

Systemes d'Information pilotés par des Ontologies

Contact: Pascal.Neveu@inra.fr





Data Challenges

More and more data!

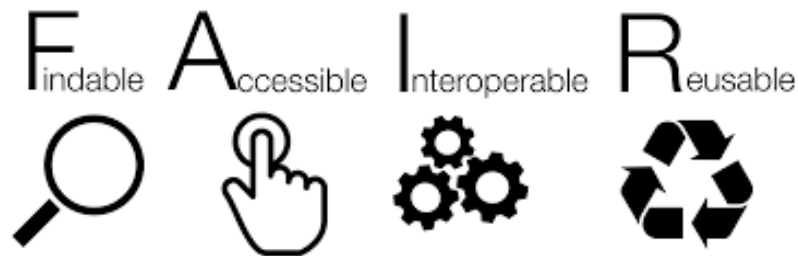
- Evolution of storage and network technologies
1 Gigabyte : \$400K in 1980, \$10 in 2000, \$0,01 in 2017
- Plenty and various devices, IoT, simulations, crowdsourcing, etc.
- Internet data sources (Open, partnerships,)

Make data valuable!

- **Decision support**
- **Knowledge discovery**
- **New services**

- *Population treatment → individualized treatment*
- *When data did not quite match what we expect!*
- *Which theories/models are consistent and which ones are not !*
- ...

Need: A new generation of Information Systems



Findable: **PID**, standardized metadata and indexed in portals

Accessible: open and standardized protocols (internet protocols), authentication* (if not open)

Interoperable: standardized formats and shared vocabularies (technology, syntactic, **semantic**)

Reusable: provenance, domain relevant **metadata for understanding across disciplines**

High Throughput Plant Phenotyping: searching for the most adapted genotypes

High frequency observation of many genotypes in various environments

Searching for the most adapted species/varieties for field challenges

- Food security
- Climate Change adaptation
- AgroEcology
- Reduce inputs / natural resource preservation
- Safe and healthy food
- ...

Decision support

- Links genomics with plant ecophysiology and agronomy
- Phenotype-driven gene function discovery

European Research e-infrastructure

- Deals with several Petabytes of Phenomics data
- Allow data reuse and data integration

FAIR data → Open technologies and standards
(MIAPPE, BrAPI, etc)

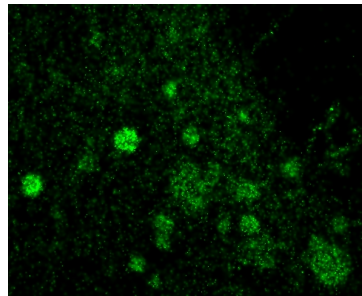
- ✓ Standardized Identification
- ✓ Standardized Semantic
- ✓ Provenance and reproducibility data processing





Phenomics Data

Different scales



Intra-cellular



Organ



Plant



Field

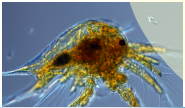


Region



Phenomics Data

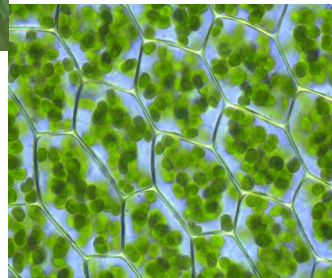
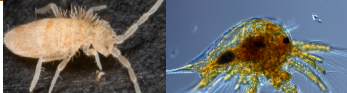
Different interactions





Phenomics Data

Different stages and transformations





Phenomics Data

Various contexts

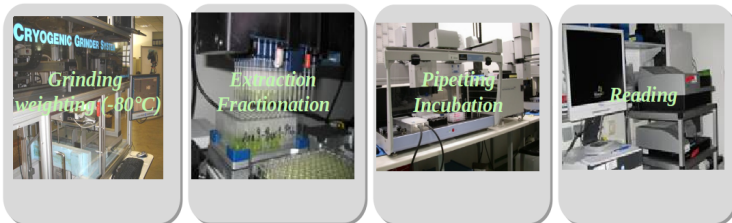
« omics » Platforms

Various data complex types

Genomics

Composition and the structure of biopolymers

Quantification of metabolites and enzyme activities



Field Platforms

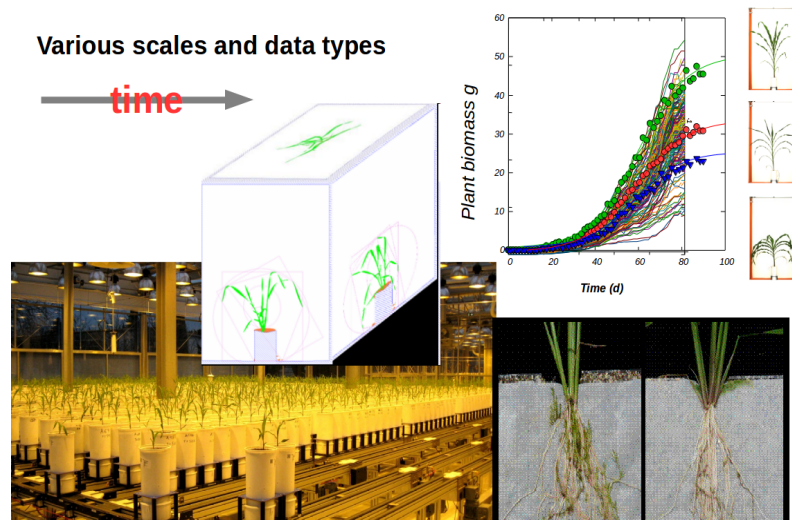
Various scales and data types

- Cell, organ, plant, population
- Images, hyperspectral, spectral, sensors, human readings...



