Australian Phenotyping Platforms:
the Phenomics Ontology Driven Data Repository (PODD)

Dr Xavier Sirault

Research Team Leader | Plant imaging and image analysis
Co-project Leader | Phenomics Informatics and Modelling Tools

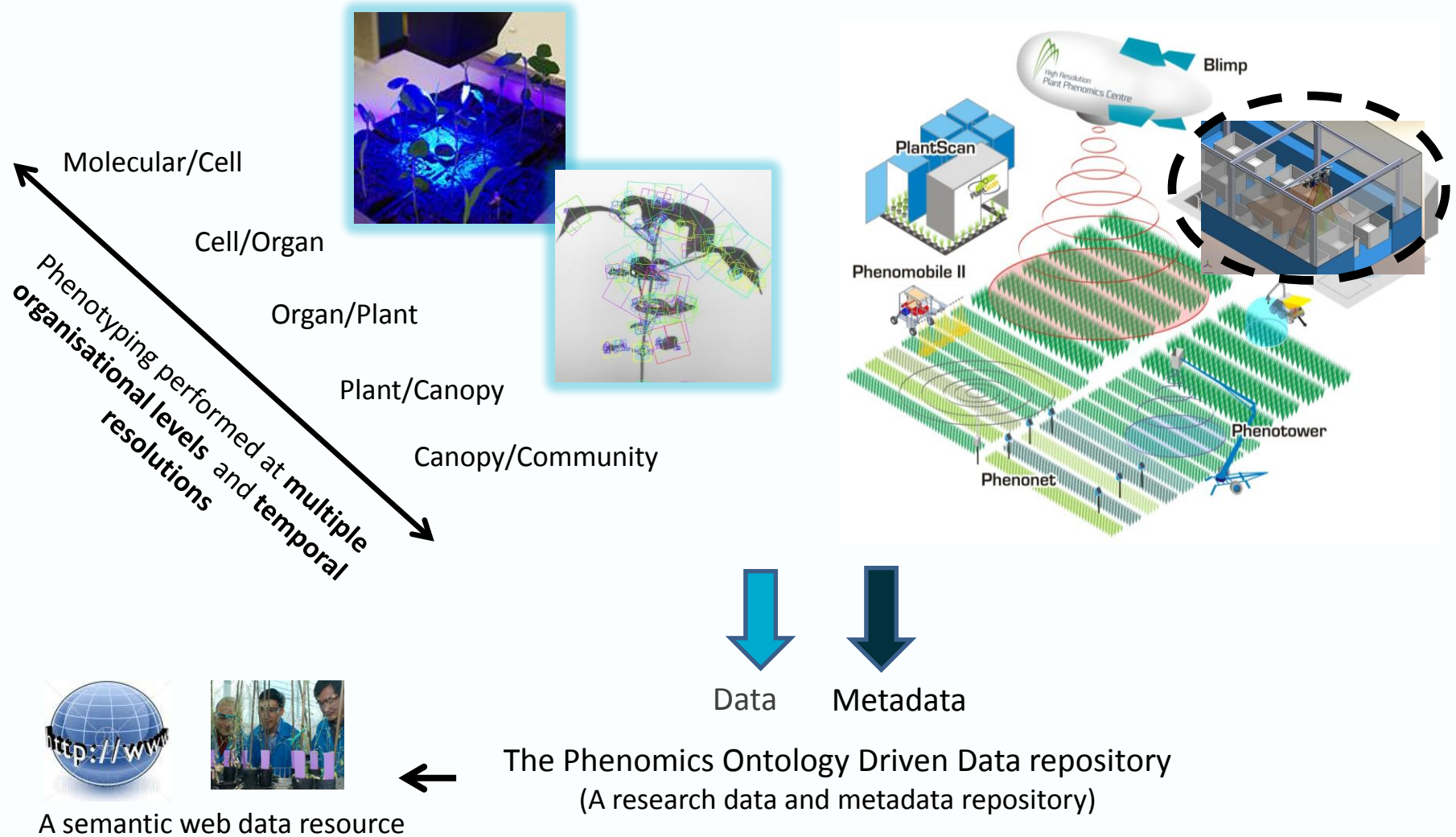**CSIRO Plant industry / High resolution plant phenomics centre**

