

## High-throughput data management and analysis in systems biology

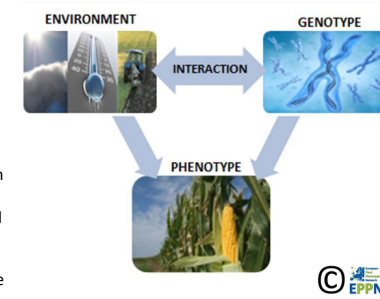
Filippo Geuna  
Università degli Studi di Milano



From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The complex and fascinating world of -omics analysis

Plant phenotyping is an emerging science that links genomics with plant ecophysiology and agronomy. The functional plant body (PHENOTYPE) is formed during plant growth and development from the dynamic interaction between the genetic background (GENOTYPE) and the physical world in which plants develop (ENVIRONMENT). These interactions determine plant performance and productivity measured as accumulated biomass and commercial yield and resource use efficiency.



<http://www.plant-phenotyping-network.eu>

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The complex and fascinating world of -omics analysis

Plant phenotyping has emerged as a comprehensive field of research as the result of significant advancements in the application of imaging sensors for high-throughput data collection. The flip side is the risk of drowning in the massive amounts of data generated by automated phenotyping systems.

Currently, the major challenge lies in data management, on the level of data annotation and proper metadata collection, and in progressing towards synergism across data collection and analyses. Progress in data analyses includes efforts towards the integration of phenotypic and -omics data resources for bridging the phenotype-genotype gap and obtaining in-depth insights into fundamental plant processes.

Imaging methodologies used in plant phenotyping generate huge amounts of complex data.

The major challenge is **data management**: metadata collection and data annotation.

Implementation of **standard ontologies** is key to integrate data efficiently. Data-driven approaches are promising to **generate new scientific insights**.

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI) <https://www.sciencedirect.com/science/article/pii/S2452310017300069>

## An overview of the talk

### The Genome level

- Creation of variants (recombinants and mutants)
- Genome sequencing (*de novo*)
- Transcriptomics (RNAseq analysis)
- Genome annotation
- Genome re-sequencing
- Variant analysis

### Genotype-Phenotype correlations

#### Data integration and mining

- Data curation and standardization
- Databases
- Statistical and biometrical software

### The Phenome level

- Transcriptomics (RNAseq analysis)
- Proteomics
- Ionomics
- Lipidomics
- Glycomics
- Imaging (hyperspectral)
- Environmental metadata

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

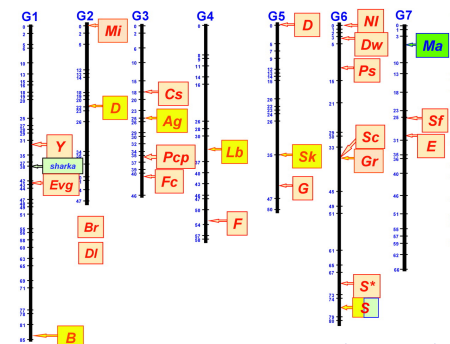
### Variability at the base of genetics



From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### Variability at the base of genetics

Approximate position of 28 major genes mapped in different populations of apricot (blue background), peach (orange background), almond or almond x peach (yellow background), and Myrobalan plum (green background) on the framework of the *Prunus* reference map.

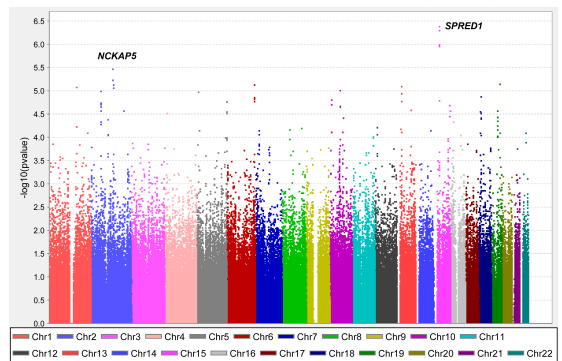


From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

Dirlewanger E. et.al. PNAS 2004; 101: 9891-9896

### Variability at the base of genetics

#### Genome-Wide Association Studies (GWAS)



[https://dfzjdn9uc3pi.cloudfront.net/2013/66/1/Supplementary%20figure%202\\_mod.jpg](https://dfzjdn9uc3pi.cloudfront.net/2013/66/1/Supplementary%20figure%202_mod.jpg)

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### The astonishing numbers of biological data

#### Encyclopedia Genomica: UK Scientists Print the Book of Life in 130 Volumes



Nowadays it is cheaper to sequence the human genome than to print out a full hard copy. We know this because Cas Kramer and a group of scientists at the University of Leicester decided to print out a complete version of the human genome.

The genome print out spans **130 volumes**, with each page printed on **both sides in 4-point font**, with precisely **43,000 characters per page**. The X chromosome is made up of seven volumes, while the Y chromosome occupies one. The exercise costs a little less than 4,000 pounds (about \$6,000).

University of Leicester staff with the human genome (hard copy). Cas Kramer and Ed Hollox, back row (right).

<http://www.bio-itworld.com/2012/12/28/encyclopedia-genomica-uk-scientists-print-book-of-life-in-130-volumes.html>

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)



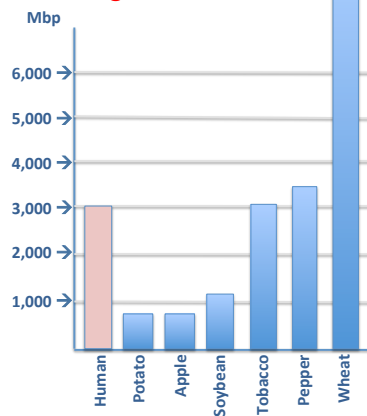
### The astonishing numbers of biological data

When size matters...

Beyond the problem of the many species to be studied (compared to the human genome), there is the problem with genome sizes.

