Ontologies for data integration and data reuse

Georgios Gkoutos
John Doonan

University of Birmingham, Aberystwyth University, United Kingdom

Can we adapt biomedical phenotype data standardisation efforts to plant phenotyping data?

- Example showing utility of inter-species phenotype data to identify mechanisms of rare genetic defects in humans
- Prospects for utilising similar approaches in crops

the Potential Power of the Phenotype: an example from biomedicine

Phenotypic space is very information rich but how can we exploit it, systematically?

A machine-readable phenotypic database that can be queried across species

Objective: Facilitate phenotype-driven gene function discovery and comparative pathobiology (in humans)

Aim: to construct "A platform for facilitating mutual understanding and interoperability of phenotype information across

- species,
- domains of knowledge,

And amongst people and computers

PATO - Phenotype And Trait Ontology – as the basis of integration and interrogation

PATO is now widely used as the computationally compatible community standard for phenotype description

- many consortia (e.g. Phenoscape, The Virtual Human Physiology project (VPH), IMPC, BIRN, NIF)
- most of the major model organism databases, (e.g. Flybase, Dictybase, Wormbase, Zfin, Mouse genome database (MGD))
- international phenotyping and clinical genetics projects
A layered approach to using PATO

- Phenotypic data is too diverse
- Different communities use different terminology
- Multicellularity and development add complexity
- Environment, experimental conditions, etc add complexity

→ use a layered approach with PATO as the kernal

Semantic Components Layer - the collection of subsidiary ontologies necessary for PATO function

PATO Conceptual Layer provides the kernal

A Unification Layer for interoperability between species specific phenotype ontologies
A Formalisation Layer allows semantic reasoning of large scale ontologies

Cross species integration framework: PhenomeNET

- A PATO-based cross species phenotype network based on experimental data from 6 model organisms and human disease phenotypes (includes clinical data)
- Integration of anatomy and phenotype ontologies
  - more than 3,000,000 classes and 6,000,000 axioms
- PhenomeNET forms a network with more than 500,000 complex nodes

Can we achieve phenotype data standardization for plants?

Mouse (MP) Normal heart Tetralogy of Fallot
Human (HPO) Normal heart Tetralogy of Fallot

- Used to predict all known human and mouse disease genes
- Reveals previously unknown heart function for mouse genes (Adam19 and Fgf15) and other new candidate genes for mammalian heart disease
- Extended to many other diseases, particularly effective for rare human genetic diseases

Plant PhenomeNet as a demo that semantic similarity measures can be applied for plant data

Algae  Lower land plants  Arabidopsis  Rice  wheat & other crops

MODELS CROPS

Anika Oellrich et al., Sanger Centre
Planteome / cROP project - Common Reference Ontologies and Applications for Plant Biology

New Version Release: 46m annotations, 1.8m objects covering 87 species

Acknowledgements

Michael Ashburner
Robert Hochstrad
ERA-CAPS expert working Group
Uli Stuhlf
Björn Usadel
Ruth Baxton
Christian Bachem
Quesneville Hadi
Grassien Wilhelm
Paul Wiley
Theo Saat
José Pereira Leal

Eslam Al-Herish
Chuan Lu
Anyela Camargo-Rodriguez

IMPC
Steve Brown
Ann-Marie Malton

Gene Ontology Consortium
Gene Ontology Consortium

Phenotype RCN
Paula Mabee

Phenotype RCN consortium
Fabio Fiorini

PPP group
Evolve
Steve Cannon
Anika Oeltrich
Laurel Cooper
Fabio Fiorini
David Grant
Lisa Harper
Scott Kalberer
Carolyn Lawrence
Johnny Lloyd
David McKin
Noa Mendoza
Laura Moore
Luca Mueller
Rex Nelson
Ramona Walls