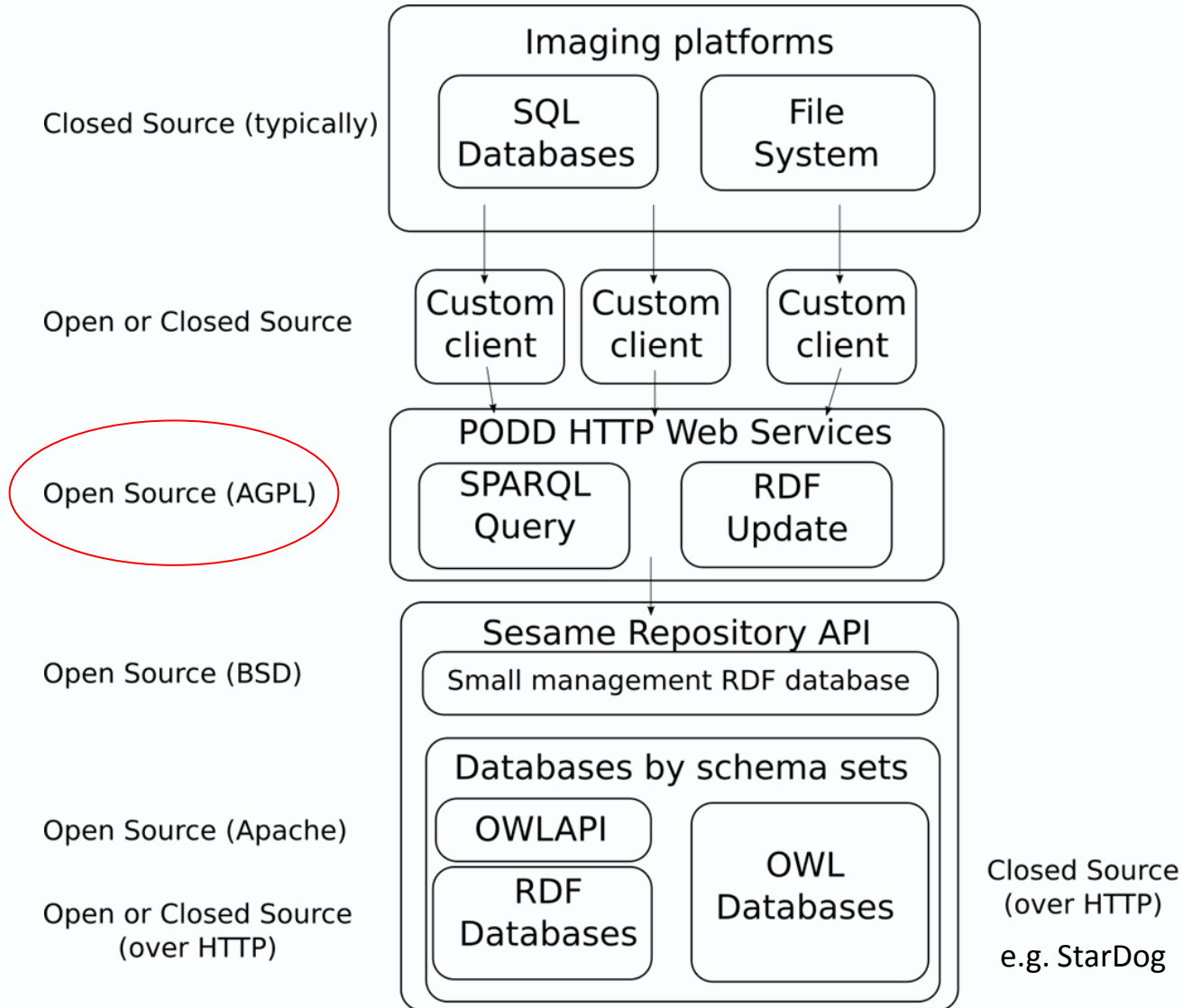Crop Ontology and Phenotyping Data Interoperability Workshop
Montpellier (3rd April 2014)

An Australian Government Initiative
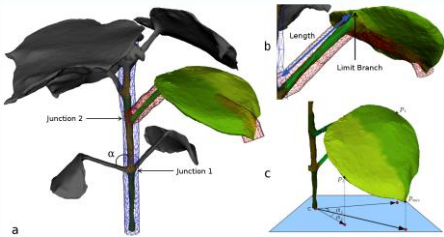National Collaborative Research Infrastructure Strategy

