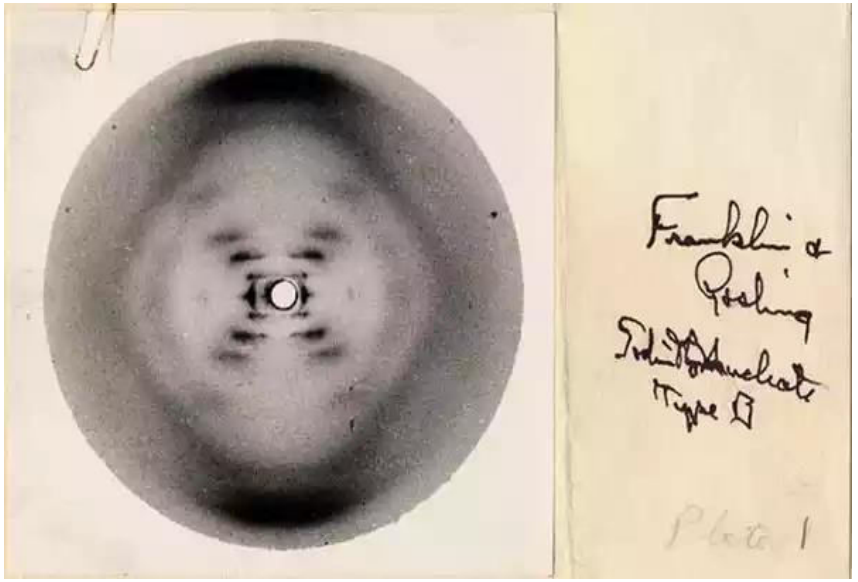
International Potato Center

Sweetpotato ontology status

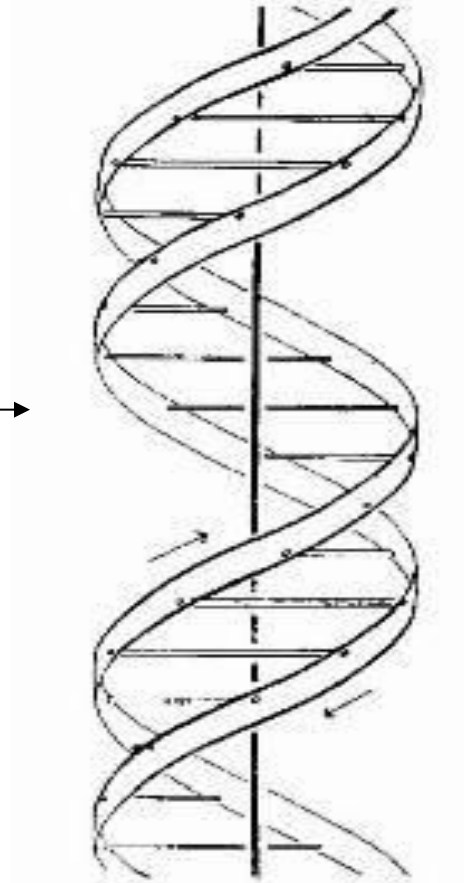
Reinhard Simon, CIP

GT4SP/SASHA workshop, Nairobi, Kenya, 6-9th of June, 2016


BIG DATA
— GOT IT...
NOW WHAT?

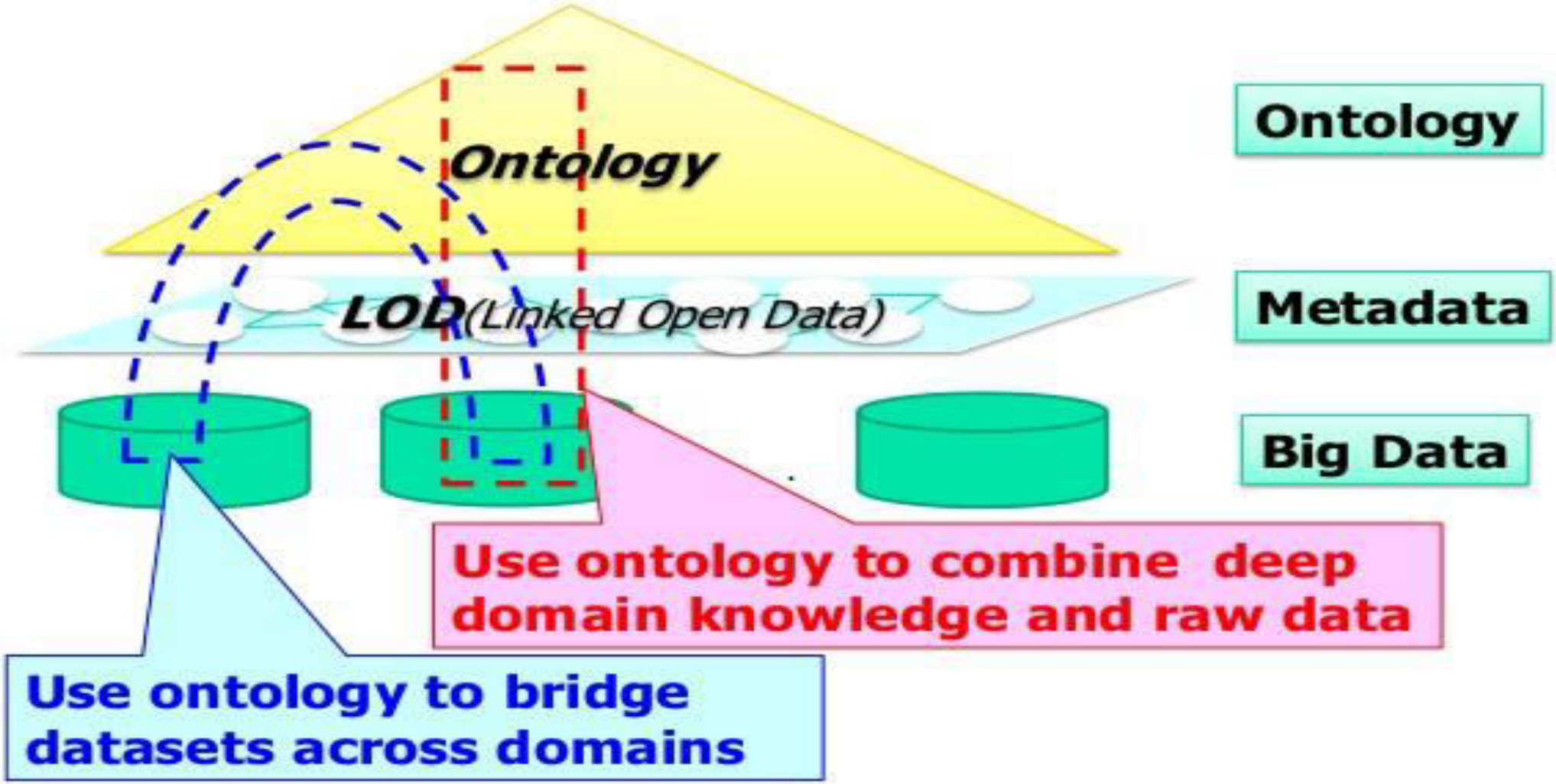


A case of big data …
The DNA model



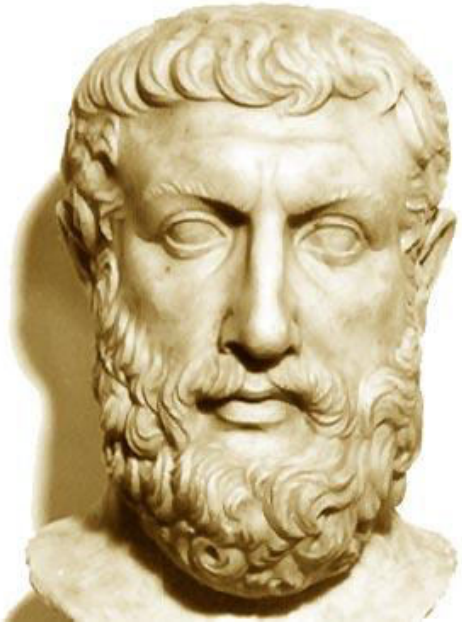
This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