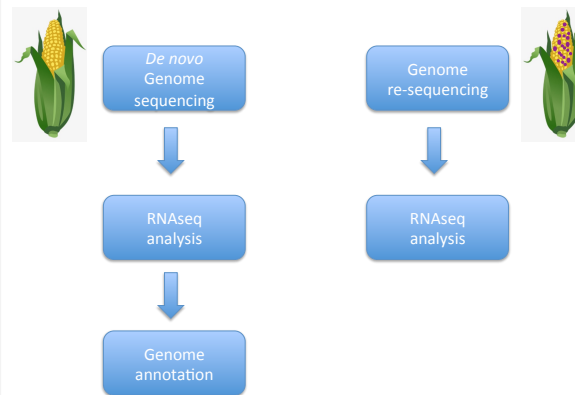
Tobacco genome is comparable to the human genome.

Wheat genome (out of scale in the picture) is nearly six fold the human genome.



From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

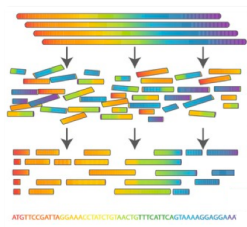
### The pipeline to assemble and annotate genomes



From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

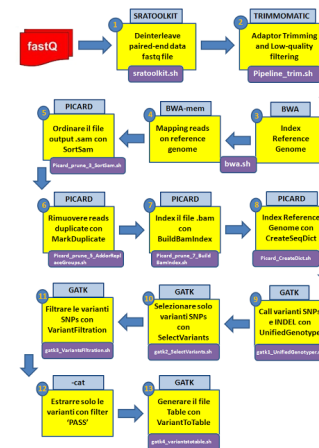
### Assembling genomes

...is like reconstructing a puzzle

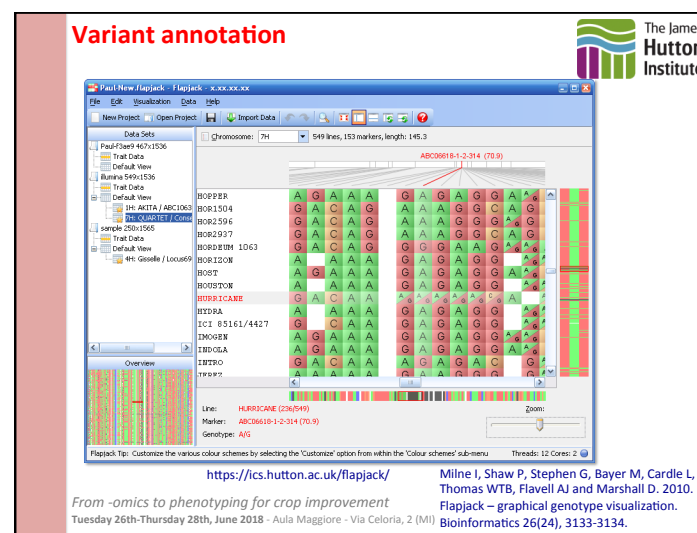
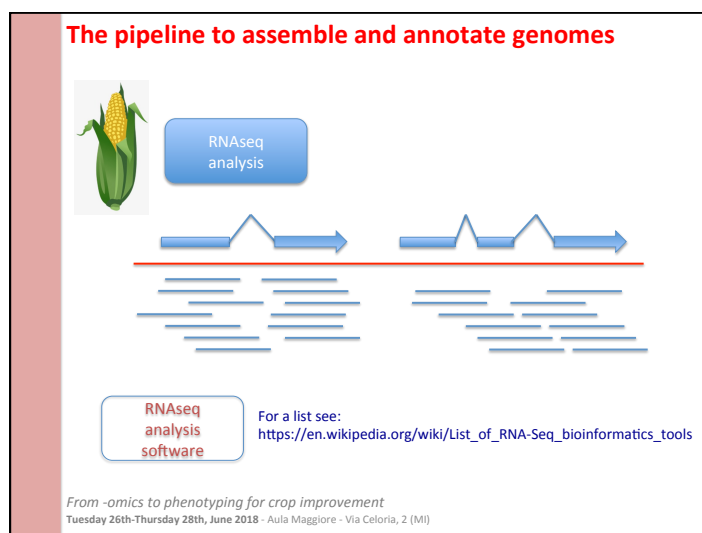
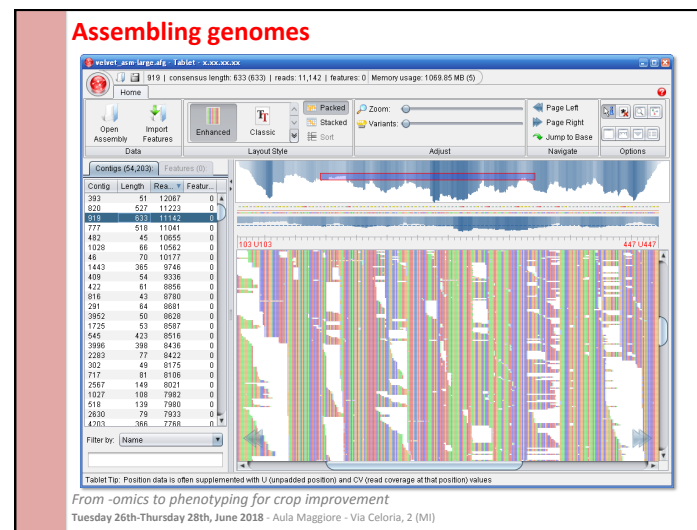
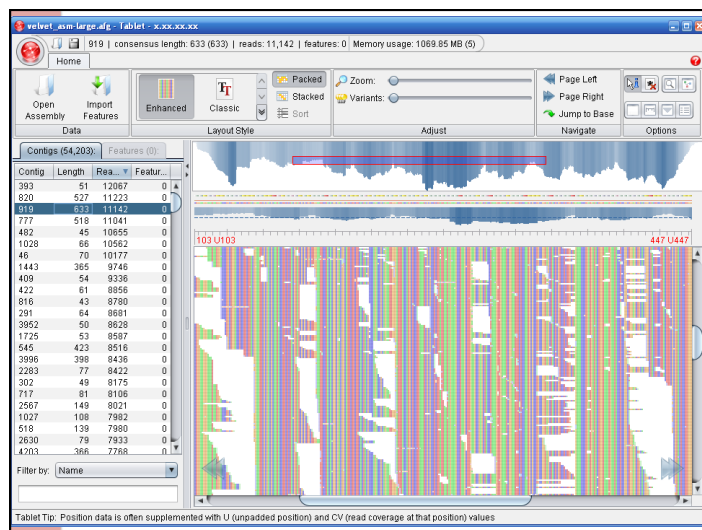


