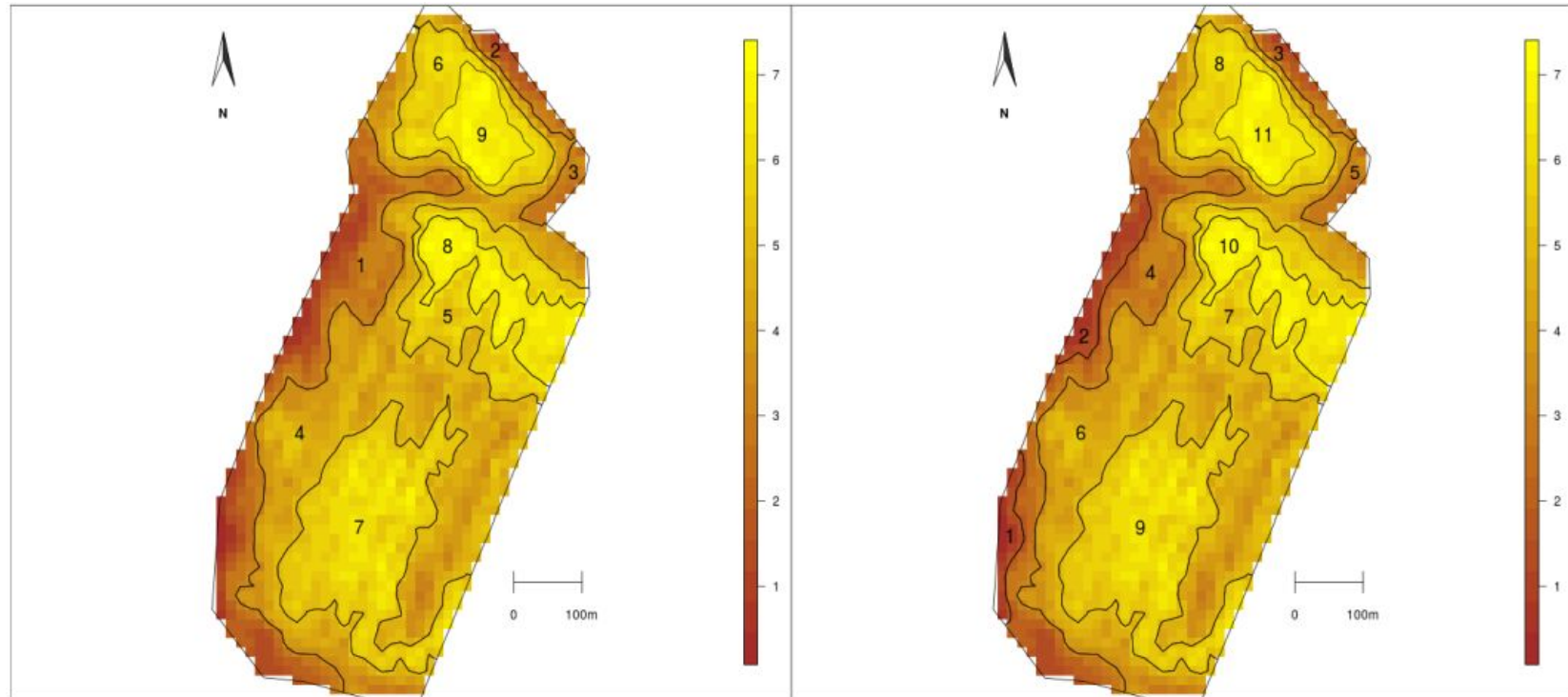


Geostatistics & Spatial tools: Zoning

Zoning: Historical phenotyping data for experiment implementation

Value: Yield (N-1)

(C. Leroux & H. Jones)

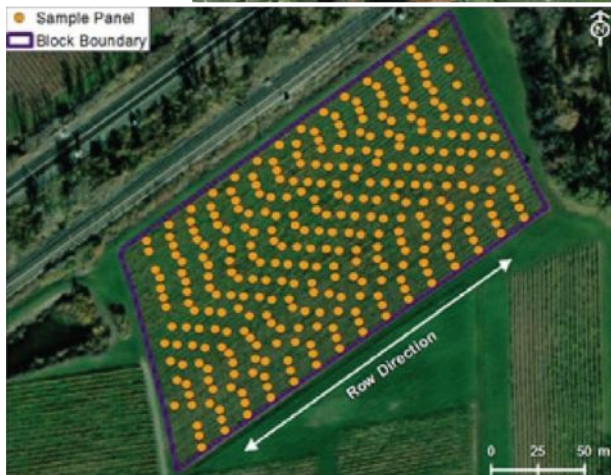


Geostatistics & Spatial tools

Vineyard yield forecasting based on machine learning methods (H. Jones et al.,)

Data (historical & spatio-temporel)