Two possible way to use ontology for big data



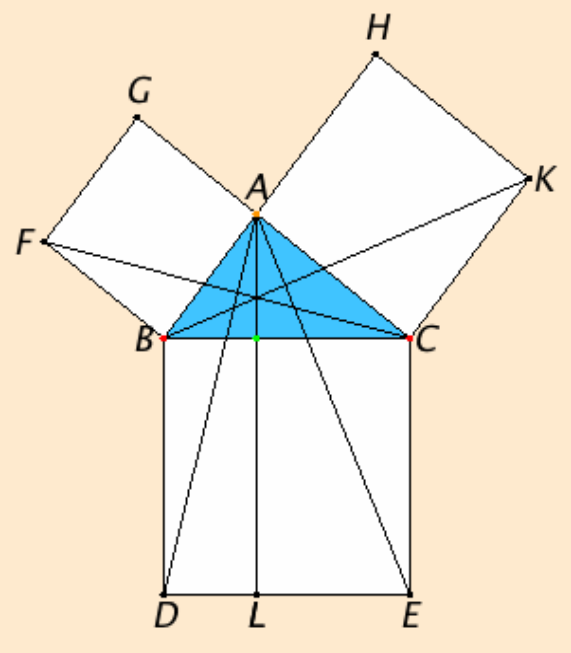
What is ‘ontology’?

- From Greek:
  - 'ontos' ~ the authentic essence of being / reality
  - 'logos' ~ word / reasoning
- Ontology: 'Reasoning about reality' (as opposed to mere opinions)
  - Formal system of knowledge organization
  - Definition of terms and relationships

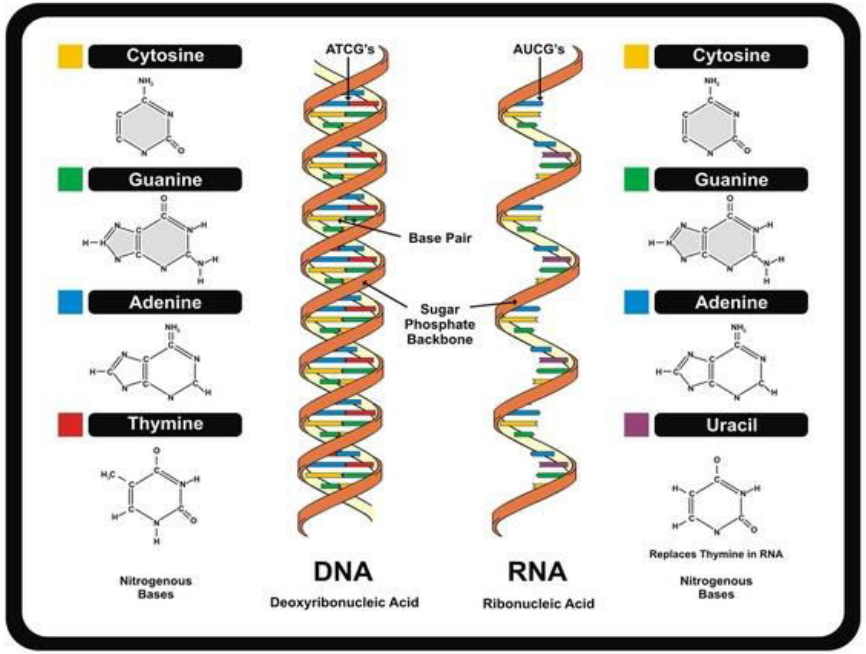


Parmenides: ~500 BC

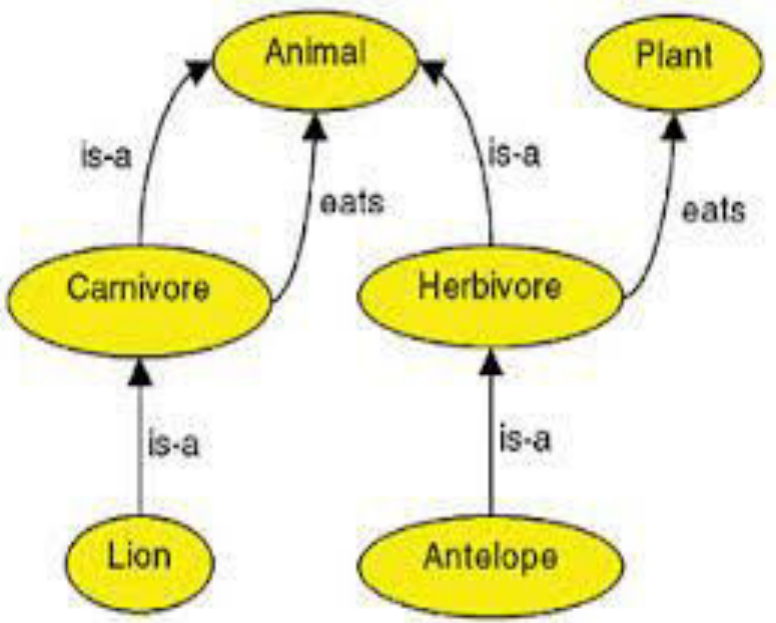
Ontology examples



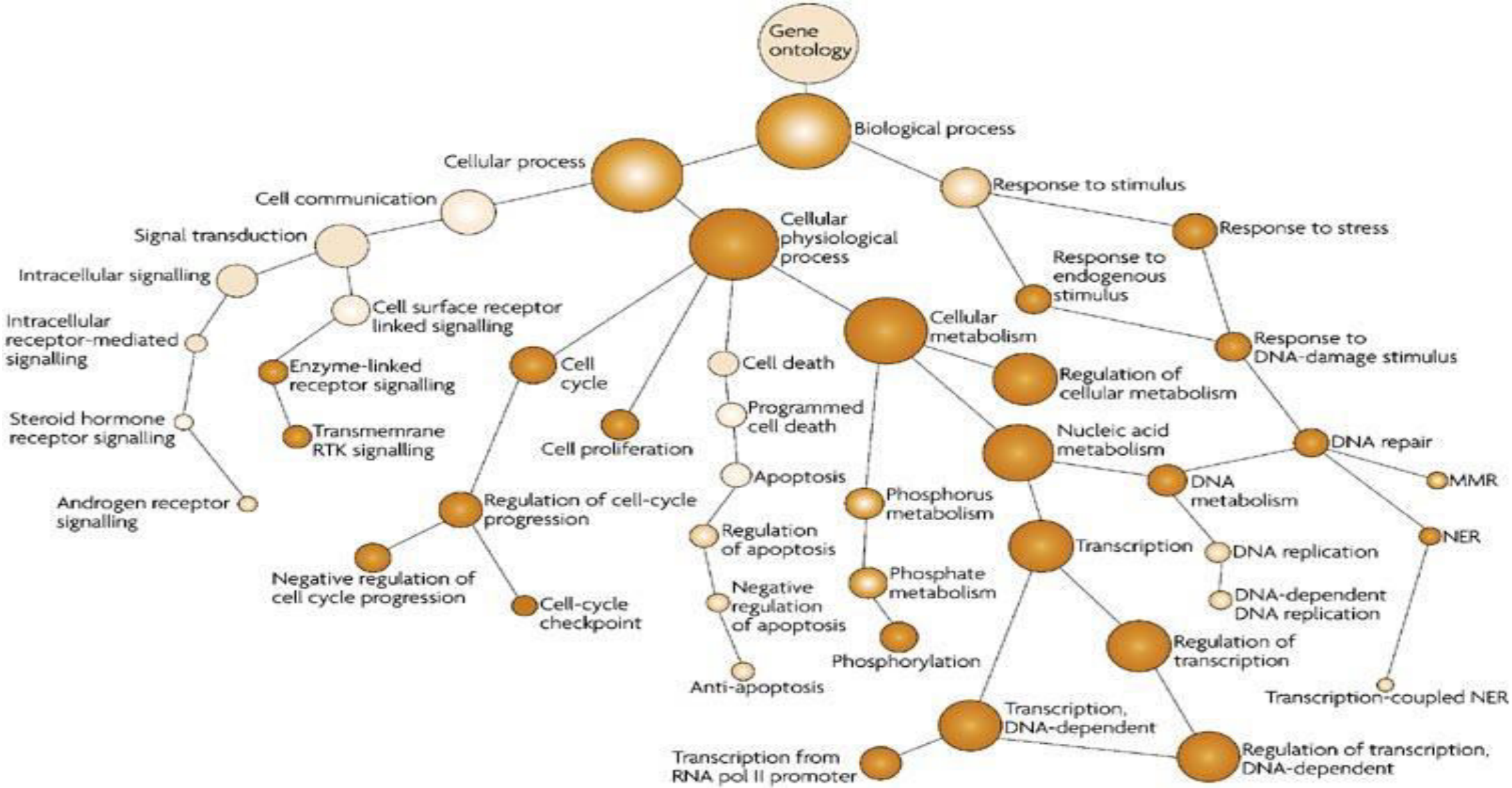
Euclid's geometry as an early example of consistent use of definitions and relationships