From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### Assembling genomes



From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)



## High-throughput phenotyping facilities

IPK Institute Gatersleben - Macrophenomics Facility and Microphenomics Facility

Currently, the plant phenotyping infrastructure at the IPK is being expanded by the new Plant Cultivation Hall. On an almost 500 m<sup>2</sup> area, plants will be cultivated under field-like precisely adjustable environmental conditions, without undesigned external influences. This will allow to repeat experiments under highly reproducible conditions. In addition to a container system, a so-called rhizotron-system will be installed in the Plant Cultivation Hall which will allow both phenotyping of shoot and root features.

<https://www.youtube.com/watch?v=psYu-xECDU>

With the Macrophenomics facility leaf damages caused by pathogens and visible to the naked eye, can automatically and precisely be quantified. A camera creating RGB images and taking pictures in four different wavelength ranges is applied to study the number and extent of fungal colonies per leaf area and the response of their host. Equipped with a multispectral 2D imaging station Macrobot (developed in cooperation with Fraunhofer IFF Magdeburg), up to 500 genotypes can be analyzed in a high-throughput process in 3-4 weeks.

<https://www.youtube.com/watch?v=8Xp0f2B7mRM>

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The 'Ontology' concept - BioPAX

computational  
BIOLOGY

PERSPECTIVE

### The BioPAX community standard for pathway data sharing

Emek Demir et al.

NATURE BIOTECHNOLOGY VOLUME 28 NUMBER 9 SEPTEMBER 2010

Biological Pathway Exchange (BioPAX) is a standard language to represent biological pathways at the molecular and cellular level and to facilitate the exchange of pathway data. The rapid growth of the volume of pathway data has spurred the development of databases and computational tools to aid interpretation; however, **use of these data is hampered by the current fragmentation of pathway information across many databases** with incompatible formats. BioPAX, which was created through a community process, solves this problem by making pathway data substantially easier to collect, index, interpret and share. BioPAX can represent metabolic and signaling pathways, molecular and genetic interactions and gene regulation networks. Using BioPAX, millions of interactions, organized into thousands of pathways, from many organisms are available from a growing number of databases. This large amount of pathway data in a computable form will support visualization, analysis and biological discovery.

From -omics to phenotyping for crop improvement <https://www.nature.com/articles/nbt.1666.pdf>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The 'Ontology' concept - BioPAX

computational  
BIOLOGY

PERSPECTIVE

NATURE BIOTECHNOLOGY VOLUME 28 NUMBER 9 SEPTEMBER 2010

### What is an ontology?

An ontology is a formal system for representing knowledge. Such representation is required for computer software to make use of information. Example ontologies include organism taxonomies and the Gene Ontology. A formal representation allows consistent communication of knowledge among individuals or computer systems and helps manage complexity in information processing as knowledge is broken down into clear concepts that can be considered independently.

Ontologies also enable integration of knowledge between independent resources linked on the World Wide Web. Such linked, structured data form the basis of the semantic web, an extension of the web that promises improved information management and search capability.

Representing and sharing knowledge using ontologies is simplified by availability of the standard **web ontology language (OWL; <http://www.w3.org/TR/owl-features/>)**. Tools to edit OWL, such as Protégé, have been developed by the semantic web community and adopted in the life sciences. Implementing BioPAX using OWL enables both the ontology and the individuals and values to be stored in the same XML-based format, which makes data transmission easier. An ontology is composed of classes, properties (representing relations) and restrictions and is used to define individuals (instances of classes, also known as objects) and values for their properties. Classes (also known as concepts or types) are often arranged into a hierarchy (or taxonomy) where child classes are more specific than, and inherit the properties of, parent classes.

From -omics to phenotyping for crop improvement <https://www.nature.com/articles/nbt.1666.pdf>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

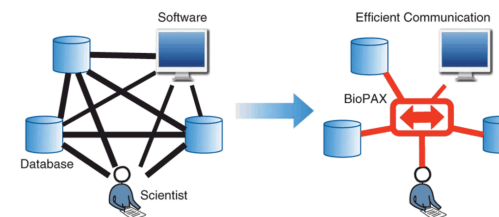
## The 'Ontology' concept - BioPAX

MENU nature  
biotechnology

Search Edit Link Log In

Figure 1: BioPAX is a shared language for biological pathways.

From: The BioPAX community standard for pathway data sharing

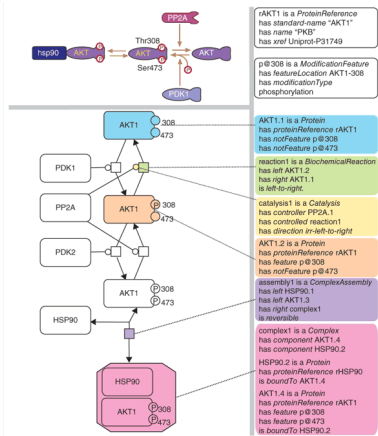


BioPAX reduces the effort required to efficiently communicate between pathway users, databases and software tools. Without a shared language, each system must speak the language of all other systems in the worst case (black lines). With a shared language, each system only needs to speak that language (central red box).