Green house Platforms

Various scales and data types



Farm Platforms

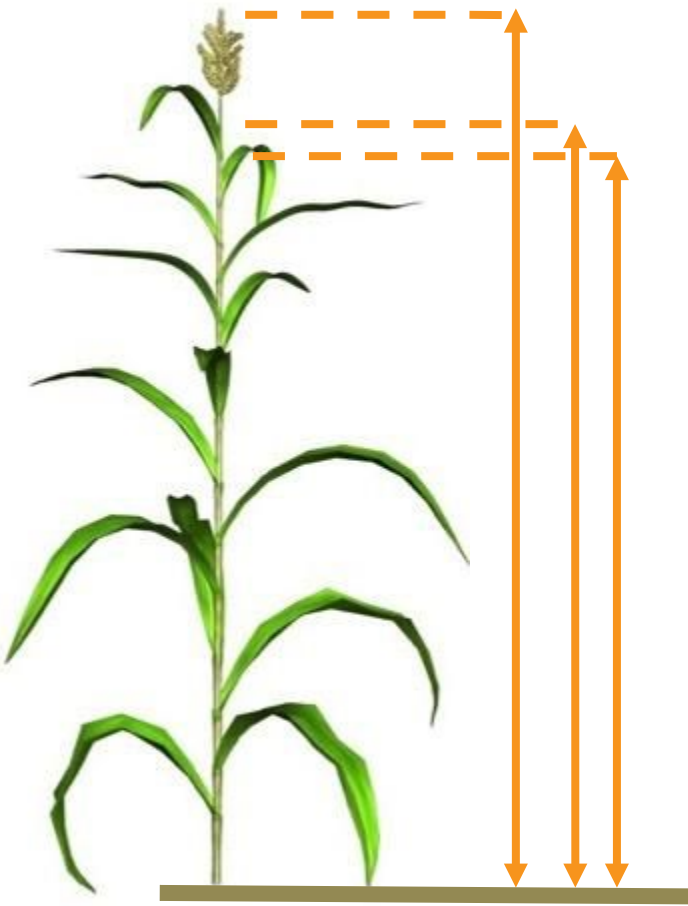
Various scales and data types from thousands of farms

- organ, plant, population, site
- Images, sensors, human readings...



Phenomics Data

Different ways



The example of *Plant Height*

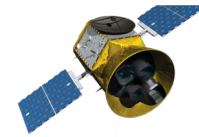
- Plant with root?
- Plant with flower?
- Stem and leaves?
- Only the stem?
- Dry or not?
- In the morning or the end of the day?

Figure From D. Pot, CIRAD

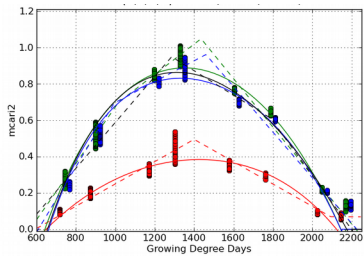
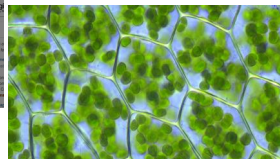
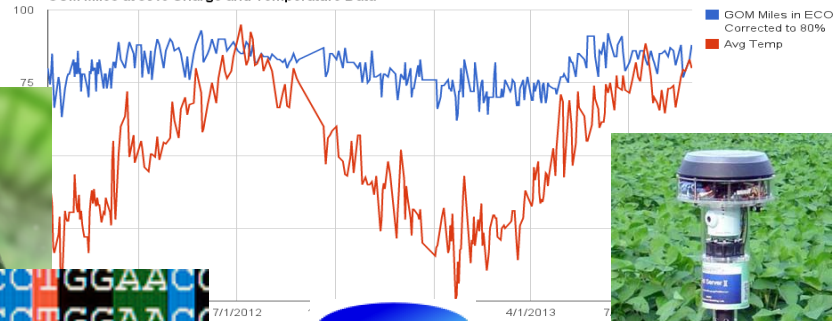
Plant Height



Phenomics Data

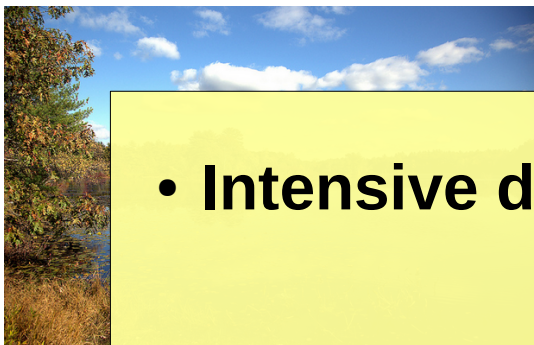
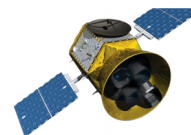


GOM Miles at 80% Charge and Temperature Data

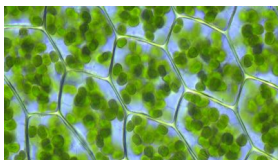
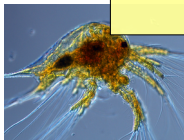