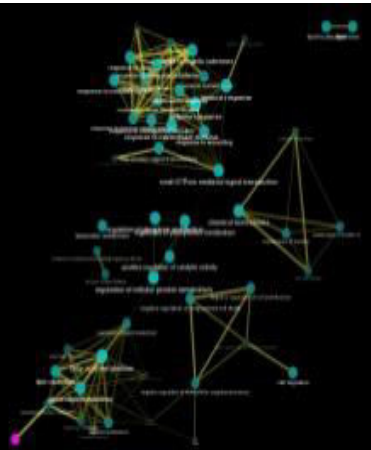


Chemical structures to capture domain knowledge in a very dense way.



Ontologized visualization of an ecological network with well-defined relationships.


Gene ontology
Biological process
Cellular process
Response to stimulus
Cell communication
Signal transduction
Cellular physiological process
Response to stress
Response to endogenous stimulus
Intracellular signalling
Cell surface receptor linked signalling
Cellular metabolism
Response to DNA-damage stimulus
Intracellular receptor-mediated signalling
Enzyme-linked receptor signalling
Cell cycle
Cell death
Regulation of cellular metabolism
Programmed cell death
Steroid hormone receptor signalling
Transmemrane RTK signalling
Cell proliferation
Nucleic acid metabolism
DNA repair
Apoptosis
DNA metabolism
MMR
Androgen receptor signalling
Regulation of cell-cycle progression
Phosphorus metabolism
Regulation of apoptosis
Transcription
DNA replication
NER
Negative regulation of cell cycle progression
Cell-cycle checkpoint
Phosphate metabolism
Negative regulation of apoptosis
DNA-dependent DNA replication
Regulation of transcription
Phosphorylation
Anti-apoptosis
Transcription-coupled NER
Transcription, DNA-dependent
Transcription from RNA pol II promoter
Regulation of transcription, DNA-dependent



Gene Ontology Network Enrichment Analysis

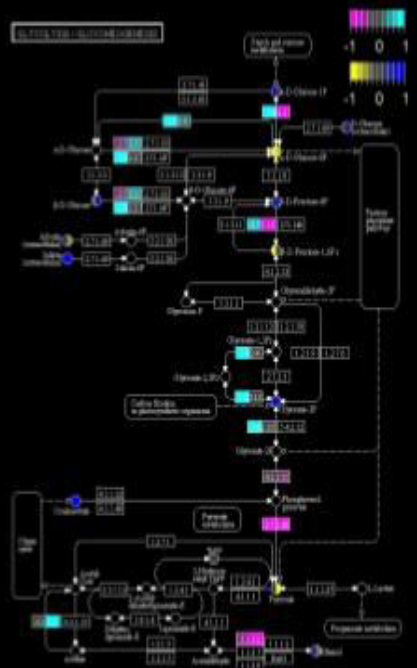
Dmitry Grapov, PhD

UCDAVIS GENOME CENTER

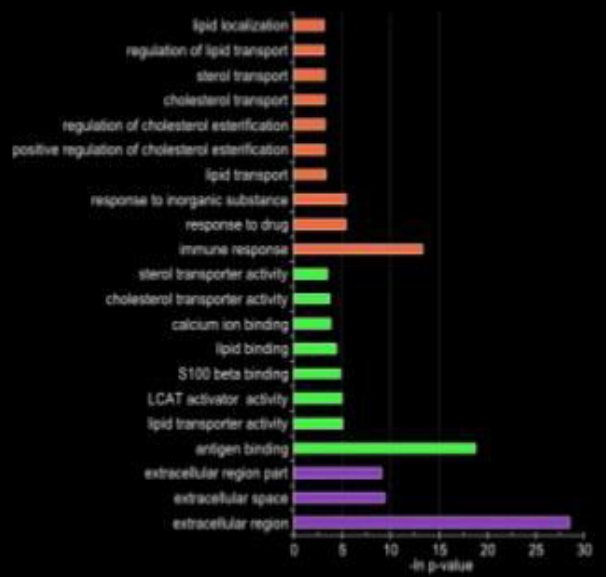
Enrichment or Overrepresentation analysis

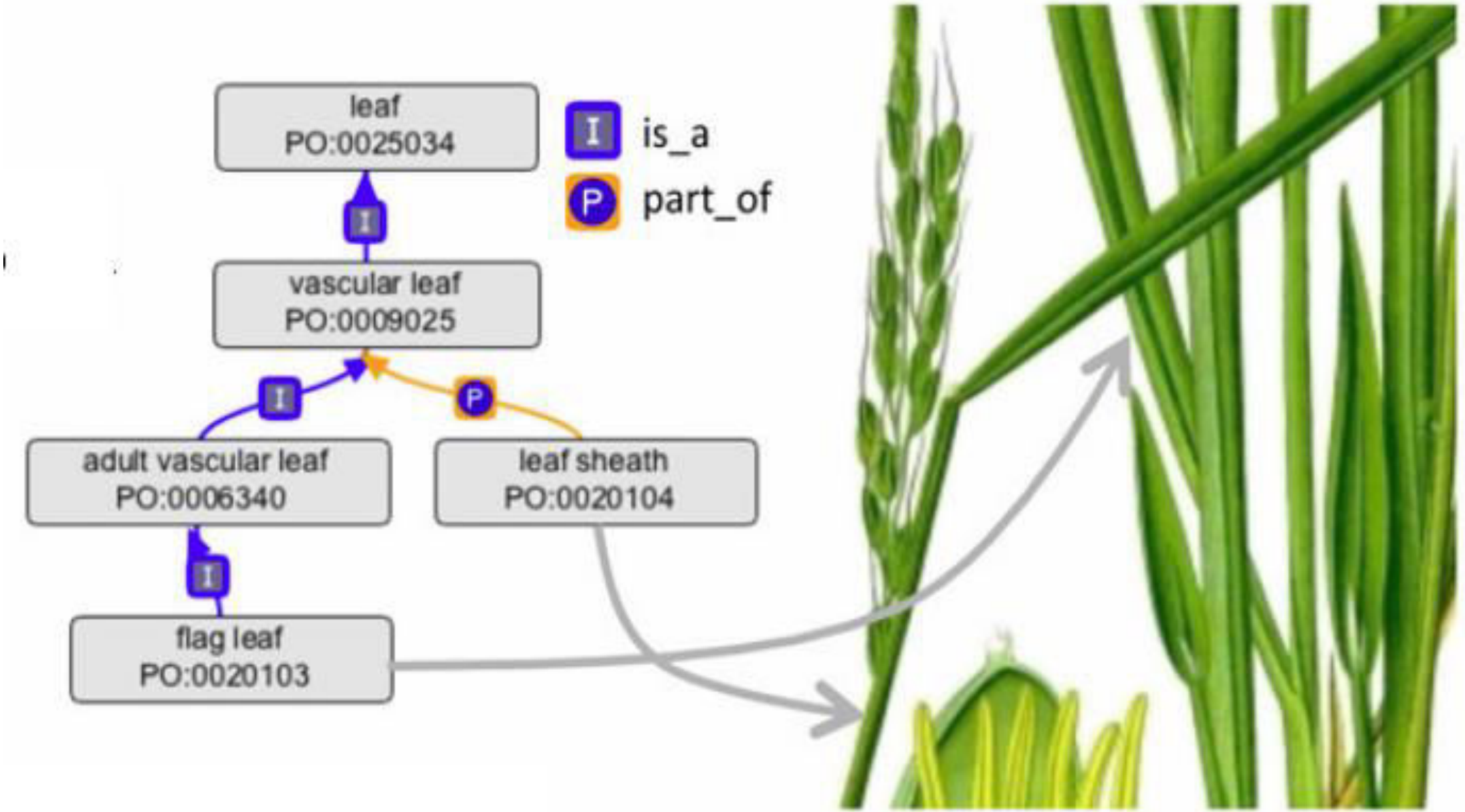
BioCyc · REACTOME · KEGG · UniProt

Biochemical Pathway



Biochemical Ontology




leaf
PO:0025034
I
is_a
P
part_of
I
vascular leaf
PO:0009025
I
P
adult vascular leaf
PO:0006340
leaf sheath
PO:0020104
I
flag leaf
PO:0020103