From -omics to phenotyping for crop improvement <https://www.nature.com/articles/nbt.1666/figures/1>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### The 'Ontology' concept - BioPAX

Figure 3: The AKT pathway as represented by a traditional method (top left; from <http://www.biocarta.com/>), a formalized SBGN diagram (left; from <http://www.sbn.org/62>) and using the BioPAX language (right).



An important advantage of the BioPAX representation is that it can be interpreted by computer software and used in multiple ways, including automatic diagram creation, information retrieval and analysis. Online documentation at <http://www.biopax.org/> contains more details about how to represent diverse types of biological pathways.

From *-omics to phenotyping for crop improvement* <https://www.nature.com/articles/nbt.1666/figures/3>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### The 'Ontology' concept - BioPAX

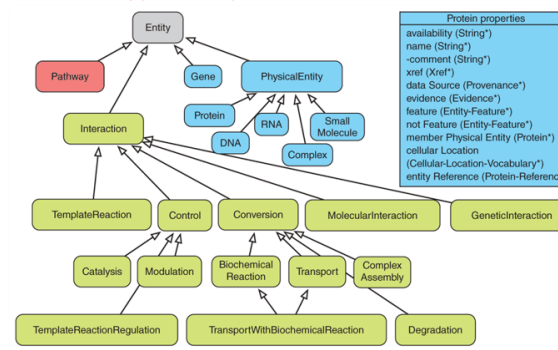
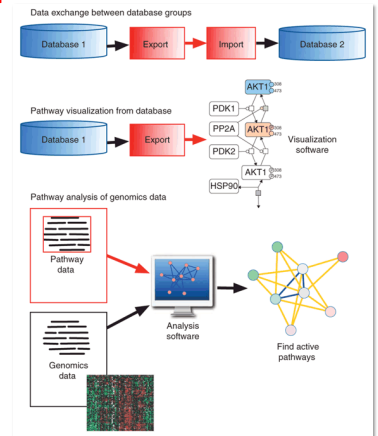


Figure 4: High-level view of the BioPAX ontology. Classes, shown as boxes and arrows, represent inheritance relationships. The three main types of classes in BioPAX are Pathway (red), Interaction (green) and PhysicalEntity and Gene (blue). For brevity, the properties of the Protein class only are shown as an example at the top right. Asterisks indicate that multiple values for the property are allowed.

From *-omics to phenotyping for crop improvement* <https://www.nature.com/articles/nbt.1666/figures/4>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### The 'Ontology' concept - BioPAX

Figure 5: Example uses of pathway information in BioPAX format. Red-colored boxes or lines indicate the use of BioPAX.



From *-omics to phenotyping for crop improvement* <https://www.nature.com/articles/nbt.1666/figures/5>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### The 'Ontology' concept - BioPAX

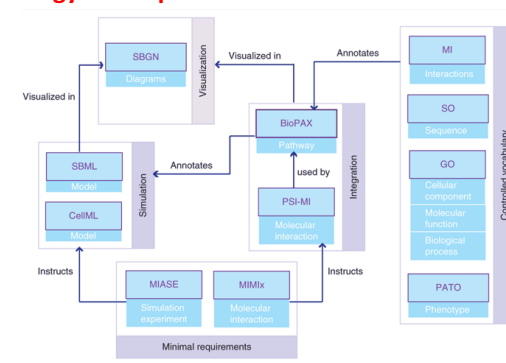


Figure 6: The relationship among popular standard formats for pathway information. BioPAX and PSI-MI are designed for data exchange to and from databases and pathway and network data integration. SBML and CellML are designed to support mathematical simulations of biological systems and SBGN represents pathway diagrams.

From *-omics to phenotyping for crop improvement* <https://www.nature.com/articles/nbt.1666/figures/6>  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)



## The 'Ontology' concept - Pathguide

Pathguide the pathway resource list

Complete Listing of All Pathguide Resources

Pathguide contains information about 782 biological pathway related resources and molecular interaction related resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BiOPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

Protein-Protein Interactions

Database Name (Order alphabetically) (by web popularity)	Full Record	Availability	Standards
292ids - The Protein-Protein Interaction Inhibition Database	Details	Free	
3D-Interologs - 3D-Interologs	Details	Free	
3DID - 3D interacting domains	Details	Free	
ACSN - Atlas of Cancer Signalling Network	Details	Free	BiOPAX
ADAN - Prediction of protein-protein interaction of modular domains	Details	X	
AHDZD - Arabidopsis Hormone Database 2.0	Details	Free	
AIFuse - Functional Associations of Proteins in Complete Genomes	Details	X	
iMAZE - Protein Function and Biochemical Pathways Project	Details	X	
ANAP - Arabidopsis Network Analysis Pipeline	Details	Free	
ANA - Association and Integrated Analysis of the 14-3-3 Interactome	Details	Free	
AnimalTFDB - Animal Transcription Factor Database	Details	Free	
AntiJen - AntiJen a Kinetic, Thermodynamic and Cellular Database	Details	Free	

<http://www.pathguide.org>

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celeria, 2 (MI)

## The Systems Biology concept - SBGN

SBGN

About

- Get involved
- SBGN Editors
- SBGN Acting Editors
- SBGN Scientific Committee
- Funding
- Ownership

Learn to Use SBGN

Symbol Highlights

Published Maps

Examples

Templates

Contribute

Competition

Get involved

Regardless of your experience or involvement level, we value your ideas and feedback. For reporting issues, suggesting enhancements or requesting features please use the dedicated [trackers on our code repositories](#). For general discussions about SBGN; please join the [sbgn-discuss mailing list](#) and bring up the topic there. For specific questions about this website, workshops, or other SBGN resource management topics: please send a message to the [sbgn-editors mailing list](#). SBGN meetings and workshops provide the best environment for learning about the process and getting involved. Please check the [Events page](#) for the upcoming events.