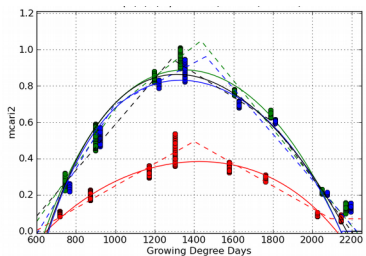
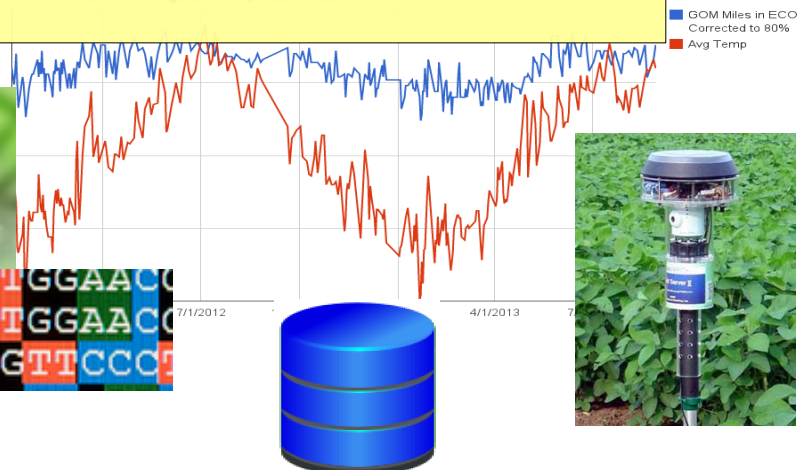
Phenomics Data



- Intensive data integration needs

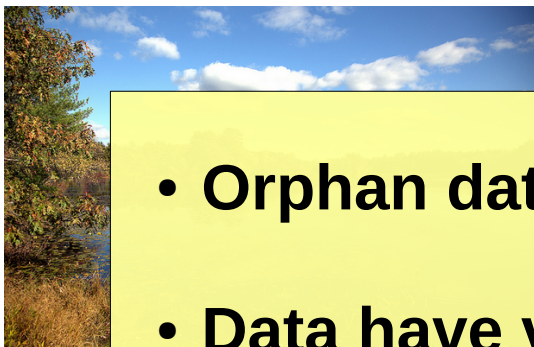
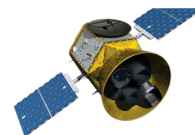


GOM Miles at 80% Charge and Temperature Data

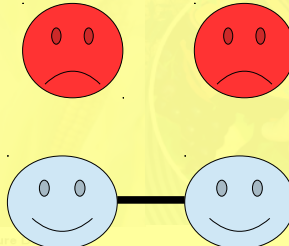
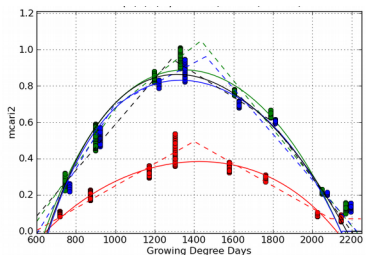
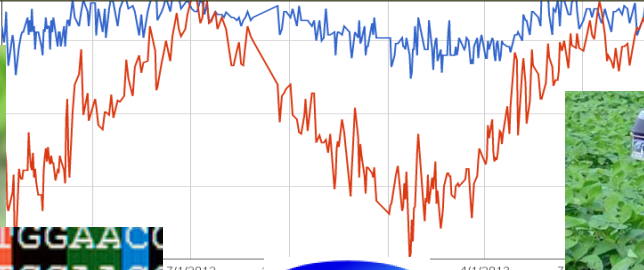
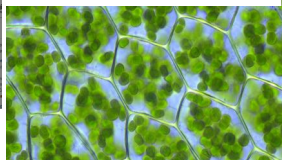
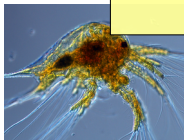




Phenomics Data



- Orphan data → Worthless!
- Data have value if they are grouped



Phenomics Data

How to structure data?

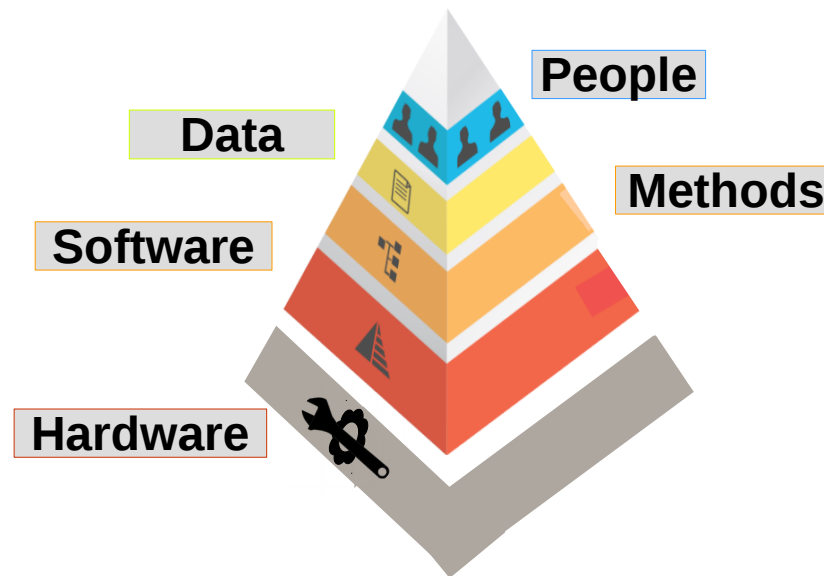




OpenSILEX

OpenSILEX is an Open source software set

- Methods, tools, components to implement information systems for experimental data in agriculture and environment
- for organisation, collection, structuration, storage, exchange and treatment of information





OpenSILEX - PHIS

→ PHIS is an instance of OpenSILEX

→ Designed for data management in phenotyping platforms

- Management of huge, complex and heterogeneous data (millions of images, sensor data, from different sites, etc)

→ Implement good practices of data management

- Make FAIR data
- Flexible
- Ability to understand and reproduce data processing
- Ability to enforce DMP and Open Science





OpenSilex approach

Scientific objects (plant, plant organ, plot, etc.)

Events (management, faults, meteo, etc :)

Equipement (Facilities, Softwares, sensors, etc.)

Organizational (experiement, study, project, etc.)