Australian National University

CSIRO

Managing Phenomics data from multiple heterogeneous, high volume, high resolution data generation platforms

Molecular/Cell

Cell/Organ

Organ/Plant

Plant/Canopy

Canopy/Community

Phenotyping performed at **multiple organisational levels** and **temporal resolutions**

Data   Metadata

The Phenomics Ontology Driven Data repository
(A research data and metadata repository)

A semantic web data resource

CSIRO

PODD Workflow

Managing Data workflow

>10GB data per plant (multi-modal images)
~200 plants per day

PlantScan

Images

Long term image store

(Peta-Bytes data archive)

Images

Run logging, calibration parameters

HPC: compute and process

Bagged images and meshes

Data Access Portal collection

(searchable, citable via DOIs)

Processes, data image/mesh location

PODD

Bag location

https://data.csiro.au/dap/home

Concept of events

CSIRO
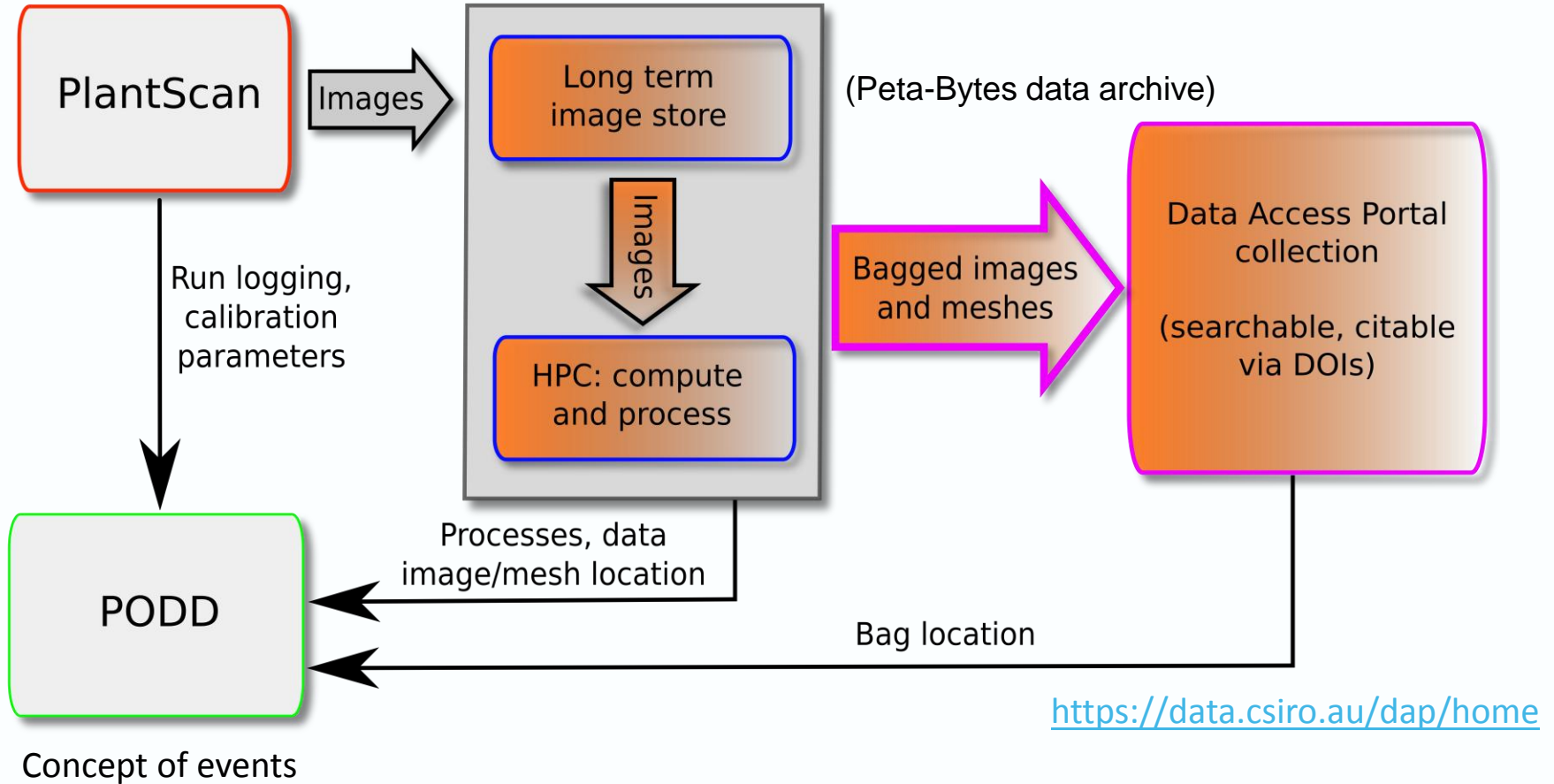
PODD Events

- Event created in PODD for each set of images from a platform

- Pots linked to images using events

- Pots currently linked to:

  - Genus, species, genotypes

  - Planting date

- Pots will be linked to more events in future:
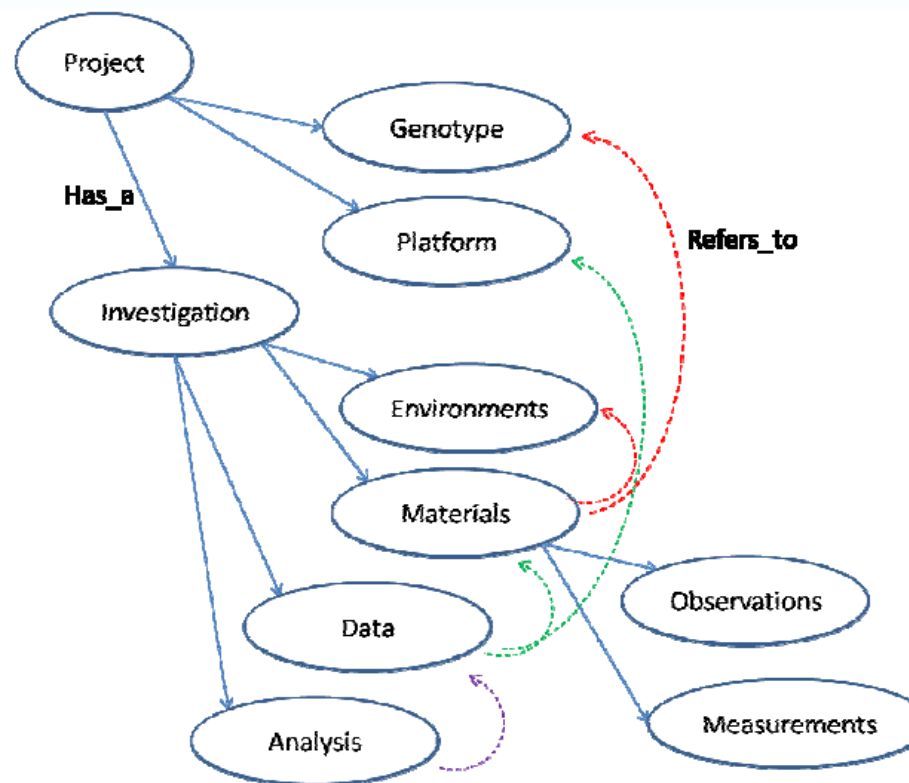
  - Watering

  - Client reviews

**Ontology of events
(co-implementation)**
Dr Pascal Neveu (INRA - MISTEA)
Mr Alexandre Mairin (INRA -MISTEA)

# *Modelling phenomics metadata with PODD ontologies*

In the PODD Ontology we model every thing as objects:

- Experiments (Investigations)
- Plants (Materials)
- Treatments
- Environments
- Measurement Platforms
- Temporal Events
- Raw Data (Data)

We then define the relationships between objects:

- Investigation has Material
- Material has Observations
- Material references Genotype
- Data references Material

