



Bioplatforms Australia

Digitisation Workshop

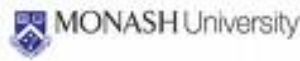
14th April 2015

www.bioplatforms.com

NCRIS

National Research
Infrastructure for Australia

An Australian Government Initiative



National Network International Outreach



BIOPLATFORMS
AUSTRALIA



Genomics

The “Blueprint”

Static Data

Permanent Value

Metabolomics

The “Fuel and Waste”

Dynamic and Volatile Data

Proteomics

The “Machinery”

Mixture of structural and functional

Dynamic Data

Bioinformatics

Embedded Production

Contextual Support

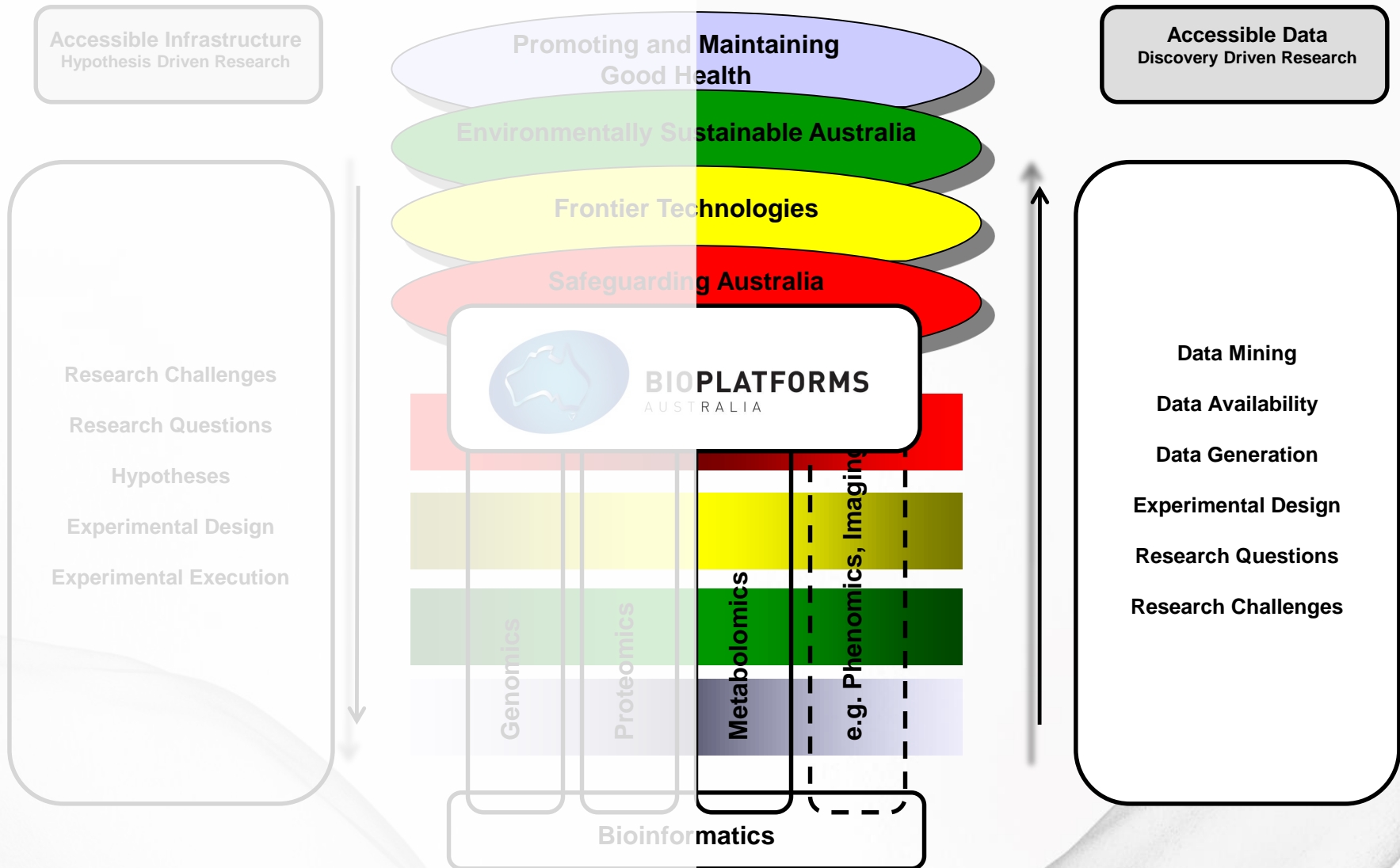
Biological Analysis

Data Management

Data Publication

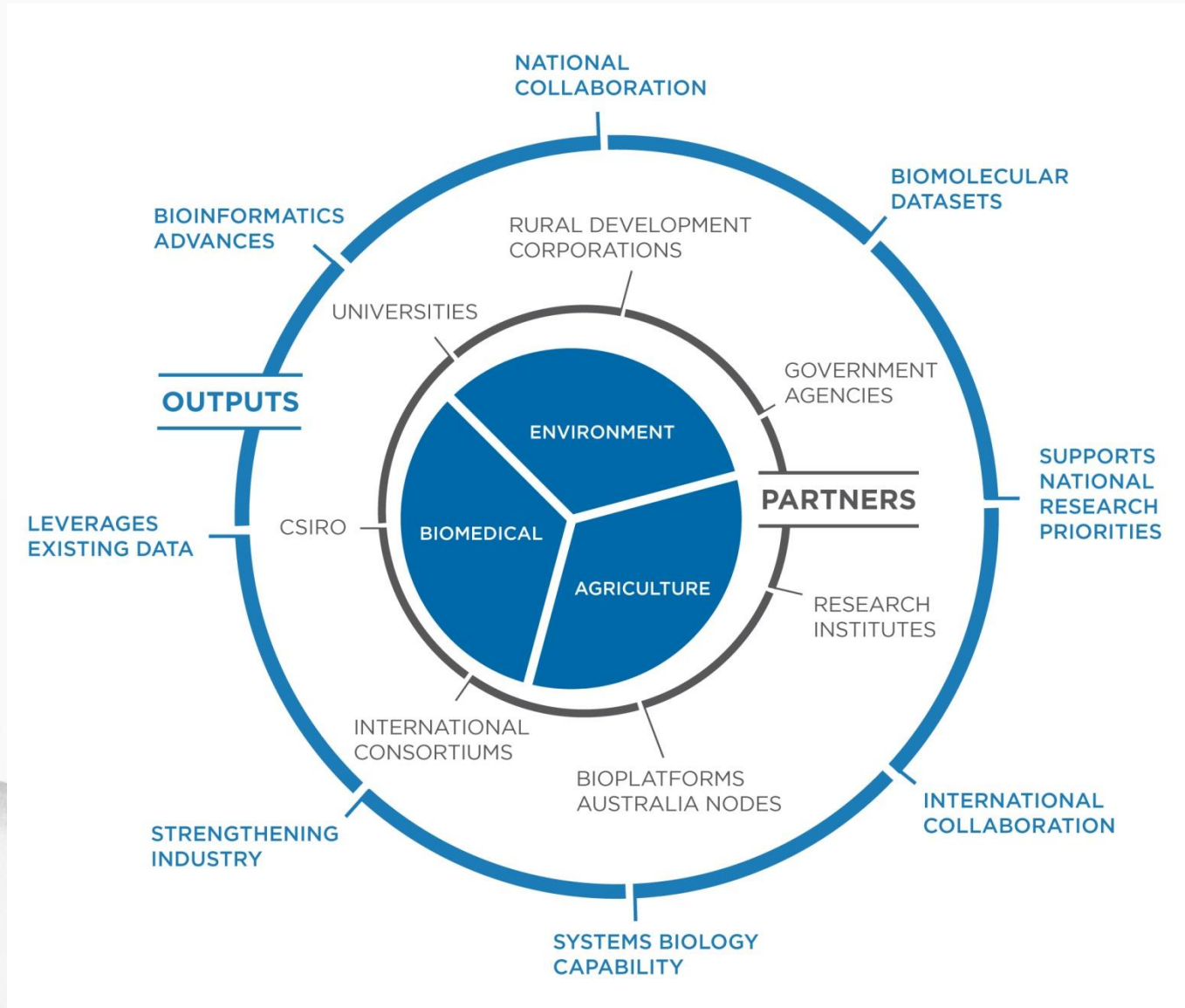
Conduit to eResearch

Hypothesis Driven and Discovery Approaches



- Large and expensive data point but of enduring value
- Molecular data needs contextualising through comprehensive metadata strategies
- Powerful descriptor of structural and functional diversity
- Allows exploitation of collections through an additional dimension
- Price is falling dramatically – now feasible for large scale surveying of new or existing collections