- Identified by **URI** → standardized, unambiguous, shared, etc

Organisation and linking → objects and events with a controlled semantic (Ontology) such as a context specific **application Ontologies (RDF*, OWL*)** and allows to **reuse or link reference ontologies (SKOS*)**

Measurements, Documents, Observations, Metadata are associated with these Objects and Events

* Semantic Web languages

OpenSILEX-PHIS Identification

URI: string used to identify a resource (Web standardized syntax, prefix)

`http://www.phis.inra.fr/path/identifier`

- Unambiguous
- Resolvable (dereferencing)
- Persistence (ePIC)

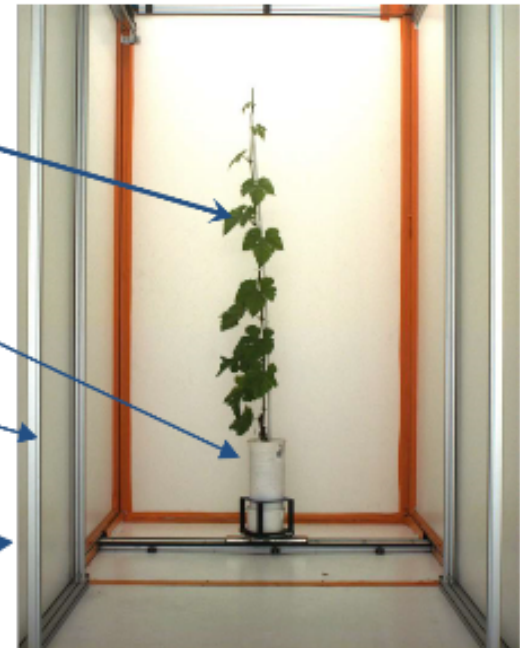
URI of plant :
`mp3:arch/2014/pl/000000012`

URI of pot :
`mp3:arch/2001/pt/000001542`

URI of cabin :
`mp3:arch/2010/ca/cabine2`

URI of camera :
`mp3:arch/2011/ss/00003312`

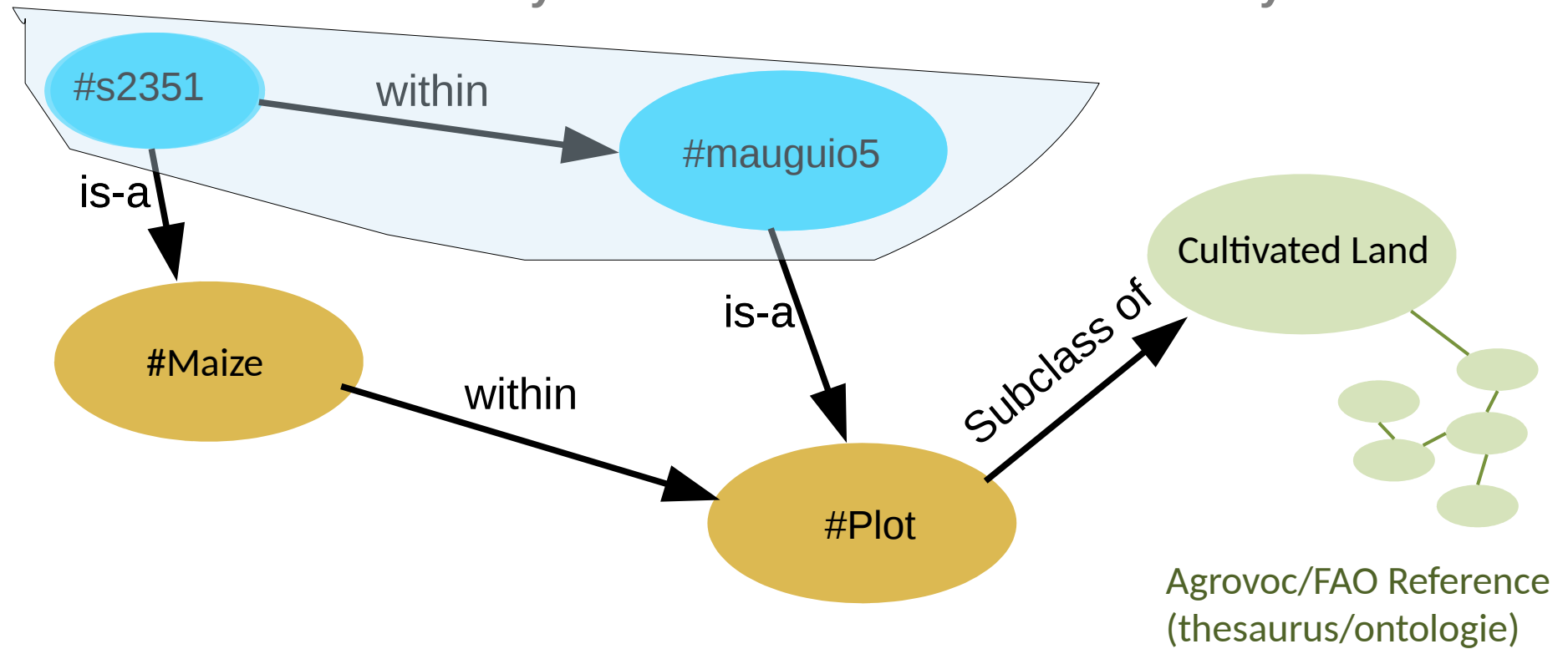
URI of image :
`mp3:arch/2015/im/000000564`