Plant Ontology Consortium

Advanced Search | Browse | Help | Request PO Term | Home

Search PO [ ] ◉Ontology ○Annotations ☐Exact Match [Submit]

Filter tree view

Filter by aspect
Aspect: All / Plant Anatomy / Plant Development Stage

Filter Annotation Objects Counts
Data source: AgBase / All / COSMOSS / Gramene Genes

[Set filters] [Remove all filters]

- **all : all [110950]**
  - **PO:0025131 : plant anatomical entity [110899]**
    - PO:0025117 : plant anatomical space [5]
    - PO:0009011 : plant structure [110899]
    - PO:0025161 : portion of plant substance [2]
  - **PO:0009012 : plant structure development stage [102919]**
    - PO:0025338 : collective plant structure development stage [49645]
    - PO:0025339 : plant organ development stage [46618]
    - PO:0025423 : plant tissue development stage [1]
    - PO:0001170 : seed development stage [30763]
    - PO:0025368 : trichome development stage [50]
    - PO:0007033 : whole plant development stage [101903]

Graphical View
Permalink
Download as XML
Download as flat file

has participant

Last updated 2012-04-04, SVN Version: 1491
How can we improve the PO? Send us your suggestions OR Contact the Plant Ontology Consortium

The Plant Ontology: Linking Genomic and Phenomic Data Across Plant Taxa

Plant Ontology Consortium Members and Curators*:

Laurel D. Cooper*, Justin Elser, Justin Preece and Pankaj Jaiswal*:
Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR
Ramona L. Walls* and Dennis W. Stevenson: The New York Botanical Garden, Bronx, NY
Maria A. Gandolfo: Department of Plant Biology, Cornell University, Ithaca, NY

Ontology Consultants:
Chris Mungall: Gene Ontology, Lawrence Berkeley National Lab, Berkeley, CA
Barry Smith: OBO Foundry, Department of Philosophy, University at Buffalo, NY

www.plantontology.org

In summary...

Why to build an ontology?

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To analyze domain knowledge

In summary...

Why to build an ontology?

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To analyze domain knowledge

How to build an ontology?

Basic ideas

- Yours will be different from mine
- Iterative process
- Initially, start with nouns and verbs
- A noun will be a class, attribute or instance
- A verb will be the relation
- Iterations are needed to further clarification

http://www.slideshare.net/rajith89/guidelines-to-create-an-ontology

Methodology

- Define concept
- Organize them to taxonomy
- Define relations among the classes
- Define attributes and their values
- Define instances
- Define axioms and function

Building the ontology …

- Sources to use
  - Experts
    - Ask everything want to know
    - Always keep manners
    - Grab their terminology
  - Documents
    - literatures, documents, technical information, etc.
    - Highlight underlying nouns and verbs
  - Existing ontologies
    - You are not the first one to think about that domain
    - Existing ontologies can be fully or partially reused

Building the ontology …

- Sources to use
  - Experts
    - Ask everything want to know
    - Always keep manners
    - Grab their terminology
  - Documents
    - literatures, documents, technical information, etc.
    - Highlight underlying nouns and verbs
  - Existing ontologies
    - You are not the first one to think about that domain
    - Existing ontologies can be fully or partially reused

http://www.cropontology.org

The Crop Ontology

a resource for enabling access to breeders' data

Elizabeth Arnaud[1*], Luca Matteis[1], Marie Angelique Laporte[1], Herlin Espinosa[2], Glenn Hyman[2], Rosemary Shrestha[3], Arlett Portugal[4], Pierre Yves Chibon[5], Medha Devare[6], Akinnola Akintunde[7], Jeffrey W. White[8], Mark Wilkinson[9], Caterina Caracciolo[10], Fabrizio Celli[10], Graham McLaren[4]

[1]Bioversity International, France, [2]International Center for Tropical Agriculture (CIAT), Colombia, [3]Genetic Resources Program (GRP), Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), Mexico, [4]Generation Challenge Programme (GCP) c/o CIMMYT, [5]UR Plant Breeding, Univ. of Wageningen, The Netherlands, [6]International Maize and Wheat Improvement Center - South Asia Regional Office (CIMMYT-SARO), Nepal, [7]International Black Sea University (IBSU) Georgia, [8]Centro de Biotecnología y Genómica de Plantas UPM-INIA, Spain, [10]Food and Agriculture Organization (FAO) of the United Nations, Office for Partnership, Italy

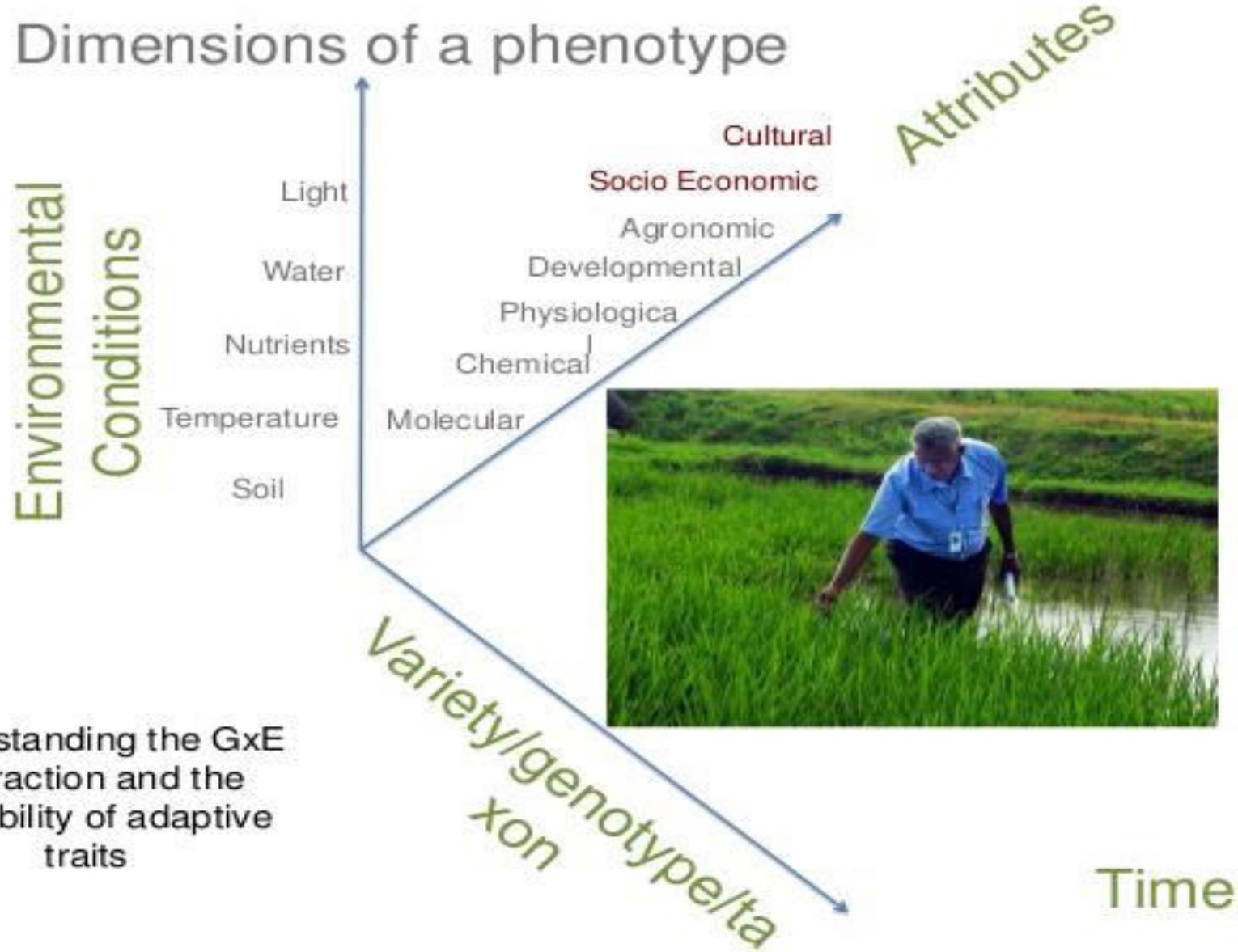
Generation Challenge Programme Workshop, 13th January 2014