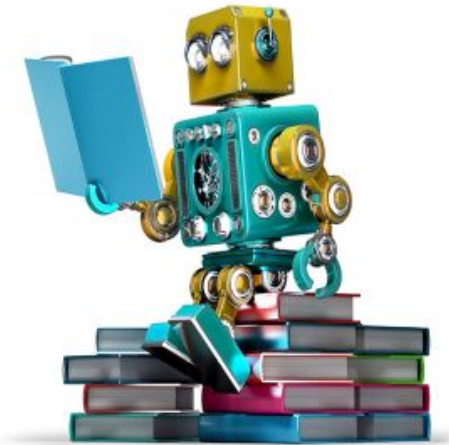
- Yield sensors
- Crop sensors (VIs)
- Soil sensors
- Pruning weight



Method Random Forest

Machine learning algorithm

- Supervised learning (need for examples for inputs/output to learn)
- Possible to use on
 - Classification (predict a class)
 - Regression (predict a value)

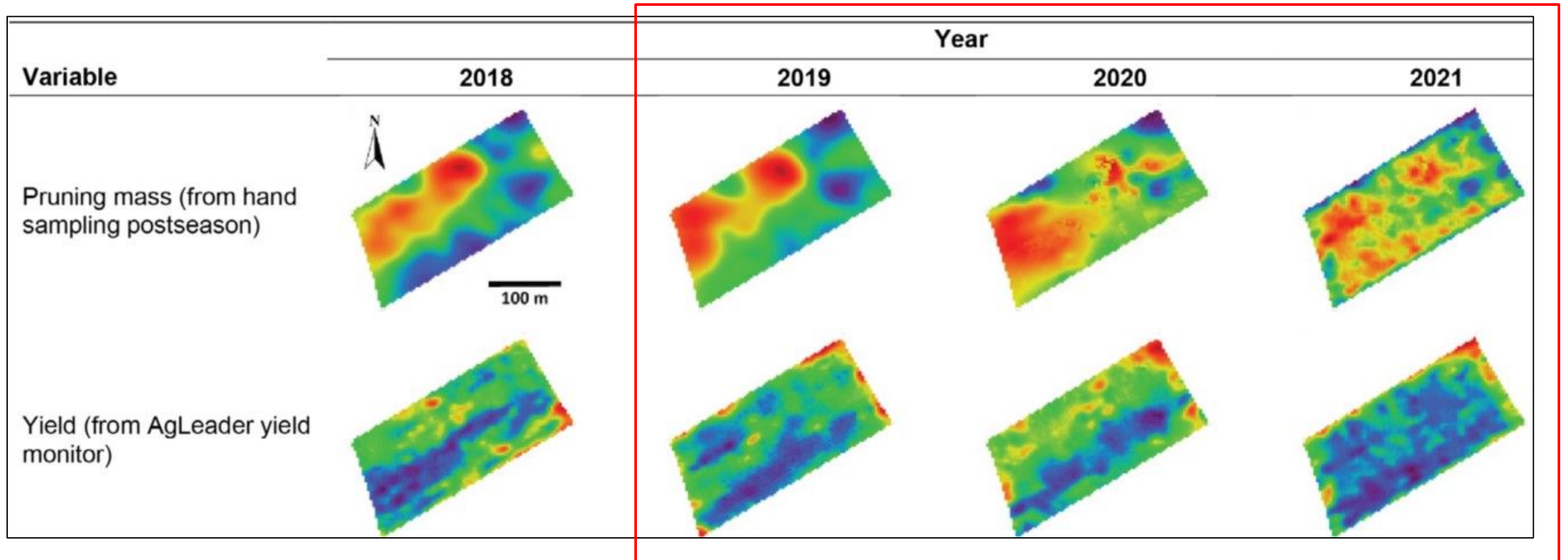


What & When are the best data for yield reporting?

Geostatistics & Spatial tools

Vineyard yield forecasting based on machine learning methods (H. Jones et al.,)

What is to be predicted?