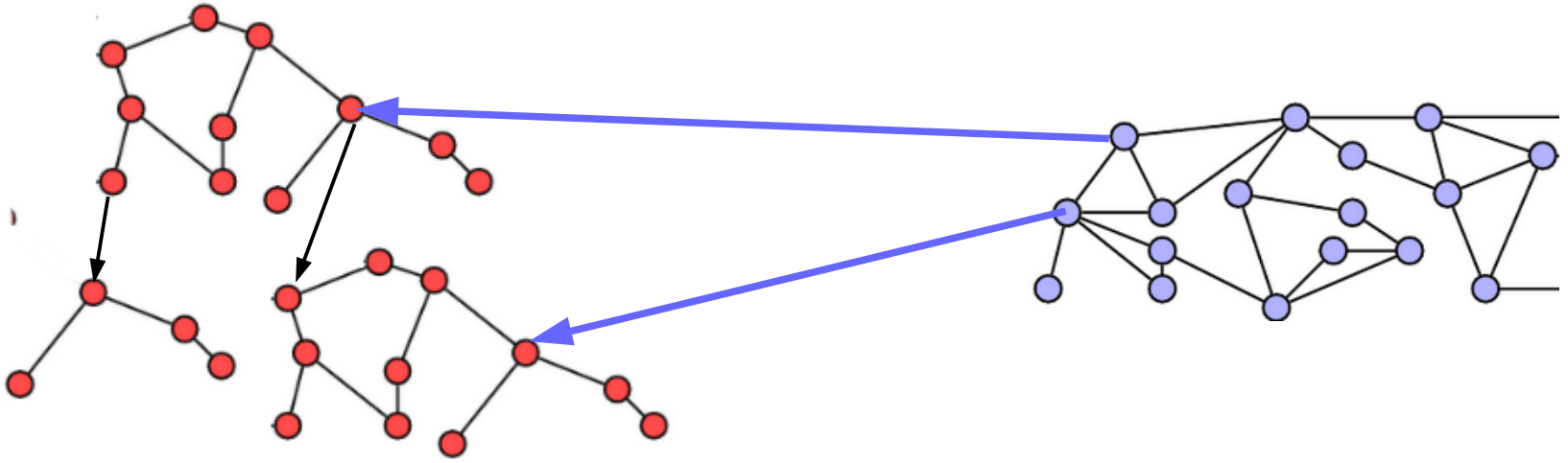
OpenSILEX - PHIS

- Metadata / ontologies provide the meaning of data
 - Link each data element to a controlled, shared, vocabulary and **machine readable** vocabulary





OpenSILEX - PHIS

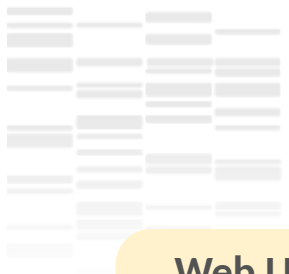


Reference ontologies

Application ontologies

Published in





OpenSILEX - PHIS

Web User Interface

Software agents

Web Service LAYER

Semantic Services

Data LAYER

NoSQL database mongoDB.

Triplestore **rd4j**

e-infrastructure LAYER

ESI Distributed storage system

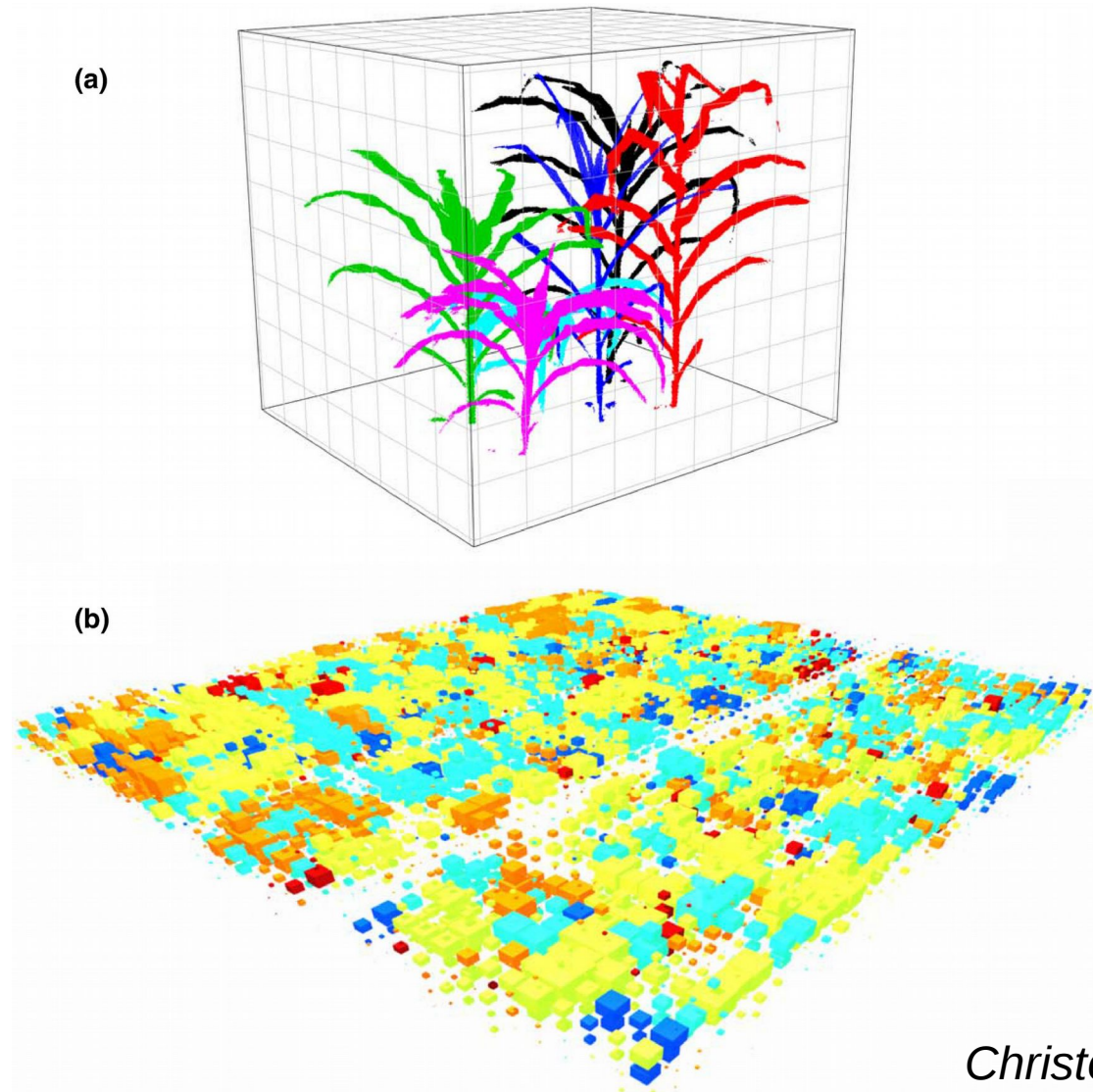
Scientific Computation and Workflow LAYER