<http://sbgn.github.io/sbgn/>

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celeria, 2 (MI)

## The Plant Ontology initiative

The Plant Ontology (PO; <http://www.plantontology.org/>) is a publicly available, collaborative effort to develop and maintain a controlled, structured vocabulary ('ontology') of terms to describe plant anatomy, morphology and the stages of plant development. The goals of the PO are to link (annotate) **gene expression** and **phenotype data** to plant structures and stages of plant development, using the data model adopted by the Gene Ontology.

From its original design covering only rice, maize and Arabidopsis, the scope of the PO has been expanded to include all green plants. The PO was the first multispecies anatomy ontology developed for the annotation of genes and phenotypes. Also, to our knowledge, it was one of the first biological ontologies that provides translations (via synonyms) in non-English languages such as Japanese and Spanish.

As of Release #18, there are about 2.2 million annotations linking PO terms to >110,000 unique data objects representing genes or gene models, proteins, RNAs, germplasm and quantitative trait loci (QTLs) from 22 plant species. The focus is on the plant anatomical entity branch of the PO, describing the organizing principles, resources available to users and examples of how the PO is integrated into other plant genomics databases and web portals.

Two examples of comparative analyses, demonstrate how the ontology structure and PO-annotated data can be used to discover the patterns of expression of the LEAFY (LFY) and terpene synthase (TPS) gene homologs.

Source: Cooper L, Walls RL. 2013. The Plant Ontology as a Tool for Comparative Plant Anatomy and Genomic Analyses. *Plant & Cell Physiology*. 54:1-23

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celeria, 2 (MI)

## The Crop Ontology initiative

Crop Ontology Community Website

Search this site

About

Crop and Agronomy Ontology Community >

2018 PhenHarmoniS

Semantics for Harmonization and Integration of Phenotypic and Agronomic Data  
14-18 May 2018 - Agropolis Scientific Park, Montpellier, France

PhenoHarmoniS workshop

Workshop of the CoP

2018 PhenHarmoniS

Agro

Talks Summaries

Presentations Files

Reports of the sessions

Sessions descriptions

Speakers

View of CROD High Tech Platforms

Outing & Discoope Visit

Participants

Programme & Logistics

Committees for PhenHarmoniS2018

Itinerary to the venue

Recommended Hotels

2018 Workshop

PhenoHarmoniS

Video of the presentations

Presentations (slides)

Tools

Posters

Agenda

Sessions descriptions

From -omics to phenotyping for crop improvement  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celeria, 2 (MI)

<https://sites.google.com/a/cxchange.org/cropontologycommunity/home/2018-phenharmonis>

## The Crop Ontology initiative

Platform for Big Data in Agriculture

About Communities of Practice Events News Opportunities Press

### Welcome

To the Ontologies Data Community of Practice

This Community of Practice (CoP) is led by Bioversity International and has been launched as a part of the CGIAR Platform for Big Data in Agriculture.

This CoP on Ontology will span the knowledge domains that are relevant to the Platform with the objective of establishing best practices, and guidelines in the selection.

It will also focus on the use and application of semantics for data harmonization at the levels of collection and storage, and for data interoperability and data discovery following the FAIR principles.

This space can be used as a discussion area, share and request relevant information and contribute towards building the community as a whole.

### Community News

<https://bigdata.cgiar.org/ontologies/>

*From -omics to phenotyping for crop improvement*  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The Crop Ontology initiative

Given the Multi-disciplinary nature of the CGIAR Big Data For Agriculture, the Ontology Community of Practice (CoP) will span the knowledge domains of importance for the platform with the objective of establishing best practices, guidelines in the selection, **use and application of semantics** for data harmonization at the collect and storage levels, for data interoperability and data discovery following the FAIR principles. The CoP will contribute to make the best practices and recommended ontologies a long-term reference resource for the CGIAR Big Data in Agriculture in particular and for other platforms. The INSPIRE projects will offer opportunities to test the power of the ontologies and best practices for combining data sets to address a research question.

**Figure 1: semantics are needed from the data collection to the data access**

*From -omics to phenotyping for crop improvement*  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

[https://bigdata.cgiar.org/wp-content/uploads/2017/05/CGIAR-Ontology-CoP\\_Workplan-2017.pdf](https://bigdata.cgiar.org/wp-content/uploads/2017/05/CGIAR-Ontology-CoP_Workplan-2017.pdf)

## The Crop Ontology initiative

**Figure 2: In colour, knowledge domains addressed by the Big Data platform. In grey, knowledge domains addressed by CGIAR Excellence in Breeding and Genebank platforms.**

*From -omics to phenotyping for crop improvement*  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

[https://bigdata.cgiar.org/wp-content/uploads/2017/05/CGIAR-Ontology-CoP\\_Workplan-2017.pdf](https://bigdata.cgiar.org/wp-content/uploads/2017/05/CGIAR-Ontology-CoP_Workplan-2017.pdf)

## The Crop Ontology initiative

### Crop Ontology Curation Tool

Home About Feedback

CO\_322 [Add New Terms](#) [API](#) [Help](#) [Agrtrials](#) [Annotation Tool](#) [Register](#) [Login](#)

[http://www.croponontology.org/ontology/CO\\_322/Maize#](http://www.croponontology.org/ontology/CO_322/Maize#)

*From -omics to phenotyping for crop improvement*  
Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## The Crop Ontology initiative

The screenshot shows the Crop Ontology interface. On the left, a tree view displays categories like 'Abiotic stress traits', 'Agronomical traits', and 'Leaf rolling severity'. The right panel provides details for the selected term 'Leaf rolling severity', including its identifier (CO\_322:0000160), trait description, attribute (rolling severity), entity (Leaf), main trait abbreviation (LRollSev), and trait class (Abiotic stress traits).

[http://www.croponontology.org/ontology/CO\\_322/Maize#](http://www.croponontology.org/ontology/CO_322/Maize#)

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (M)

## The OBO Foundry initiative

The image shows the cover of a 'PERSPECTIVE' article in Nature Biotechnology. The title is 'The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration'. The authors listed are Barry Smith, Michael Ashburner, Cornelius Rose, Jonathan Bard, William Bugé, Werner Custer, Louis Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, the OBI Consortium, Nicolas Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna Assunta Sansone, Richard H Schoenemann, Nigam Shah, Patricia L Whetzel, and Susanna Lewis.