In Plant and Animal Genomics Conference, San Diego, USA, 11-15th January 2014

CGIAR Crop Lead Centers

Since 2008

Dimensions of a phenotype



Understanding the GxE interaction and the heritability of adaptive traits

Harmonization and access to data

Breeders' data are often unstructured data - Complex free text used for phenotypes description

- No semantic coherence :
  - Same trait given different names by scientists
  - One trait named the same way for various species but refers to different plant structures
- Data and metadata are NOT interoperable and often not online

'Fruit colour'

Bean pod color

Rice grain or caryopsis colour

Maize Kernel Colour

Phenotype

It is a composite of an entity (e.g. fruit) and an attribute (e.g. shape) with a value (e.g. round):

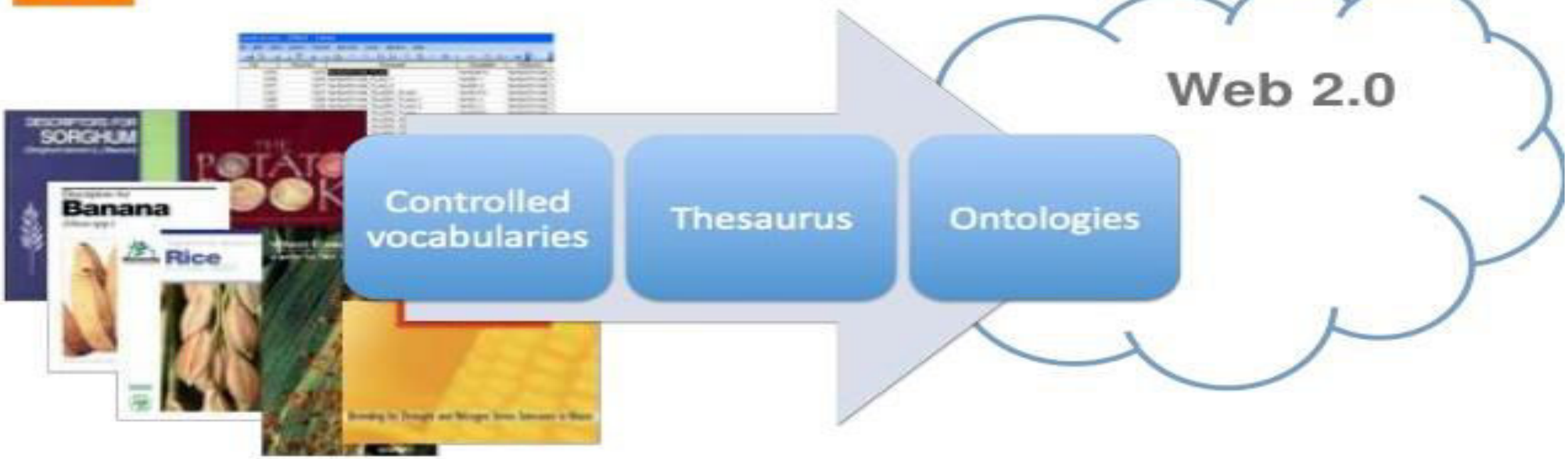
Entity + Attribute = Trait

Entity + (Attribute + Value) = Phenotype (observed)

fruit + (shape + round) = fruit shape round

-> round fruit is the phenotype

A range of controlled vocabularies



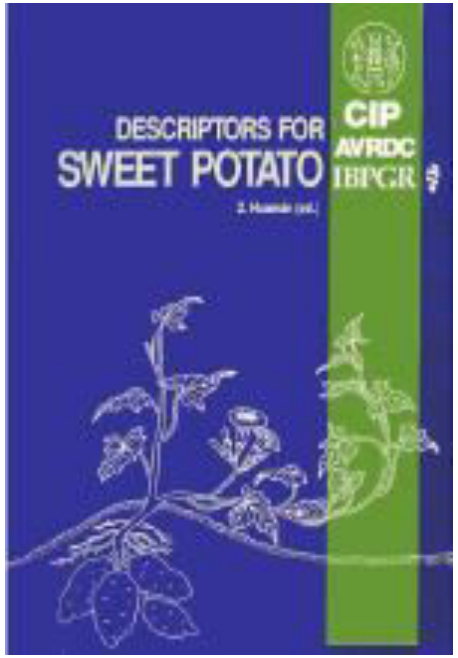
- From the controlled vocabularies build valid semantic ontologies consumabke by Web 2.0
- Best practices

Crop Ontology themes

- General germplasm information
- Phenotype and traits
- Plant anatomy and development
- Location and environment
- Trial management and experimental design
- Structural and functional genomics

Sweetpotato ontology

- Based on
  - FAO/IBPGRI/CIP descriptors
  - CIP/SASHA
  - NCSU
- Total: 206 variables
- Http://cropontology.org



Recent activities

- Continued consultation with cropontology community on best practices (PAG 2016, PhenoHarmonis conference, May)

- GT4SP meeting at NCSU adding/harmonizing NCSU traits & reviewing existing traits

- adding traits from drought research on sweetpotato

- update on crop ontology web-site on-going (to enable review at this meeting)

Crop Ontology Curation Tool

Home About Feedback

Crop Ontology for agricultural data

Integrated Breeding Platform
Today's tools for tomorrow's crops.

News

- Workshop PhenoHarmonIS '*Semantics for Harmonization and Integration of Phenotyping and Agronomy Data*', 9th-13th May 2016, Montpellier
- **Release of the Trait Dictionary template v5 (TDv5)!** integrates standard variables which are the combination of a trait, a method and scale. *NB: upload of TD in version 4 not supported anymore.*
- **Trait Dictionaries version 5: cassava, chickpea, cowpea, groundnut, lentil, pigeon pea, soybean, sweet potato, rice and wheat.** Click on a term to display the variables.
- **Mapping CO terms to high level reference ontologies** (TO, PO, PATO, CHEBI, EO, PDO, etc) is in progress
- Crop Ontology is partner of the **Planteome project** (NSF-awarded project #1340112). Planteome site
- Crop Ontology participates to the Agroportal **pilot project** - release of a beta version 1.0
  To upload an ontology
- Guidelines are available at the Crop Ontology wiki (new guidelines coming soon for the template 5)
- To get the code of your crop ontology , please contact helpdesk: helpdesk@cropontology-curationtool.org
  Crop and Agronomy Ontology Community Web Site: http://tinyurl.com/ho4j922

Search

Add New Terms API Help Agtrials Annotation Tool Register Login

Latest

OBO Ontology Trait Dictionary

General Germplasm Ontology

FAO/Bioversity Multi-Crop Passport Descriptors 109 terms ELIZABETHARNAUD
FAO/Bioversity Multi-Crop Passport Descriptors V.2.1 [MCPD V.2.1] standard, developed and co-authored by Alercia, A; Diulgheroff, S; Mackay, M.

Germplasm 386 terms SHRESTHA
germplasm

ICIS germplasm method 166 terms SHRESTHA
ICIS germplasm methods

Phenotype and Trait Ontology

Banana 67 variables RCRICHTON
Banana Trait Dictionary in template 5 - Bioversity & IITA - May 2016 - characterization and breeding variables

Location and Environmental Ontology

Country and Location 1118 terms SHRESTHA
Describes official ISO 3166-1 alpha-2, alpha-3 and numeric country codes along with location names.

