Flexible scientific data management for plant phenomics research

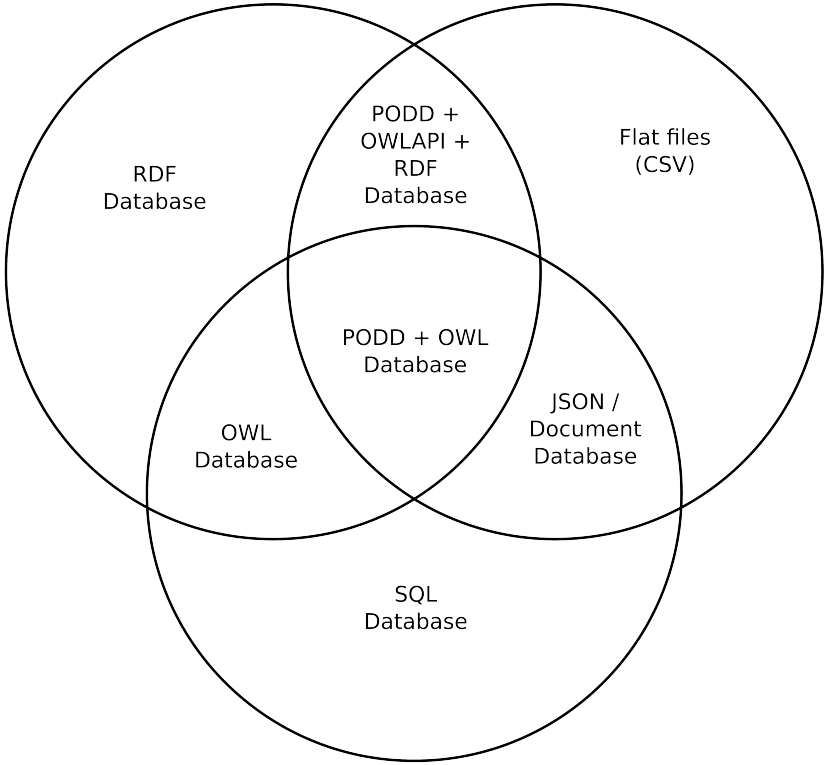
Dr Xavier Sirault

CSIRO Plant Industry
www.csiro.au

Database Management Tradeoffs

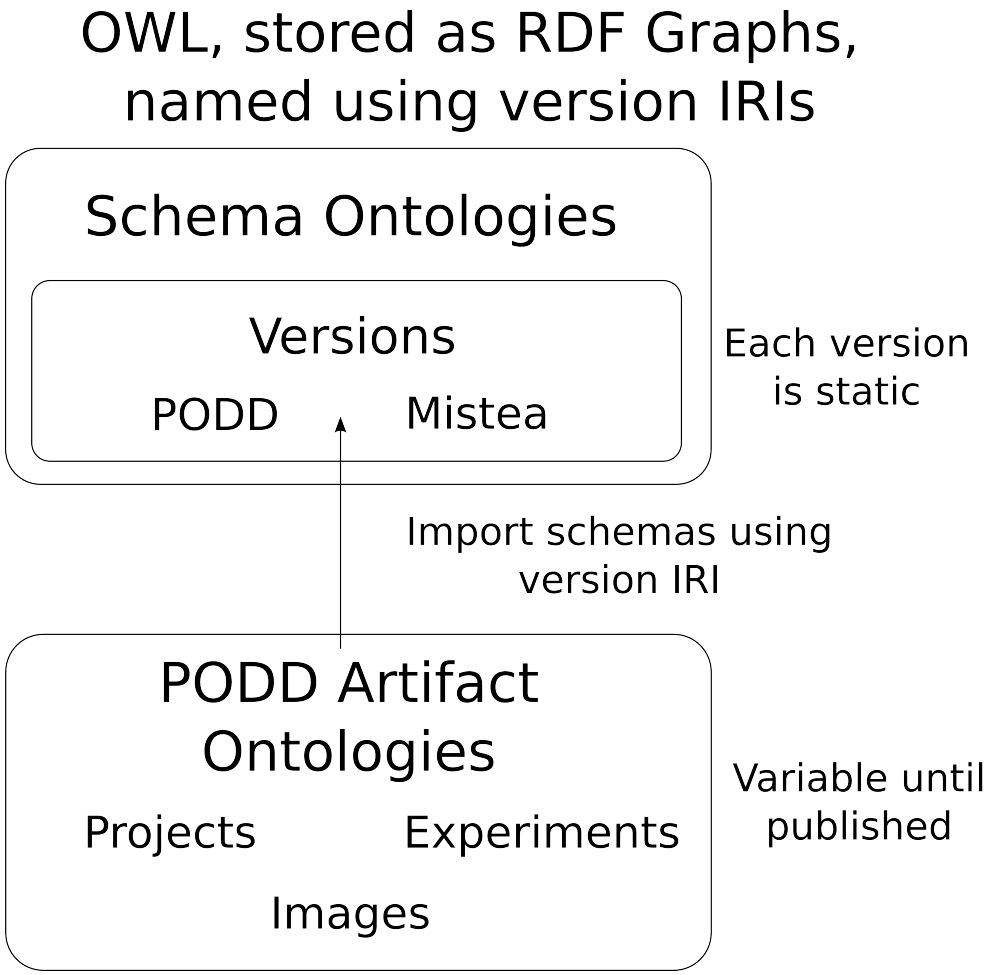
Quality

Flexibility



Performance

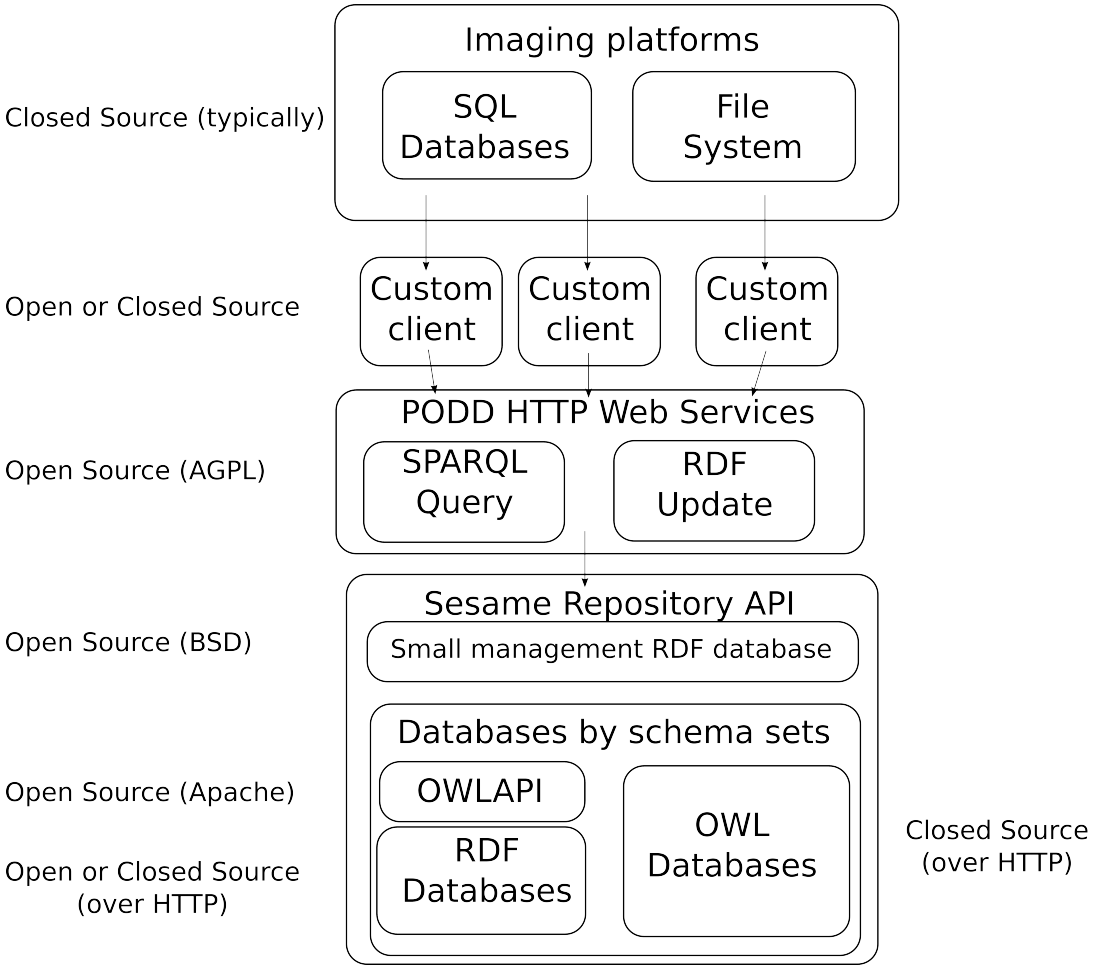
PODD Object Model



PODD Data Management

- Quality
  - Semantic and structural
  - Both correct and complete
- Flexibility
  - Controlled, optional migration of data
  - Based on OWL Version IRIs and OWL Imports
- Performance
  - Scales by OWL performance for updating
  - Scales by SPARQL performance for querying

PODD Workflow



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, and wild type
  - Planting date
- Pots will be linked to more events in future:
  - Watering
  - Client reviews

Conclusion

- Overall goal to allow for federated SPARQL queries across PODD instances to reuse existing plant phenomics data whereever possible
  - Already 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 they are currently using
  - Results would be limited of course by the practicality of the query for the particular schema ontologies in use