Galaxy

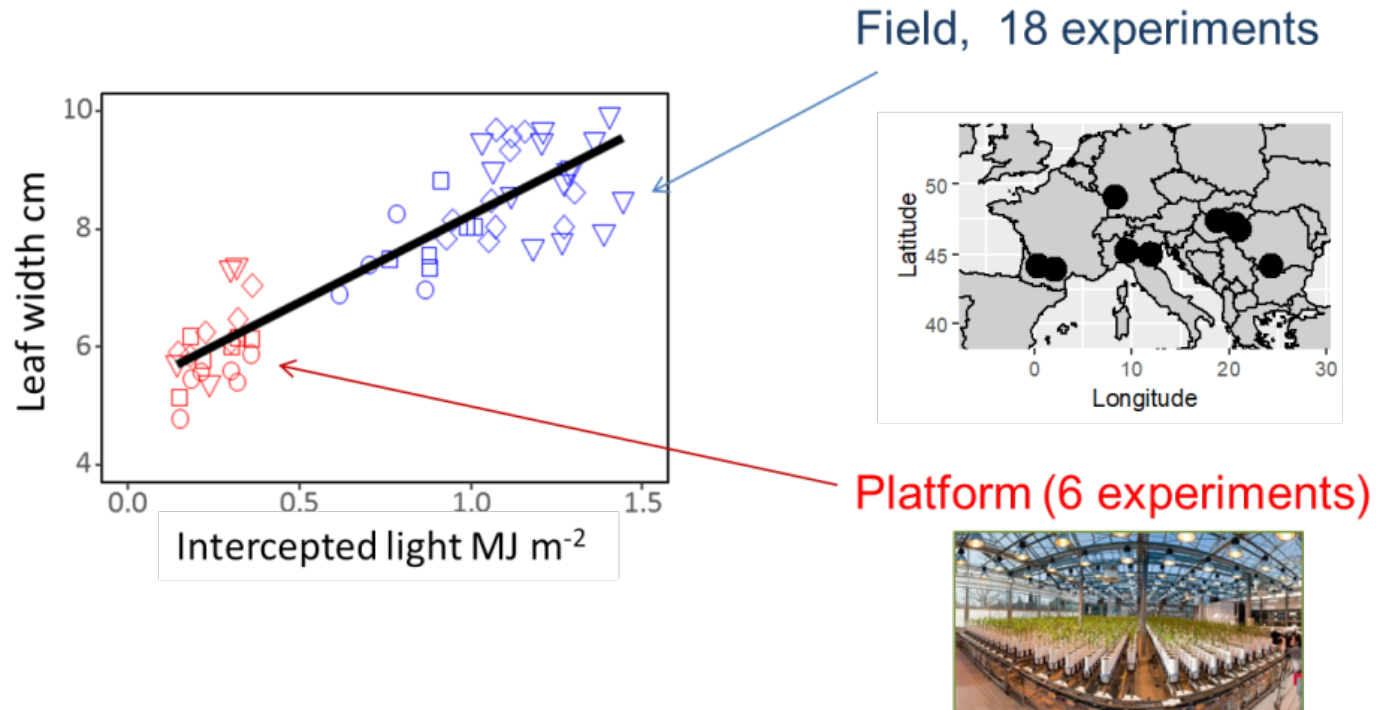


Knowledge Discovery Illustration

PHIS provides contextualisation: intercepted light value computation



Knowledge Discovery Illustration



A common relationship between leaf width and intercepted light per plant accounted for variations in width between fields, and for the difference between field and greenhouse



OpenSILEX

- ✓ Management of huge and complex data
- ✓ Enables and facilitates cloud computing (data center, EGI)
 - distributed computing, distributed storage, backup
- ✓ Free software and Open technologies
- ✓ International identification (URI and DOI)
- ✓ Semantic management (ontologies, standardized vocabularies)
- ✓ Provenance and reproducibility for data processing
- ✓ Flexible design
- ✓ R interface mainly based on OpenCPU package
- ✓ 10 instances for various installations (field and greenhouse)
 - Phenoarch instance → Over 300 Tb of data + 10 plant species
 - INRA, WUR, CIRAD, Univ of Tokyo, Nottingham
- ✓ MISTEA team and Start-up project for support and development