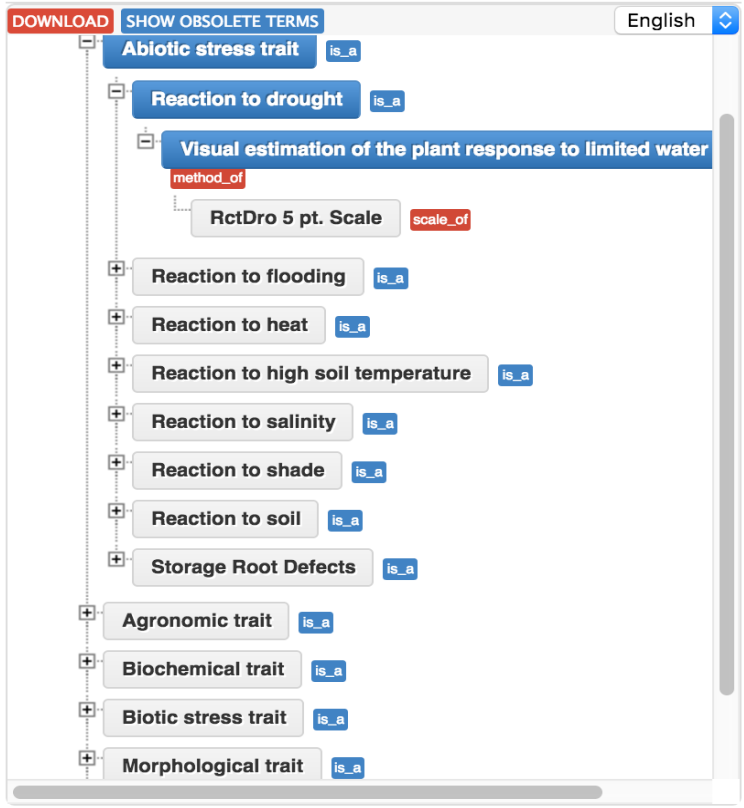
Crop Research 256 terms SHRESTHA
Describes experimental design, environmental conditions and methods associated with the crop study/experiment/trail and their evaluation.

Plant Anatomy & Development Ontology

Banana Anatomy 149 terms ELIZABETHARNAUD
Banana Anatomy

Plant Ontology 1710 terms COOPERL
The Plant Ontology describes plant anatomy and morphology and stages of development for all plants. The goal of the PO is to establish a semantic framework for meaningful cross-species queries across gene expression

Traits, methods and scales



Variables

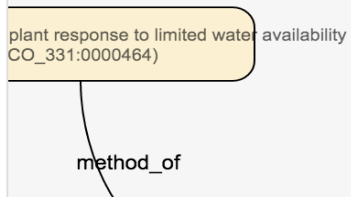
RnDrt_Et_1to9

Term information

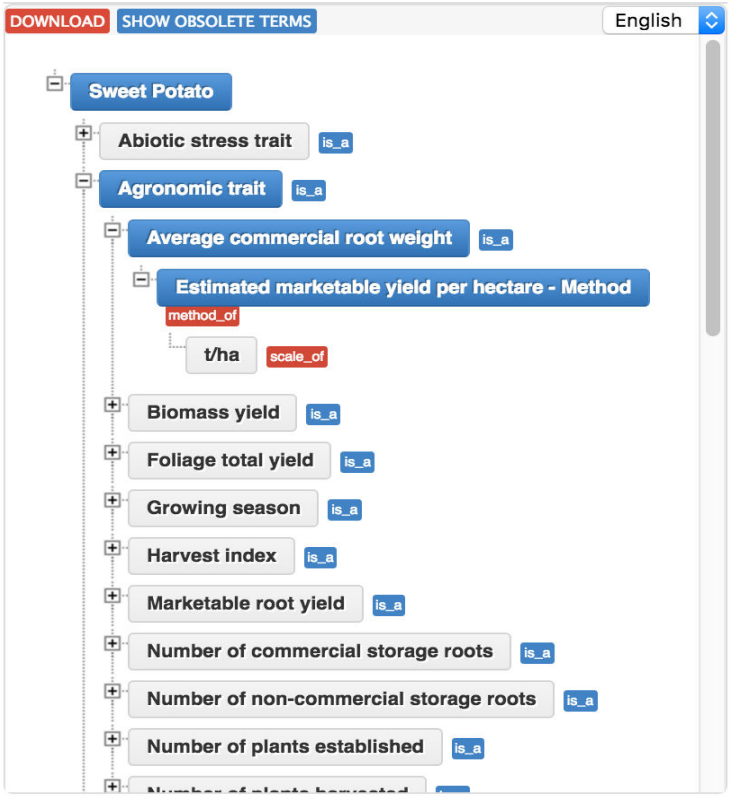
Visual estimation of the plant response | Permalink | General | 0 Comments

Identifier	CO_331:0000464
Method class	Estimation
Method description	The reaction is estimated by observing visually the appearance of plants in response to limited water availability
Method name	Visual estimation of the plant response to limited water availability
Method reference	Huaman1991:Def_7.1
created_at	Mon Jun 06 12:09:24 UTC 2016
name	Visual estimation of the plant response to limited water availability

Add a new attribute



Traits, methods and scales



Variables

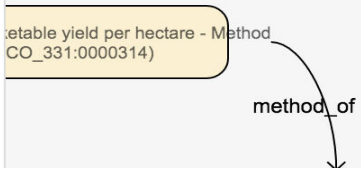
RtACRW_Cp_g

Term information

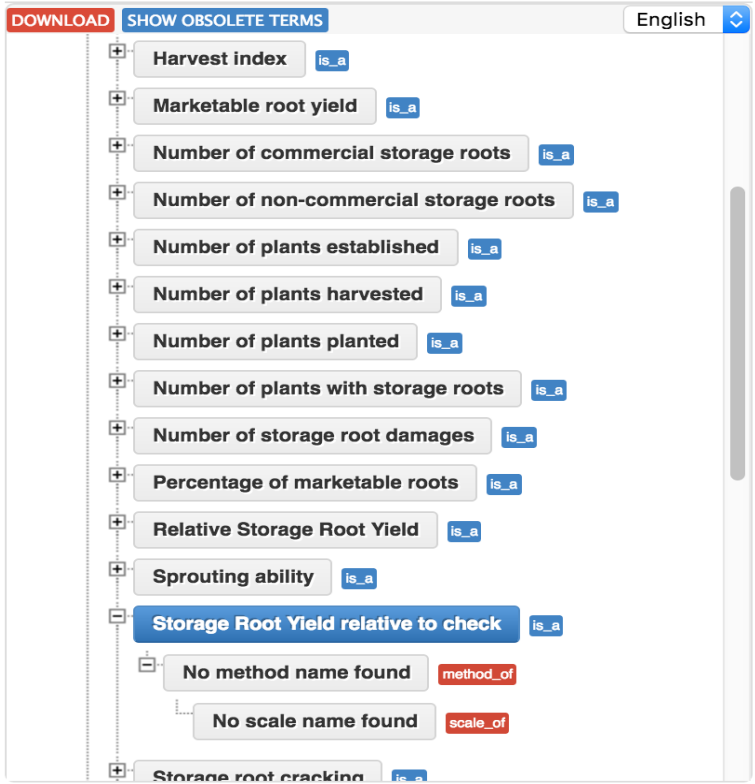
Estimated marketable yield per hectare | Permalink | General | 0 Comments

Identifier	CO_331:0000314
Formula	ACRW = CRW / NOCR
Method class	Computation
Method description	(Weight of commercial storage roots/ Number of non-commercial roots
Method name	Estimated marketable yield per hectare - Method
Method reference	Grueneberg2016:Def_36
created_at	Mon Jun 06 12:08:28 UTC 2016
name	Estimated marketable yield per hectare - Method

