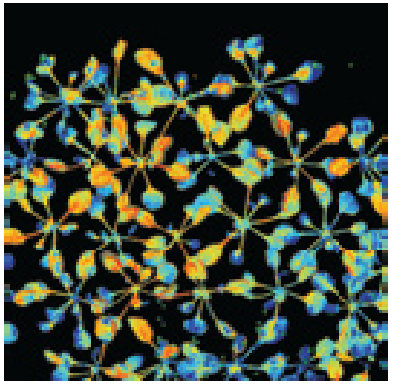
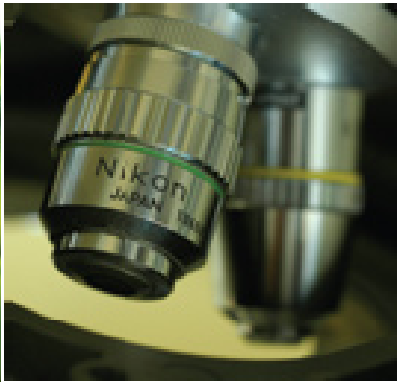
Australian Plant Phenomics Facility

The High Resolution Plant Phenomics Centre



www.plantphenomics.org.au

THE UNIVERSITY OF ADELAIDE AUSTRALIA

CSIRO

ANU THE AUSTRALIAN NATIONAL UNIVERSITY

An Australian Government Initiative
National Collaborative Research Infrastructure Strategy

Government of South Australia

ACPFG

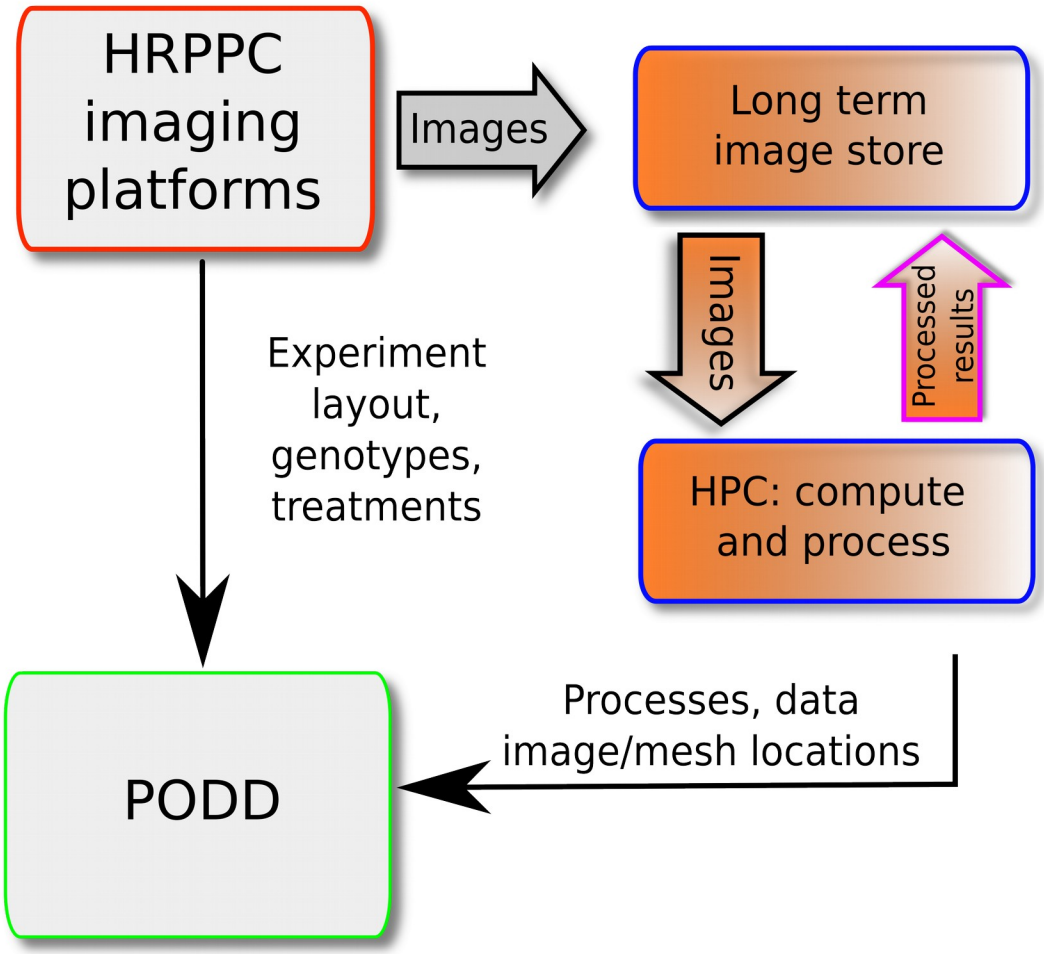
Data acquisition and management

Dr Peter Ansell

Data Scientist, BSc/BBus, BIT (Hons), PhD

Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Overview



Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Data acquisition

- **Experimental layouts agreed on and encoded before experiment begins**
- **Layout uploaded to machine databases and PODD**
- **PODD is our metadata store to enable systematic integration of all of our platforms**
- **Images and other measurements collected regularly using PlantScan, TrayScan, CabScan, Cropatron, etc.**
- **Images stored on disks, specific to each platform, and acquisition metadata initially stored in SQL databases for each platform**

Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Data management and processing

- **Images are batched up and sent to the CSIRO Datastore**
- **CSIRO Datastore is backed up to a long term tape library**
- **Images are processed on the CSIRO Pearcey cluster and in the near future on the NCI node of the Nectar Research Cloud**
- **Results are sent to the CSIRO Datastore and backed up on tape**

Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Data analysis

- **Results are aggregated based on the experimental layout that is found in PODD**
- **Categories include:**
  - **Genotype**
  - **Pot**
  - **Treatment**
  - **Replicate**
- **For each category, the average, maximum, minimum, standard deviation and count are computed for each day where data was acquired**

Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Discussion

- **CSIRO Datastore capacity and a very large allocation to the HRPPC makes it possible to permanently store our very large datasets**
- **Experimenting with new technologies and new species**
- **Experimental layouts submitted to PODD and machine databases must be accurate for analysis to be useful**
- **All species need to be tested in each of the desired platforms before data and image analysts can say whether analysis is likely to be useful**

Australian
Plant Phenomics Facility
The High Resolution Plant Phenomics Centre

Future work

- Further visualisation of data and results using Sapphire, including aggregation based on the experimental layout in PODD
- Integration with genomics through the experimental layout and PODD
- Acquire experimental layouts in a common form from experiments run on all platforms and store them in PODD
- Expand the number of platforms that use the CSIRO Datastore for permanent archival of data
- Expand the number of platforms that use fully automated cluster processing
- Define data standards to enable other scientists to send us their images and experimental layouts for us to attempt to process them, in particular, for Mini PlantScan replicas
- Reuse relevant specifications for dataset metadata, such as the W3C HCLS Dataset Description
- Have an integrated approach to data management and processing