Phenotypic Data Exchange and Processing Using ISA-Tab Standard: transPLANT Project Use Case with GnpIS.Ephesis and Bii

Cyril Pommier, Pawel Krajewski, Thomas Letellier, Hanna Cwiek, Célia Michotey, Pierre Roumet, Raphaël-Gauthier Flores, Guillaume Merceron, Erik Kimmel, Michael Alaux, et al.

To cite this version:
Cyril Pommier, Pawel Krajewski, Thomas Letellier, Hanna Cwiek, Célia Michotey, et al.. Phenotypic Data Exchange and Processing Using ISA-Tab Standard: transPLANT Project Use Case with GnpIS.Ephesis and Bii. 23. Plant and Animal Genome (PAG), Jan 2015, San Diego, United States. hal-02739236

HAL Id: hal-02739236
https://hal.inrae.fr/hal-02739236
Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Phenotypic data exchange and processing using ISA-Tab standard

TransPlant project use case with GnpIS.Ephesis and Bii.

**TRANSPLANT, ISATAB FOR PHENOTYPING**

TransPLANT

- 4 year EU FP7-funded project (DG CONNECT) coordinated by EMBL-EBI,
- Until septembre 2015
- An I3 project with elements of coordination, service and RTD
- [http://www.transplantdb.eu/](http://www.transplantdb.eu/)
- Aims
  - Provide of an integrating point of interactive access to diverse data sets
  - Develop common standards for use within transPLANT and a wider community
  - Engage with other related communities to share experiences, tools and roadmaps

**Transplant Standardization**

- Work Package 3
  - Community standards for the interoperability of data resources
  - Aim: To develop recommendations concerning annotation and formatting methods of data sets containing phenotypic observations
- Main deliverables
  - Recommended ontology set for use in phenotype description and epigenetic variability
  - Format specifications for data exchange by flat file and web services

**Recommended ontology set**

- Ontologies
  - References
    - Plant Ontology, cROP
  - Application Ontologies
    - Species/domain specific
    - Crop Ontology: [www.cropontology.org](http://www.cropontology.org)
    - Vitis : Grape Ontology
    - Wheat : Wheat ontology
    - INRA Ontologies
  - EO, to be developed
  - Chebi, Brenda Enzyme, Golm Metabolome
  - User specific ontology

- [References Ontologies](#)
  - Applied/Variables Ontologies
  - Thesaurus
  - BreedWheat
    - [Leaf Area](#)
    - [BreadWheat Method](#)
    - Yield
  - Vitis
    - [Leaf Area](#)
    - Budbreak date
  - Wheat
    - [Leaf Area](#)
    - [BreadWheat Method](#)
    - Yield

2/3/2015
ISA-TAB format: Investigation/Study/Assay

http://www.isa-tools.org
Sansone & Rocca-Serra (2012)
- text files
  - structures/terms:
    - source/sample
    - factor
    - characteristics
    - protocol
    - unit
    - extract
    - comment
- supports annotation
Tools:
  ISA Software Suite
  To-BIF converter

BII USE CASE

BII

- ISA TAB Archive system
- Index capabilities
- Polapgen/Transplant instance
  - http://cropnet.pl/transplant-bii/

Bii manager

- Administrator credentials form

Bii Manager

- Data management

Bii Manager

- Loading ISA-TAB into database
GnpIS: user web interfaces

**Transversal tools**
- Quick search
- BioMart
- Galaxy

**Thematic tools**
- Genomic
- Synteny
- Sequences
- Polymorphisms
- Markers
- Maps
- GWAS
- Genomic Selection
- Expression
- Genetic Resources
- Phenotypes

GnpIS USE CASE

- Integrative information system
  - Plant and bioaggressors
  - Multi thematic
    - Genetic
    - Phenotyping
  - Public and private data: partners, international consortium
- INRA: French National Center for Agronomic Research
- URG1: INRA Research laboratory
  - Production
  - Data integration
  - Development

GnpIS: user web interfaces

**GNPIS PHENOTYPING DATA**
Phenotyping in GnpIS Ephesis application scientific perimeter

- Phenotype and environment Experimental data
  - Whole Trials
  - Generic: Ontology driven
    - Levels: Whole trial, plot, plant data in experimentation
    - Observation variables: traits, environment, ...
  - Multiple factors
    - Genotype factor is mandatory
    - 0 to n other factors: treatments, stress, etc...
  - Punctual observation or temporal series/kinetic

Phenotyping Interface Overview Ephesis

- https://urgi.versailles.inra.fr/ephesis

Search criterions overview

- Three tabs
  - Genotype
  - Variables
  - Trial
- Query helper
  - Consistent criterion combination
  - Selection in one list limits choices for the other
  - Autocompletions

Genotype/Germplasm tab

- Taxon
- Genetic resource accession
- Collections
- Panel (soon)
Variable tabs

- Ontology driven
- Reference Ontologies
- Specific ontologies
  - Trait
  - Environment

Trial details tabs

- Trial identification
  - ID
  - Name
- Year, Date
- Experimental site

Multi Trial Dataset

Trial Details

GnpIS Ephesis ISA Tab export

- One search result ➔ one ISA archive / investigation
- Each Trial ➔ one study
- Investigation
  - Ontologies declaration
  - Ephesis query parameter, date, version
- Study
  - Trial identification (Number, name, site, date, design, Contacts, ...)
  - Trial Publications
  - Experimental Factors: Cultural practice (itk), nitrogen, water
  - Plant Material list: full identification (taxon, variety, BRC, Accession Number, Lot Number) + Isa Archive short ID

GnpIS Ephesis ISA Tab export

- Assays
  - Type: phenotyping
  - Factor Values: low nitrogen, high water, ...
  - Assay Name: combination of factors: lownitrogen+accession1, lownitrogen+accession2, highnitrogen+accession1, highnitrogen+accession2
- Trait definition file
  - Reference existing ontologies, including Scale, unit, protocol (Crop Ontology, INRA Ontologies, ...)
- Raw data file
  - Not mandatory
  - Not used (volume) can be a pointer/URI to an archive in the future
- Derived Data File
  - Use ID declared in ISA and Trait files
    - Germplasm + Factor + Assay Name
    - Traits
    - As many datafile by assay as needed
    - Trace the transformation protocol (analysis tool, data transformation procedure, etc...)

Germplasm + Factor = Assay Name

Traits

As many datafile by assay as needed

Trace the transformation protocol (analysis tool, data transformation procedure, etc...)

Germplasm + Factor + Assay Name

- As many datafile by assay as needed
- Trace the transformation protocol (analysis tool, data transformation procedure, etc...)
STANDARDS AND DISTRIBUTED INFORMATIONS SYSTEMS

Transplant Phenome

GnpIS.Ephesis: Phenome Integrative portal

- French Plant Phenotyping Platform
  - High throughput Phenotyping
  - Field
  - Controlled environment
  - 9 Nodes across France

- Distributed information system

- Open API Web Service
  - Need of a common Phenotyping OpenAPI
  - Need to build a metadata standard
    - PODD
    - Crop research Ontology
    - Distributed index system, document based (SolR, Elastic Search)

Perspectives

- Isa Tab standard dissemination
- Dataset building for analysis
- Phenotype API building
  - Phenome
  - France
  - Europe
  - International
  - Triticeae Toolbox?, IBP?, PODD?, ...

Contact: cpommier_at_versailles.inra.fr

Aknowledgements

Pommier C., Krajewski P., Letelier T., Cwiek H, Michotey C., Markiewicz A., Roumet P., Flores R, Merceron G., Kimmel E, Alaux M, Steinbach D, Quesneville H.