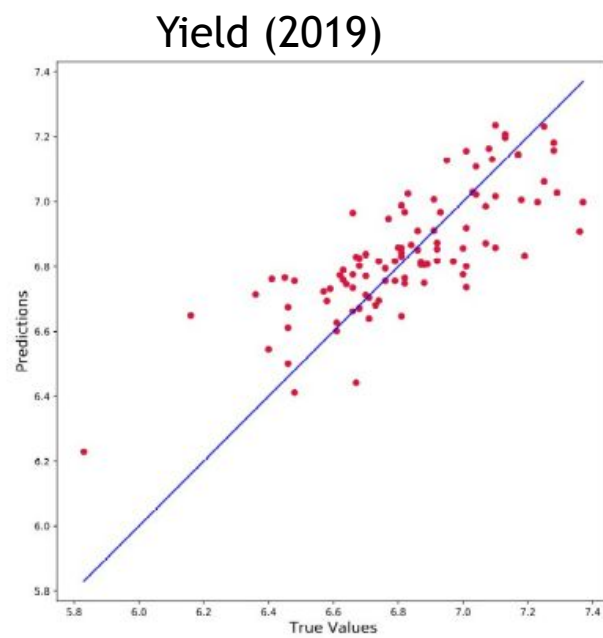


Geostatistics & Spatial tools

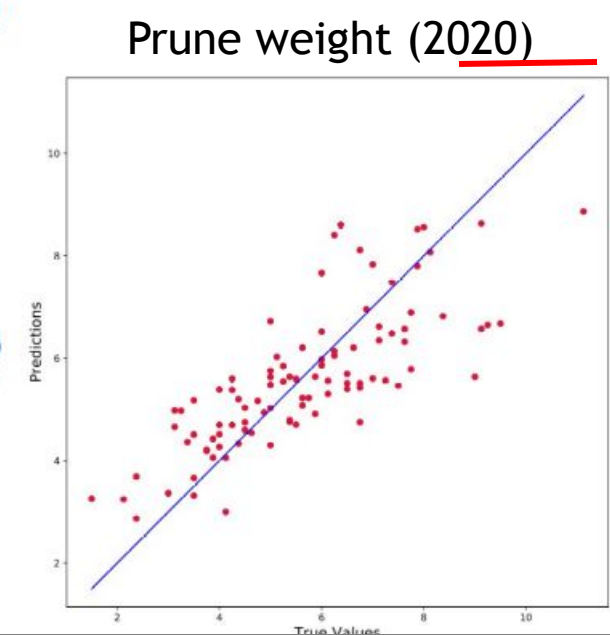
Vineyard yield forecasting based on machine learning methods (H. Jones et al.,)

Some results...

- Prediction on yield (2019)
 - Mean absolute error: 0.127
 - Mean absolute percentage error: 0.0187
 - explained_variance: 0.608
 - range y_test: [5.83 7.37]
- **feature importance :**
 - SR_310519 0.079783
 - DifVI_170619 0.058389
 - MSR_170619 0.053485



- Prediction on Prune Weight (2020)
 - Mean absolute error: 0.887
 - Mean absolute percentage error: 0.174
 - explained_variance: 0.613
 - range y_test: [1.5 11.125]
- **feature importance :**
 - PW_LbsperPanel_2019 0.281330
 - CropLoad2019_AgLeader 0.27707



Conclusion & Perspectives

Think FAIR

Structure Data

Use Standards

Think of MetaData

Use Ontology

Use semantic

Use appropriate DBMS

Use appropriate storage systems

Make good DMP

GDPR compliant

Conclusion & Perspectives

Think FAIR

Structure Data

Use Standards

Think of MetaData

Use Ontology

Use semantic

Use appropriate DBMS

Use appropriate storage systems

Make good DMP

GDPR compliant

Seems infeasible? Ask for help!

- Information systems
- Ontology
- Pipelines for data transmission & processing
- Mapping with standards
- DMP & GDPR setting
- Spatio-temporal data for advanced analysis (variables)



Conclusion & Perspectives

- Start changing habits and adopt appropriate methods & tools
- Trainings, workshops & Hackathons
 - PHENET-EMPHASIS Joint training event in 2024
 -  Training materials and survey to set the dates and content
 - BrAPI Hackathon in 2024 - <https://brapi.org/events/hackathon>
 - BioHackathon in 2024 - <https://biohackathon-europe.org/>
 - PhenoHarmonIS - Montpellier May 2024

Conclusion & Perspectives

- Start changing habits and adopt appropriate methods & tools
- Trainings, workshops & Hackathons
 - PHENET-EMPHASIS Joint training event in 2024
 -  Training materials and survey to set the dates and content
 - BrAPI Hackathon in 2024 - <https://brapi.org/events/hackathon>
 - BioHackathon in 2024 - <https://biohackathon-europe.org/>
 - PhenoHarmonIS - Montpellier May 2024

Contact us:

opensilex@gmail.com

isabelle.alic@inrae.fr

farzaneh.kazemipour@inrae.fr

pascal.neveu@inrae.fr

cyril.pommier@inrae.fr



INRAE

PHENOME
EMPHASIS | FRANCE

<https://www.phenome-emphasis.fr/>

Mistea
Mathématiques, Informatique et Statistique
pour l'Environnement et l'Agronomie

URGI

LEPSE
PLANT ADAPTATION
TO CLIMATE CHANGE



CAPTE



OpenSILEX Team - <http://opensilex.org/>

Special thanks to: Silvana Moscatelli, François Tardieu, Alexia Chiavarino, Sarra Abidri, Anne Tireau, Hazaël Jones...

 emphasis@fz-juelich.de

 emphasis.plant-phenotyping.eu

 EMPHASIS_EU

 EMPHASIS.EU

 EMPHASIS on Plant Phenomics



EMPHASIS is an ESFRI-listed project.



EMPHASIS-PREP is funded by the European Union (Grant Agreement: 739514).

EUROPEAN INFRASTRUCTURE
FOR PLANT PHENOTYPING