The value of any kind of data is greatly enhanced when it exists in a form that **allows it to be integrated with other data**. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates obstacles to integration. The Open Biomedical Ontologies (OBO) consortium is pursuing a strategy to overcome this problem. Existing OBO ontologies, including the Gene Ontology, are undergoing coordinated reform, and new ontologies are being created on the basis of an evolving set of shared principles governing ontology development. The result is an expanding family of ontologies designed to be interoperable and logically well formed and to incorporate accurate representations of biological reality.

Source: <https://www.nature.com/articles/nbt1346.pdf>

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (M)

## The OBO Foundry initiative

Ontology	URL	Custodians
<b>Table 2 OBO Foundry ontologies (as of April 2007)</b>		
<b>Ontology Scope</b>		
<b>Mature ontologies undergoing incremental reform</b>		
Cell Ontology (CL)	<a href="http://obofoundry.org/cgi-bin/detail.cgi?cell">http://obofoundry.org/cgi-bin/detail.cgi?cell</a>	Michael Ashburner, Jonathan Bard, Oliver Hofmann, Sue Rhee
Gene Ontology (GO)	<a href="http://www.geneontology.org">http://www.geneontology.org</a>	Gene Ontology Consortium
Foundational Model of Anatomy (FMA)	<a href="http://fma.biostr.washington.edu">http://fma.biostr.washington.edu</a>	J.L.V. Mejino, Jr., Cornelius Rose
Zebrafish Anatomical Ontology (ZAO)	<a href="http://zfin.org/zfin_info/anatomy/dictum.html">http://zfin.org/zfin_info/anatomy/dictum.html</a>	Melissa Haendel, Monte Westerfield
<b>Mature ontologies still in need of thorough review</b>		
Chemical Entities of Biological Interest (ChEBI)	<a href="http://www.ebi.ac.uk/chebi">http://www.ebi.ac.uk/chebi</a>	Paula Dematos, Rafael Alcántara
Disease Ontology (DO)	<a href="http://diseasesontology.sf.net">http://diseasesontology.sf.net</a>	Rex Chisholm
Plant Ontology (PO)	<a href="http://plantontology.org">http://plantontology.org</a>	Plant Ontology Consortium
Sequence Ontology (SO)	<a href="http://www.sequenceontology.org">http://www.sequenceontology.org</a>	Karen Eilbeck
<b>Ontologies for which early versions exist</b>		
Ontology for Clinical Investigations (OCI)	<a href="http://www.biontology.org/wiki/index.php/OCI:Main_Page">http://www.biontology.org/wiki/index.php/OCI:Main_Page</a>	OCI Working Group
Common Anatomy Reference Ontology (CARO)	<a href="http://obofoundry.org/cgi-bin/detail.cgi?caro">http://obofoundry.org/cgi-bin/detail.cgi?caro</a>	Fabian Neuhaus, Melissa Haendel, David Suberstad
Environment Ontology	<a href="http://www.obofoundry.org/cgi-bin/detail.cgi?envo">http://www.obofoundry.org/cgi-bin/detail.cgi?envo</a>	Norman Morrison, Dawn Field
Ontology for Biomedical Investigations (OBI)	<a href="http://obi.tl.net">http://obi.tl.net</a>	OBI Working Group
Phenotypic Quality Ontology (PQO)	<a href="http://www.phenotypeontology.org">http://www.phenotypeontology.org</a>	Michael Ashburner, Susanna Lewis, Georgios Koutrouf
Protein Ontology (PRO)	<a href="http://pro.garagetown.edu/pro">http://pro.garagetown.edu/pro</a>	Protein Ontology Consortium
Relation Ontology (RO)	<a href="http://obofoundry.org/pro">http://obofoundry.org/pro</a>	Barry Smith, Chris Mungall
RNA Ontology (RNO)	<a href="http://rnc.bgu.edu/">http://rnc.bgu.edu/</a>	RNA Ontology Consortium

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (M)

Source: <https://www.nature.com/articles/nbt1346.pdf>

## The potential of plant phenotyping data

The image shows the cover of an article in ScienceDirect, Current Opinion in Systems Biology. The title is 'Unlocking the potential of plant phenotyping data through integration and data-driven approaches'. The authors are Frederik Coppens, Nathalie Wuyts, Dirk Inzé, and Stijn Dhondt.

### Data management to enable data integration

The current technologies and methods used in plant phenotyping generate a huge amount of complex, unstructured "Big Data", which can give the impression that a lot of the phenotype data might not be retrieved anymore. In first instance, phenotypic data management requires the use of ontology terms for the unique and repeatable annotation of data in order to ensure their persistence in view of traceability and reuse under the form of data sharing and meta-analyses.

The use of ontologies therefore promotes synergism. Moreover, in contrast to repositories such as the European Nucleotide Archive (ENA) or Sequence ReadArchive (SRA) for sequencing data, there is currently no central, structured repository for phenotyping data or metadata. Although data can be uploaded to general purpose repositories such as Zenodo (<https://zenodo.org/>), FigShare (<https://figshare.com>) and Dryad (<http://datadryad.org>), these do not provide services to facilitate the description of, access to and integration of data.

As a consequence of the lack of a central repository, advanced data mining and discovery depends on the error-prone scavenging of scientific literature.