PHIS

Event annotation

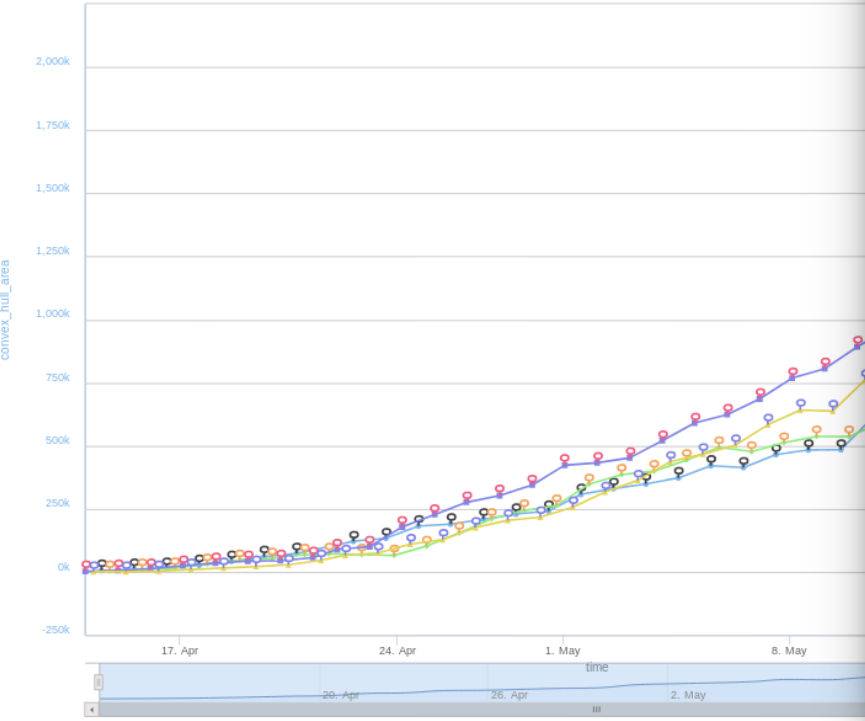
147.99.24.182/phis-dev/web/index.php?r=graphic%2Fvisu&experiment=http%3A%2F%2F 80% Rechercher

Phenotyping Hybrid Information System M3P

Experimental Organization Data Tools Pierre-Etienne Alary



Friday, Apr 28, 04:27:01
 ● 0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00



● 0010/DZ_PG_19/ZM4367/WD/Veg_1/01_10/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00 Images
 ● 0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00 Images
 ● 0063/DZ_PG_41/ZM4378/WW/Rep_1/02_03/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00 Images
 ● 0091/DZ_PG_18/ZM4373/WW/Rep_1/02_31/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00 Images

PHIS - Mozilla Firefox

147.99.24.182/phis-dev/views/graphic/commi 80%

Event or Expert Annotation

Author: pierre-etienne.alary@supagro.fr

IP: 10.146.2.250

Confidential: (oui)

Target (choose one):
 plant 0017/DZ_PG_20/ZM4344/WD/Veg_1/01_17/ARCH2017-03-30~Convex Hull Plant Area~side60~lepse: 243,378.00
 nearby image side60 2017-05-12T07:56:37+02:00

Datetime Event: 2017-05-12T07:45:07+02:00

Category:

Subject:

Content:

save

Projects Experiments Agronomical Objects Dataset Variables Tools Logout (morgane.vidal@inra.fr)



Use Alt+Shift+Drag to rotate the map. Use Ctrl+Click+Drag to select multiple elements.

Dataset(s) Visualization (On selected plot(s))

Leaf-Area-Index_LAI-Computation_m2.m2



Quantitative Variable
Leaf-Area-Index_LAI-Computation_m2.m2

Date Start
Enter date start

Date End
Enter date end

Search

Images Visualization (On selected plot(s))

Type
Hemisphericals

Show Images


Images



Trait – Images links

Home / Experiments

<http://www...>



Close

Use Alt+Shift+Drag to rotate the map. Use Ctrl+Click+Drag to select multiple elements.



Create Variable

Variable Label *

MyNewTrait_MyNewMethod_NA

Trait

Trait label



Internal Label

MyNewTrait

Comment

This is a comment for y new trait, on which my new variable is focused.

Method

Method label



Internal Label

MyNewMethod

Comment

This is a comment for my new method, used to produce the values of my new variable.

Unit

Unit label



Ontologies References

In order to fill ontological references (URI) you can go to these ontologies :

- [AGROPORAL ?](#)
- [AGROVOC ?](#)
- [PLANT ONTOLOGY ?](#)
- [PLANTEOME ?](#)
- [CROP ONTOLOGY ?](#)
- [UNIT ONTOLOGY ?](#)

Related References

Entity	Relation	Reference URI	Hyperlink
Variable	skos:closeMatch		<input type="text"/> +
Variable	skos:narrower		<input type="text"/> x
Trait	skos:exactMatch		<input type="text"/> x
Method	skos:exactMatch		<input type="text"/> x

OpenSILEX - PHIS

Data Analysis

Global Greenhouse Report

Name	Global Greenhouse Report
URI	http://www.phenome-fppn.fr/id/analysis/daglobal
Description	Visualization of a specified variable of an experiment. A HTML report is produced by this program.
Documents	

Experiment *

Trait *

View Label *

Data Analysis

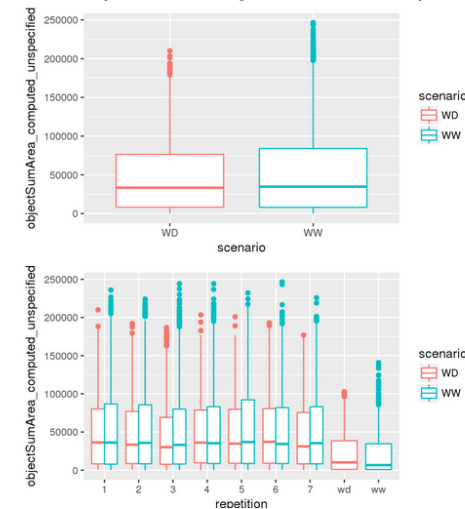
Showing 1-5 of 5 items.