Add a new attribute



Traits, methods and scales



Variables

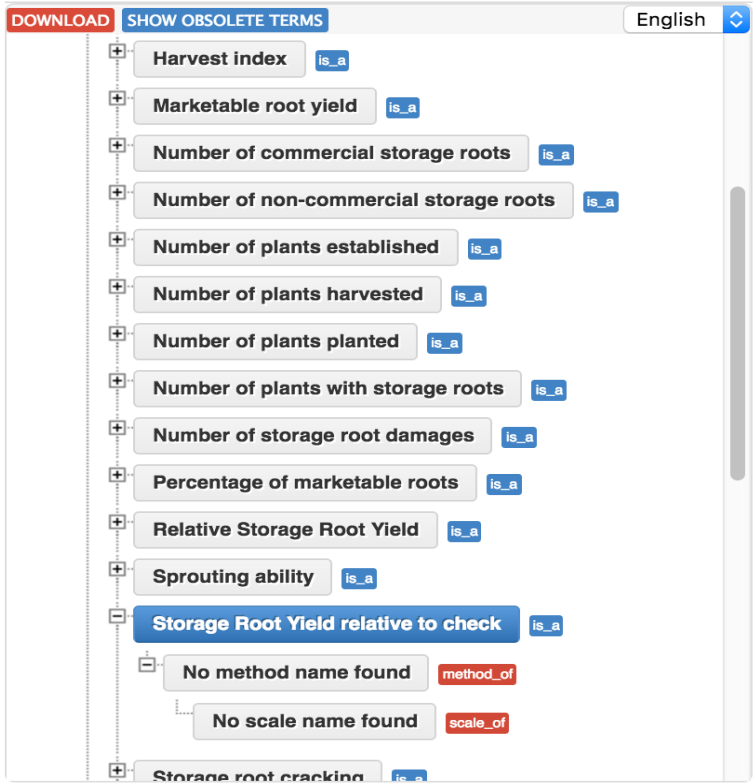
RtYldChk_Cp_pct

Term information

Storage Root Yield relative to check | Permalink | General | 0 Comments

Identifier	CO_331:0000425
Trait description	Overall calculation of relative root yield
Attribute	relative yield
Entity	Root
Main trait abbreviation	RtYldChk
Trait Xref	NCSU2016
Trait class	Agronomic trait
Trait name	Storage Root Yield relative to check
Trait status	Recommended
created_at	Mon Jun 06 12:10:42 UTC 2016

Traits, methods and scales



Variables

RtYldChk_Cp_pct

Term information

Storage Root Yield relative to check | Permalink | General | 0 Comments

Identifier	CO_331:0000425
Trait description	Overall calculation of relative root yield
Attribute	relative yield
Entity	Root
Main trait abbreviation	RtYldChk
Trait Xref	NCSU2016
Trait class	Agronomic trait
Trait name	Storage Root Yield relative to check
Trait status	Recommended
created_at	Mon Jun 06 12:10:42 UTC 2016

Traits, methods and scales

DOWNLOAD | SHOW OBSOLETE TERMS | English

- Sweet Potato
  - Abiotic stress trait (is_a)
  - Agronomic trait (is_a)
  - Biochemical trait (is_a)
    - Amylose content (is_a)
    - Anthocyanin content (is_a)
      - Anthocyanin - Method (method_of)
        - mg/g DW (scale_of)
    - Asparagine content (is_a)
    - Cyanidin content (is_a)
    - Peonidin content (is_a)
    - Phenol content (is_a)
    - Total Monomeric Anthocyanin content (is_a)
  - Biotic stress trait (is_a)
  - Morphological trait (is_a)

Variables

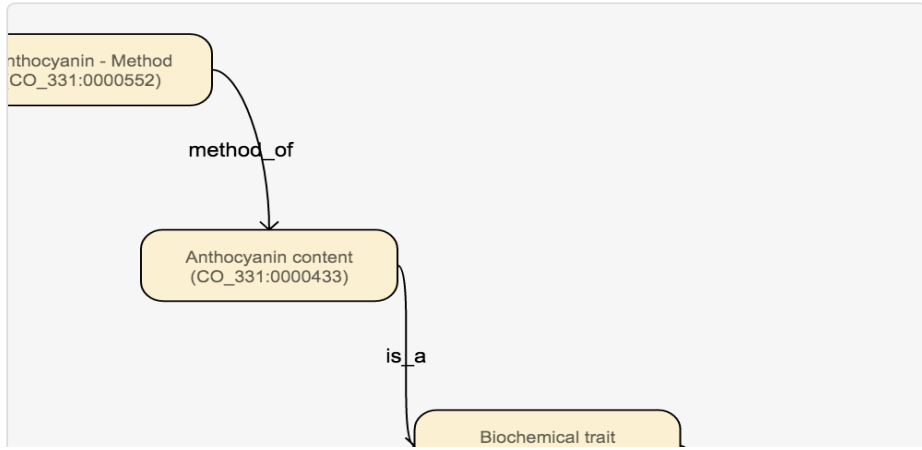
RtFlsAnt_Ms_mgpergDW

Term information

Anthocyanin - Method | Permalink | General | 0 Comments

Identifier	CO_331:0000552
Method class	Near Infrared Spectroscopy (NIRS)
Method name	Anthocyanin - Method
Method reference	NCSU-TBD2016
created_at	Mon Jun 06 12:10:50 UTC 2016
name	Anthocyanin - Method

Add a new attribute



Traits, methods and scales

DOWNLOAD | SHOW OBSOLETE TERMS | English

- Sweet Potato
  - Abiotic stress trait (is_a)
  - Agronomic trait (is_a)
  - Biochemical trait (is_a)
    - Amylose content (is_a)
    - Anthocyanin content (is_a)
      - Anthocyanin - Method (method_of)
        - mg/g DW (scale_of)
    - Asparagine content (is_a)
    - Cyanidin content (is_a)
    - Peonidin content (is_a)
    - Phenol content (is_a)
    - Total Monomeric Anthocyanin content (is_a)
  - Biotic stress trait (is_a)
  - Morphological trait (is_a)

Variables

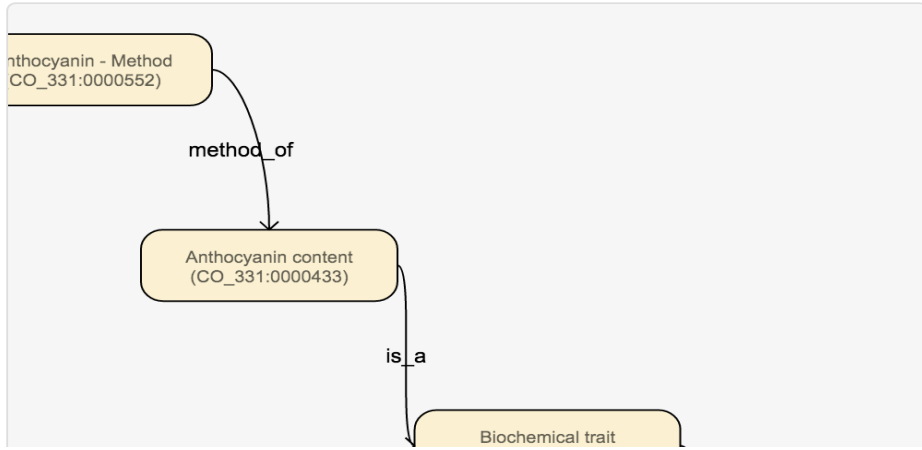
RtFlsAnt_Ms_mgpergDW

Term information

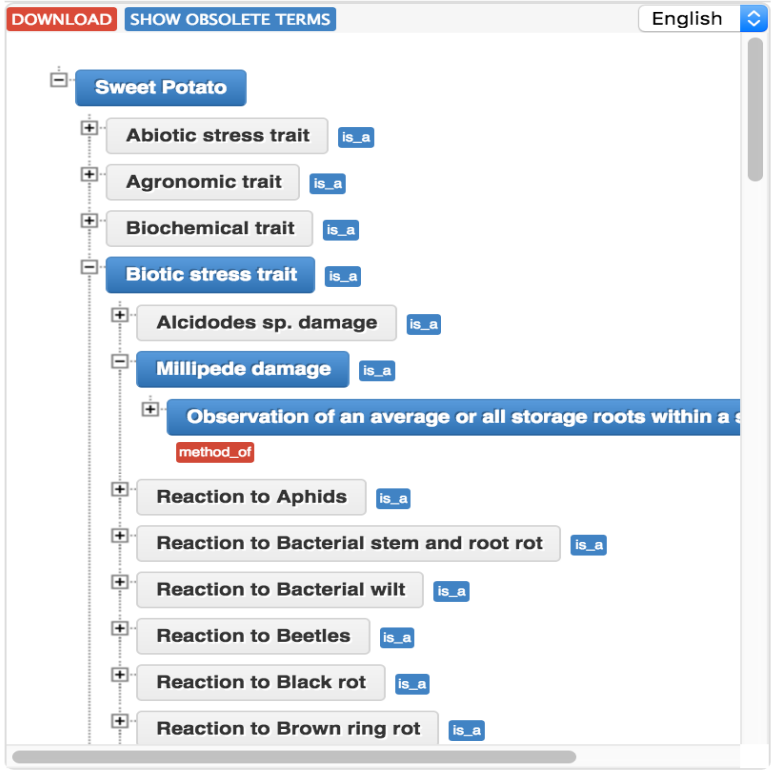
Anthocyanin - Method | Permalink | General | 0 Comments