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (M)

<https://www.sciencedirect.com/science/article/pii/S2452310017300069>

## The potential of plant phenotyping data



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)  
ScienceDirect

Current Opinion in  
Systems Biology

Unlocking the potential of plant phenotyping data through integration and data-driven approaches  
Frederik Coppens<sup>1,2</sup>, Nathalie Wuylts<sup>1,2</sup>, Dirk Inzé<sup>1,2</sup> and Stijn Dhondt<sup>1,2</sup>

As a consequence, a plethora of resources has been developed by individual research groups and consortia, ranging from resources dedicated to one species or one type of phenotyping system to more generic platforms allowing the integration of several data types. **AraPheno** provides a central repository of population-scale phenotypes for Arabidopsis accessions, whereas the **Plant Genomics and Phenomics (PGP)** research data repository is an infrastructure to comprehensively publish plant research data covering cross-domain datasets.

The **Phenomics Ontology Driven Data (PODD)** repository was developed to handle and distribute phenotyping data and metadata from Australian facilities. **ClearedLeaves DB** functions as an online database of cleared plant leaf images. **Phenopsis DB** is an information system for sharing data generated by the PHENOPSIS plant phenotyping platform and **PhenoFront** is a web-server front end to the **LemnaTec Phenotyper** platform. Whereas **BreeDB** hosts data sets of tomato and potato populations (<https://www.eu-sol.wur.nl>), **Genoplante Information System (GnplIS)** is a multispecies integrative information system dedicated to plant and fungi pests, bridging genetic and genomic data.

Many of these data resources have been built to organize a huge amount of collected phenotypic data. In the light of high-throughput phenotyping, there is a need for managing the data at the moment it is being generated.

*From -omics to phenotyping for crop improvement*

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

<https://www.sciencedirect.com/science/article/pii/S2452310017300069>

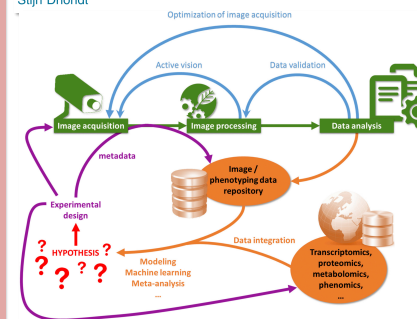
## The potential of plant phenotyping data



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)  
ScienceDirect

Current Opinion in  
Systems Biology

Unlocking the potential of plant phenotyping data through integration and data-driven approaches  
Frederik Coppens<sup>1,2</sup>, Nathalie Wuylts<sup>1,2</sup>, Dirk Inzé<sup>1,2</sup> and Stijn Dhondt<sup>1,2</sup>



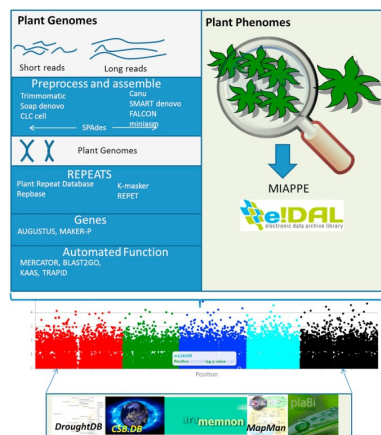
*From -omics to phenotyping for crop improvement*

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

<https://www.sciencedirect.com/science/article/pii/S2452310017300069>

**A systems biology approach in phenotypic data management.** A scientific hypothesis leads to new experiments including image-based plant phenotyping or other -omics approaches. Active vision systems can directly feedback into the image acquisition. Image acquisition features like the spatial and temporal resolution can also be optimized after data analysis. Sanity checks on the generated data help to quickly validate the image analysis. The analyzed data and images are saved along with the metadata and the experimental design in a dedicated data repository. Additional value is created by the integration of -omics data coming from private or public data resources, after which new hypotheses are generated through data-driven approaches like modeling, machine learning and meta-analysis.

## A world of databases and... a galaxy of data



*From -omics to phenotyping for crop improvement*

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## A world of databases and... a galaxy of data

The use of **advanced imaging systems** has **drastically increased the volume of data** from a couple of bytes, e.g. manually scored traits in a spreadsheet, to several megabytes (MB) or sometimes more than 100 MB, e.g. in the case of hyperspectral imaging or scene characterization by means of video capture.

**Data are also stored in a myriad of formats** on diverse types of media ranging from a researcher's hard drive to local server stations or in "the cloud". **Proper annotation of data** to ensure their continued relevance after acquisition is thus essential.

Furthermore, because the plant's phenotype is the result of a strong interaction between its genotype and the environment in which it grows ( $G \times E$ ), plant phenotyping efforts should **include the logging of environmental conditions**, which in turn requires the collection of metadata on the sensors in use.

Because of the tremendous amounts and diversity of data produced within the plant phenotyping research field, data management, storage and analysis are currently considered as the major challenges. On the other hand, large datasets may also create opportunities for data modeling and machine learning towards "**Big Data**" analyses.

*From -omics to phenotyping for crop improvement*

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

<https://www.sciencedirect.com/science/article/pii/S2452310017300069>



### Open Science

Source: Trends in Plant Science 2018 23:451-466

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### ELIXIR - distributed infrastructure for biological data

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. [www.elixir-europe.org](http://www.elixir-europe.org)

Source: <https://www.youtube.com/watch?v=stTY6kwanY>

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### Open Science: the power of collaboration

**Forum**  
 The Future of Field Trials in Europe: Establishing a Network Beyond Boundaries

Hartmut Stützel,<sup>1,\*</sup>  
 Nicolas Brüggemann,<sup>2</sup> and  
 Dirk Inzé<sup>3,4</sup>

We propose the establishment of a European Consortium for Open Field Experimentation (ECOFE) that will allow easy access of European plant and soil scientists to experimental field stations that cover all major climatological regions. Coordination and quality control of data extraction and management systems will greatly impact on our ability to cope with grand challenges such as climate change and food security.

Source: Trends in Plant Science, February 2016, Vol. 21, No. 2

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

### Open Science: the power of collaboration

**Forum**  
 The Future of Field Trials in Europe: Establishing a Network Beyond Boundaries

Hartmut Stützel,<sup>1,\*</sup>  
 Nicolas Brüggemann,<sup>2</sup> and  
 Dirk Inzé<sup>3,4</sup>

[...] A closer look at the research topics investigated at field trial sites demonstrates their **irreplaceability**: questions related to crop productivity and quality, climate change effects on crops, nutrient fluxes in agro-ecosystems, resource efficiency, stress mitigation, or the properties of resilient cropping systems cannot be investigated in test tubes in the laboratory. [...]