PODD ontologies

Scientists can reuse existing ontologies to define the parameters for their projects
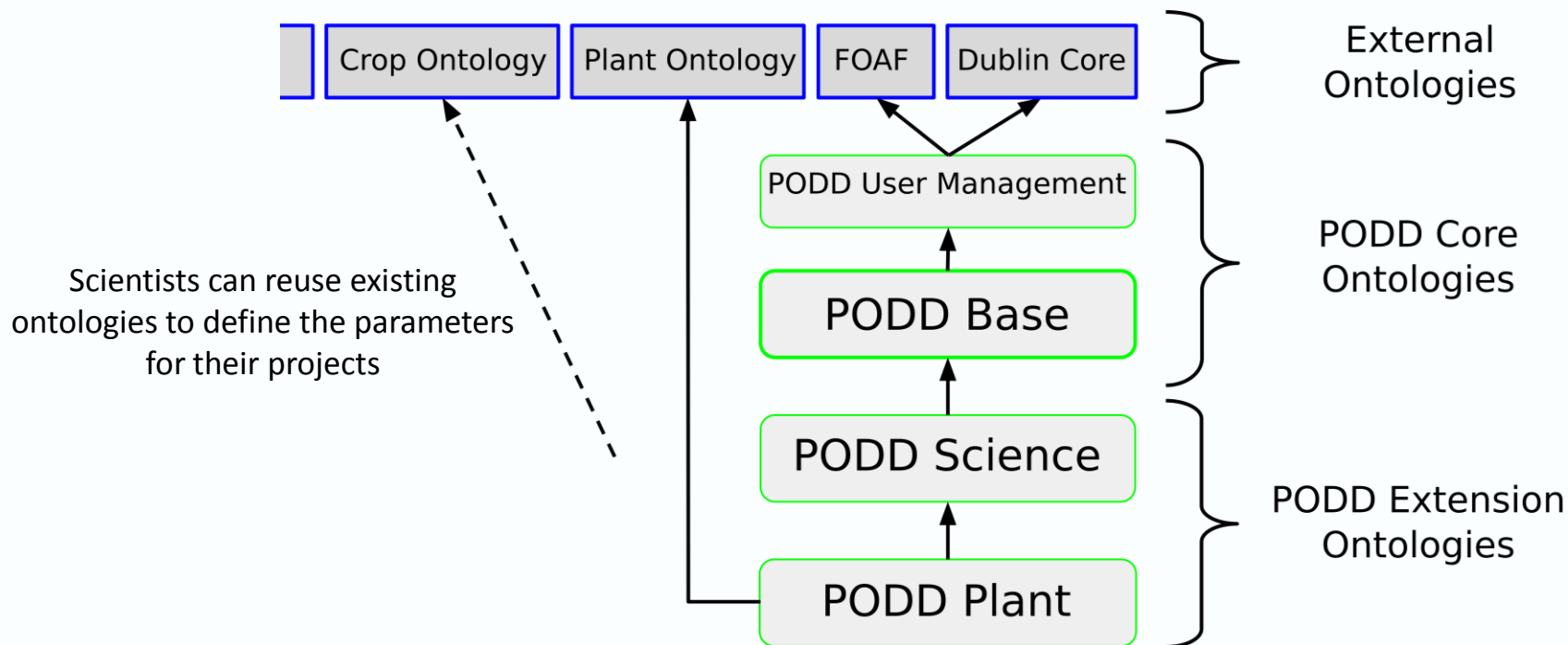
Model domain objects as ontological objects
Base ontology: domain independent
Phenomics ontology: domain specific

PODD Object Model

Objects represented semantically

➢ captured in Resource Description Framework triples (subject-predicate-object)using concepts defined in Web Ontology Language, named using version Internationalised Resource Identifier

➢ overall goal to allow for federated SPARQL queries acrossPODD instances to reuse existing plant phenomics data whereever possible (this is implemented, but datasets are small so far)

➢ flexible integration of all project management data for a research group in a single PODD instance using the most relevant schema ontologies in each project