Identifier	CO_331:0000552
Method class	Near Infrared Spectroscopy (NIRS)
Method name	Anthocyanin - Method
Method reference	NCSU-TBD2016
created_at	Mon Jun 06 12:10:50 UTC 2016
name	Anthocyanin - Method

Add a new attribute



Traits, methods and scales



Variables

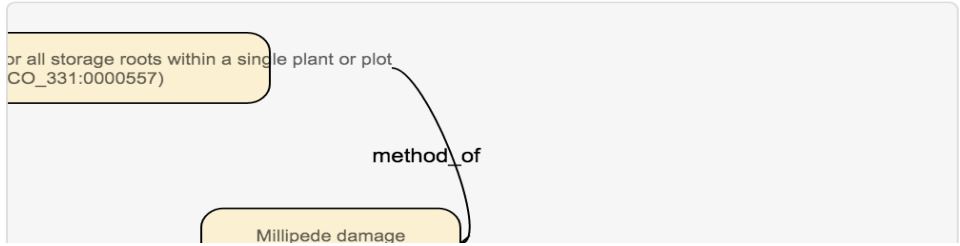
RtsMillDam_Et_1to9

Term information

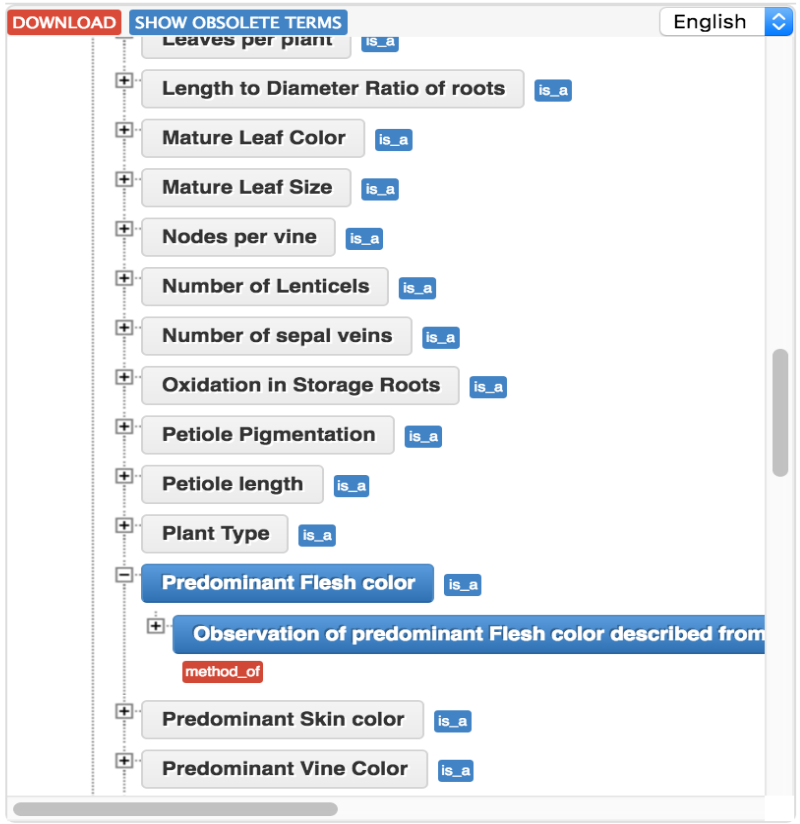
Observation of an average or all storage | Permalink | General | 0 Comments

Identifier	CO_331:0000557
Method class	Estimation
Method description	Visual estimation
Method name	Observation of an average or all storage roots within a single plant or plot
Method reference	CIP_SASHA_2016
created_at	Mon Jun 06 12:10:56 UTC 2016
name	Observation of an average or all storage roots within a single plant or plot

Add a new attribute



Traits, methods and scales



Variables

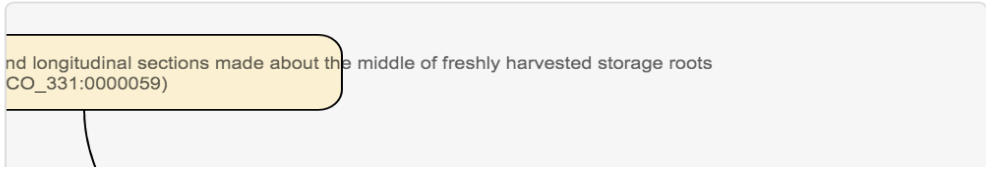
RtFlsColP_Et_1to9

Term information

Observation of predominant Flesh color | Permalink | General | 0 Comments

Identifier	CO_331:0000059
Method class	Estimation
Method description	Visual estimation
Method name	Observation of predominant Flesh color described from cross and longitudinal sections made about the middle of freshly harvested storage roots
Method reference	Huaman1991:Def_4.2.5.1, Grueneberg2010:Def_26
created_at	Mon Jun 06 12:07:50 UTC 2016
name	Observation of predominant Flesh color described from cross and longitudinal sections made about the middle of freshly harvested storage roots

Add a new attribute



On-going work & issues

- Re-view utility of traits & measurement methods
- Document measurement methods
- Harmonization of variable names
- Avoidance of trait duplication
- Maintaining backward compatibility with existing data

More research needed on

- Not including ‘sub-sampled variables’
  - That is the application has to take care of time-series variables of same type or several biological samples from a field
- How to best include photographic information

More emphasis on meta-data for an experiment

- Minimum information on a plant phenotyping experiment (MIAPPE)
- Is a **checklist** what to document about the experimental context
- http://cropnet.pl/phenotypes/?page_id=15

Minimum Information About a Plant Phenotyping Experiment (MIAPPE)

Attributes (concepts, subconcepts - in terms of ontology) marked by asterisk (*) are essential for a description of experiment (e.g. by Poorter et al. [26]); the rest forms an extended description. For some attributes examples of possible values are listed.

Checklist section	Attributes	Source list / Biosharing ID / Reference	Recommended ontologies
General metadata	Unique identifier* Title* Description* Submission date Public release date Publications Laboratory address and contact details	Default ISA-Tab configuration [1]	OBI, Ontology for Biomedical Investigations [2] CRO, Crop Research Ontology [3]
Timing and location	Timing: Start of experiment (date)* Duration (days/months/years)* Experiment location: Geographic location* Latitude and longitude Altitude Inclination and aspect Habitat	Poorter et al. [4] Morrison et al. [5] CIMR [6]: Environmental Analysis Context [7]	OBI, Ontology for Biomedical Investigations [2] GAZ, Gazetteer [9]
Biosource	Organism (taxon)* Infraspecific_name* Infraspecific_rank Common name Genotype Organism age Life stage Seed preparation: Seed source* Pretreatments Conservation conditions	MIxS Plant-associated environmetal package [10] Yilmaz et al. [11] FAO/Bioversity Multi-Crop Passport Descriptors V.2 (MCPD V.2) [12]	UNIPROT Taxonomy [13] NCBI Taxonomy [14]

Next steps

- Reviewing traits and variables for adequateness & measurement procedures (breeders)

- updating references to protocols

- vice-versa: updating protocols to CO-identifiers

- establishing a process of review