Source: Trends in Plant Science, February 2016, Vol. 21, No. 2

From -omics to phenotyping for crop improvement  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## Open Science: the power of collaboration

CellPress

Forum  
The Future of Field Trials in Europe: Establishing a Network Beyond Boundaries

Hartmut Stützel,<sup>1,\*</sup>  
Nicolas Brüggemann,<sup>2</sup> and  
Dirk Inzé<sup>3,4</sup>

[...] Usually they imply the interactions between genotype (G), environment (E), and management (M), in short: the G E M interactions. Inevitably, the investigation of G x E x M interactions requires, in addition to genotypic variation, ranges of environmental factors or gradients, and variation in agronomic management. In view of a growing world population, global climate change, and increasing strictness of environmental policies, we can expect that the above-mentioned themes will gain importance in agricultural and plant research, and it is imperative to investigate whether we have the appropriate infrastructure to meet these challenges..

[...] future research in plant and agro-ecological science **will increasingly depend on large-scale and long-term data** obtained from scientific experiments under real-world conditions. [...]

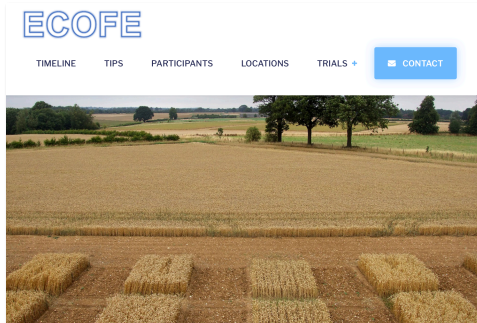
Source: Trends in Plant Science, February 2016, Vol. 21, No. 2

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## Open Science: the power of collaboration

**ECOFE – A research network to meet the needs for 21st century crop science**  
Support is sought for development and implementation of harmonized and standardized methodologies and to expand the infrastructure network in order to enhance the effectiveness and competitiveness of European crop science.



Source: <https://www.ecofe.eu>

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## Open Science

[...] Infrastructures for field trials including **data management** are a necessary prerequisite for research in plant science and agro-ecology. Setting up a common organization across Europe through an **intelligent network** of existing structures, simultaneously creating a best practice and quality control system, together with an accessible data repository, would be a major step forward in fostering a truly interdisciplinary European research arena to meet the challenges of the next decades towards food production, bio-economy, and sustainability. [...]

Source: Trends in Plant Science 2018 23, 451-466

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

## Open Science

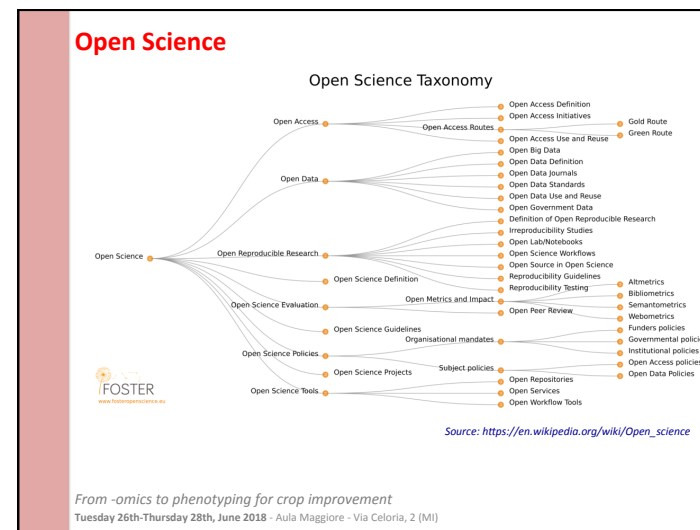
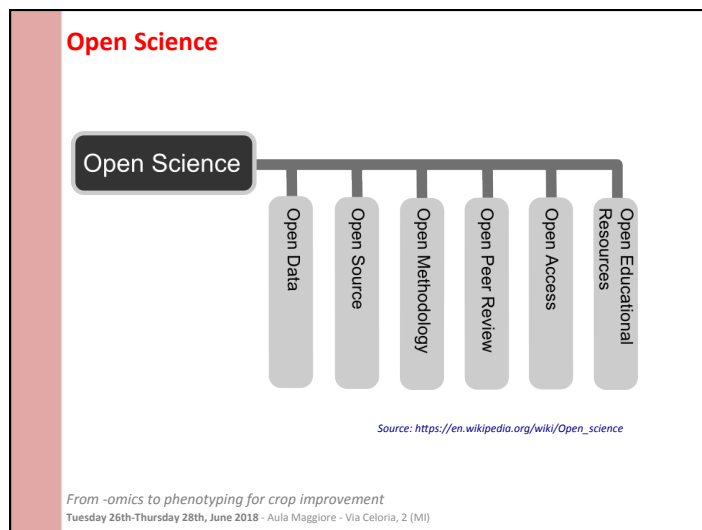
"The open sharing of research data is not widely practiced"



Source: [https://en.wikipedia.org/wiki/Open\\_science](https://en.wikipedia.org/wiki/Open_science)

From -omics to phenotyping for crop improvement

Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)



### Data consistency (and integrity...)

IS YOUR DATA ALL OVER THE PLACE?

- 50% time spent on DATA ENTRY
- 35 h annually spent on DATA SEARCHING
- 17% annually DATA LOST

**SCINOTE**  
 OPEN SOURCE ELECTRONIC LAB NOTEBOOK

*From -omics to phenotyping for crop improvement*  
 Tuesday 26th-Thursday 28th, June 2018 - Aula Maggiore - Via Celoria, 2 (MI)