#	Name ↓	URI	Description
1	Daily report Greenhouse	http://www.phenome-fppn.fr/id/analysis/dailyreportphis	Daily description of a PhenoArch experiment (imagery, environnement and so on...) to running of it. A HTML report is produced by this program.
2	Environment Report Field	http://www.phenome-fppn.fr/id/analysis/daenvirfield	Description of the environment of a field experiment (meteo...). A HTML report is produced by this program.
3	Environment Greenhouse Report Greenhouse	http://www.phenome-fppn.fr/id/analysis/daenvir	Description of environment of PhenoArch experiment (meteo...). A HTML report is produced by this program.
4	Global Report Greenhouse	http://www.phenome-fppn.fr/id/analysis/daglobal	Visualization of a specified variable of an experiment. A HTML report is produced by this program.
5	Thermal Calculation Report Time	http://www.phenome-fppn.fr/id/analysis/dathermal	For a PhenoArch experiment, a thermal time is calculated according to the user's chosen method. A HTML report and a csv file are produced by this program.

ARCH2017-03-30 - ZA17 experiment on objectSumArea parameter and side90 view

- 60 genotypes
- 2 scenarios: WD, WW
- 9 repetitions
- 1467 pots
- Scientist supervisors: Cabrera-Bosquet, Tardieu, Turc, Welcker
- Technical supervisors: Brichet, LUCHAIRE, Suard
- Experiment performed from 2017-03-30 to 2017-06-30
- Genotypes used in this experiment: IPG004, IPG007, IPG017, IPG026, IPG029, IPG062, IPG063, IPG066, IPG073, IPG077, IPG082, IPG089, IPG101, IPG103, IPG109, IPG110, IPG111, IPG116, IPG117, IPG119, IPG120, IPG121, IPG128, IPG131, IPG136, IPG138, IPG146, IPG148, IPG152, IPG153, IPG155, IPG158, IPG159, IPG164, IPG165, IPG167, IPG169, IPG173, IPG176, IPG181, IPG188, IPG189, IPG190, IPG194, IPG195, IPG202, IPG216, IPG228, IPG233, IPG234, IPG239, IPG303, IPG304, IPG310, IPG311, IPG312, IPG313, IPG314, IPG318, IPG321

Description of objectSumArea parameter



PHIS

Workflow management

Home / Currents Tasks / Clean plant height using default



Clean plant height using default

Workflow name	Clean plant height using default
Start	09-01-2018 13:53
End	09-01-2018 14:29

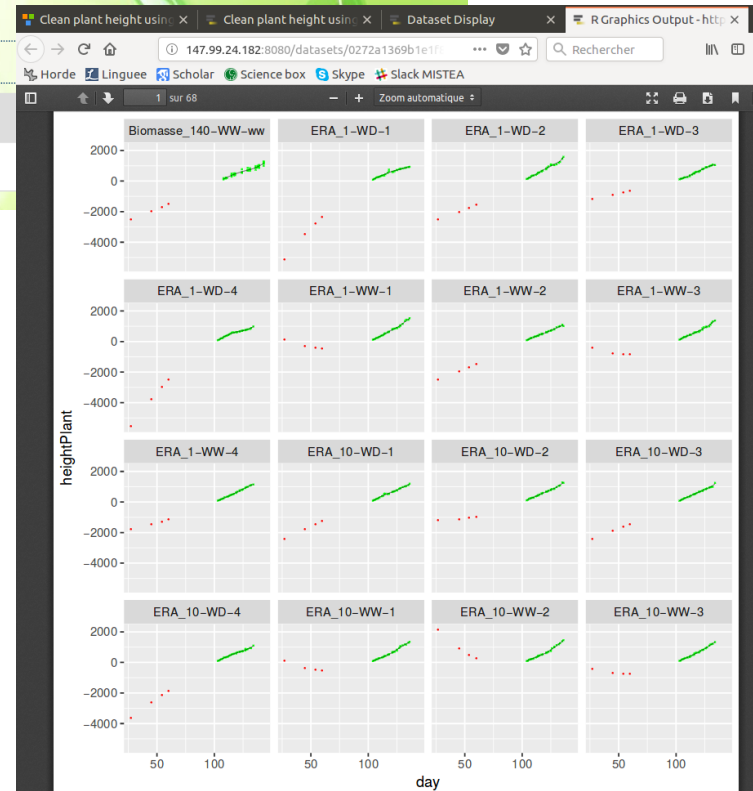
Open in Galaxy

Technical details

Invocation ID	40876639881ca029
History Id	6f91353f3eb0fa4a

The screenshot shows the Galaxy workflow interface. At the top, it displays the workflow name 'Clean plant height using default by Alamy' and the start time '13:53:00'. Below this, there are several data objects listed with their sizes and formats:

- 11. Input**: 2.8 MB, format: pdf, genome de reference: 2
- 10. Log from plotting**: 9.3 KB
- 8. Lines data**: ~100,000 lines, format: tabular, genome de reference: 2
- 7. Some data**: 8.1 KB
- 6. log**: 5.3 KB
- 5. Treated data**: 4.1 KB
- 4. log**: 3.1 KB
- 3. pipeline file**: 2.1 KB
- 2. 1st file**: 1.1 KB
- 1. request file**: 1.1 KB





OpenSILEX

➤ **PHIS** demonstration

- <http://phis.inra.fr/> Or <http://www.opensilex.org/opensilex/web/>
Research paper:
<https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15385>

➤ How to contribute to OpenSILEX?

- Github repository: <https://github.com/OpenSILEX/>
- Developer documentation: <https://opensilex.github.io/docs-community-dev/>

➤ User documentation of the version in development:

- <https://opensilex.github.io/phis-docs-community/>



Conclusion

OpenSILEX

- Provides description and structuring of data
- Ensures and makes easy FAIR data
- Recommends and implements standards,
- Provides a frame for data analysis and data publication