➢ similar SPARQL queries across all artifacts, regardless of the schema ontologies currently used (results limited by the practicality of the query for the particular schema ontologies in use)
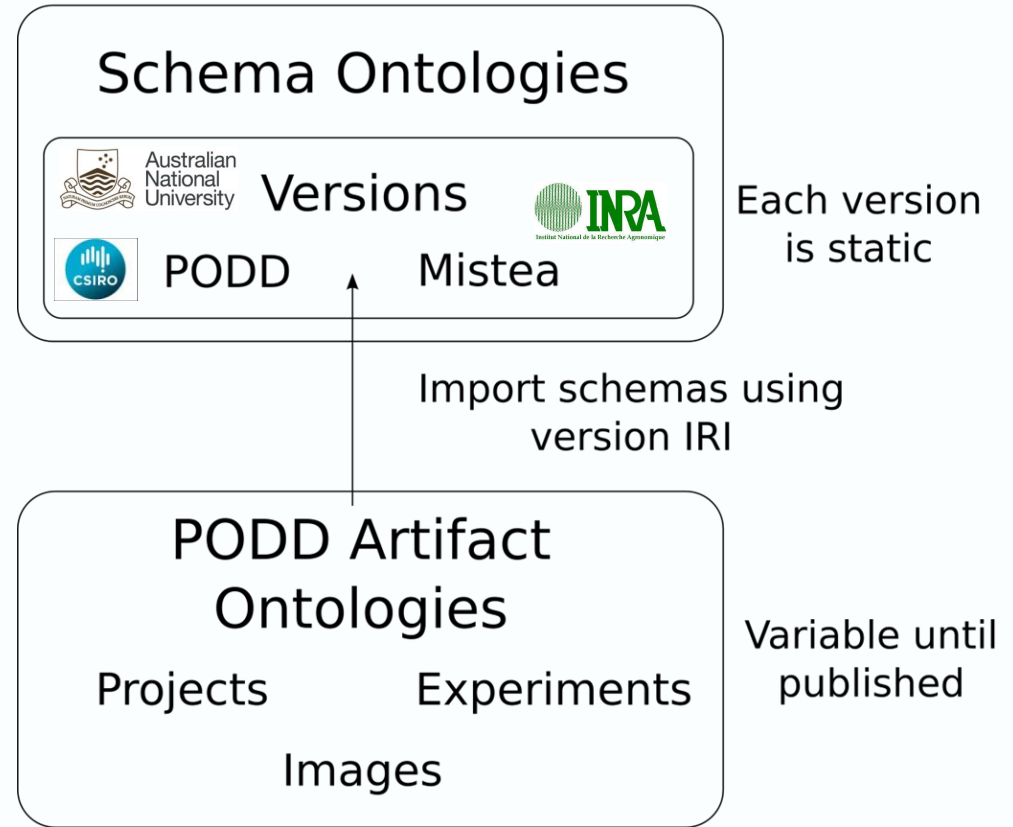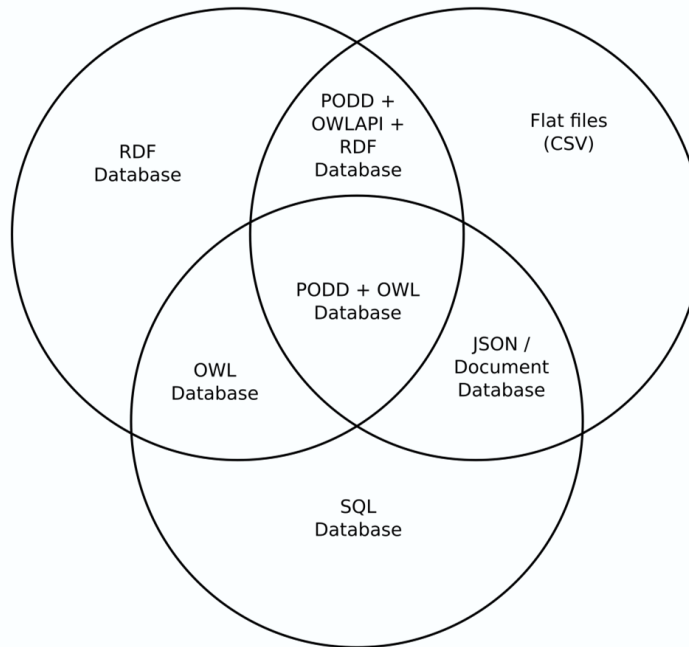
PODD in summary

Quality
Semantic and structural
Linked data both correct and complete

Flexibility
Controlled, optional migration of data
(different physical RDF database instances
for each set of schema being used – unique
to PODD)
Based on OWL Version IRIs and OWL
Imports (import other ontologies and tell
which version to import)

RDF
Database

PODD +
OWLAPI +
RDF
Database

Flat files
(CSV)

PODD + OWL
Database

OWL
Database

JSON /
Document
Database

SQL
Database

Performance
Scales by OWL performance for updating
Scales by SPARQL performance for querying

CSIRO

Acknowledgements

Co-Authors

Adjunct Prof. Robert Furbank
Dr Peter Ansell

Colleagues

Mr Kutila Gunasekera
Dr Pascal Neveu
Mr Alexandre Mairin