- Data volume and storage considerations
- Data management is key
- Access to data is key

Framework Datasets



- **Wheat reference genome and variety sequencing**
- **Wheat pathogenomics systems biology**
- **Chardonnay Sequencing**
- **Melanoma Genome Project**
- **National Centre for Indigenous Genomics**
- **Great Barrier Reef Foundation ReFUGE**
- **Marine Microbial Diversity**
- **Biome of Australian Soil Environments (BASE)**
- **Rapid Environmental Monitoring (Barcoding)**

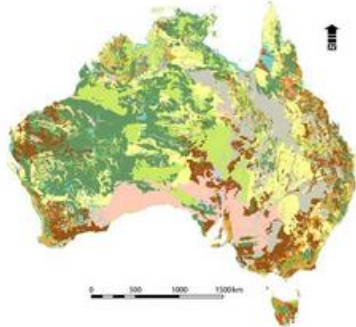
- **Wine Yeast Systems Biology**
- **Yeast 2.0 Synthetic Biology**



Bioplatforms Australia Metadata

Welcome to Bioplatforms Australia's project metadata portal

Bioplatforms Australia Downloads

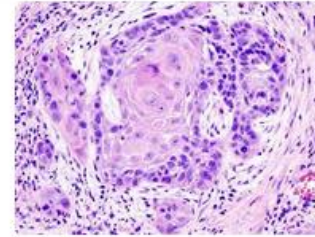


Biome of Australian Soil Environments



Great Barrier Reef

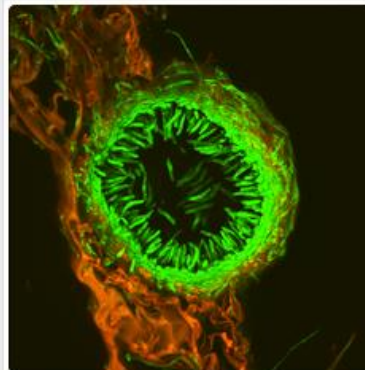
Image courtesy of Gary Cranitch



Melanoma



Wheat Cultivars



Wheat Pathogens Genomes



Continental Survey

- **1200 samples**
- **4 amplicons per site –
fungus, bacteria, archaea,
eukaryote**
- **240 GB**
- **Large scale surveying
very feasible**

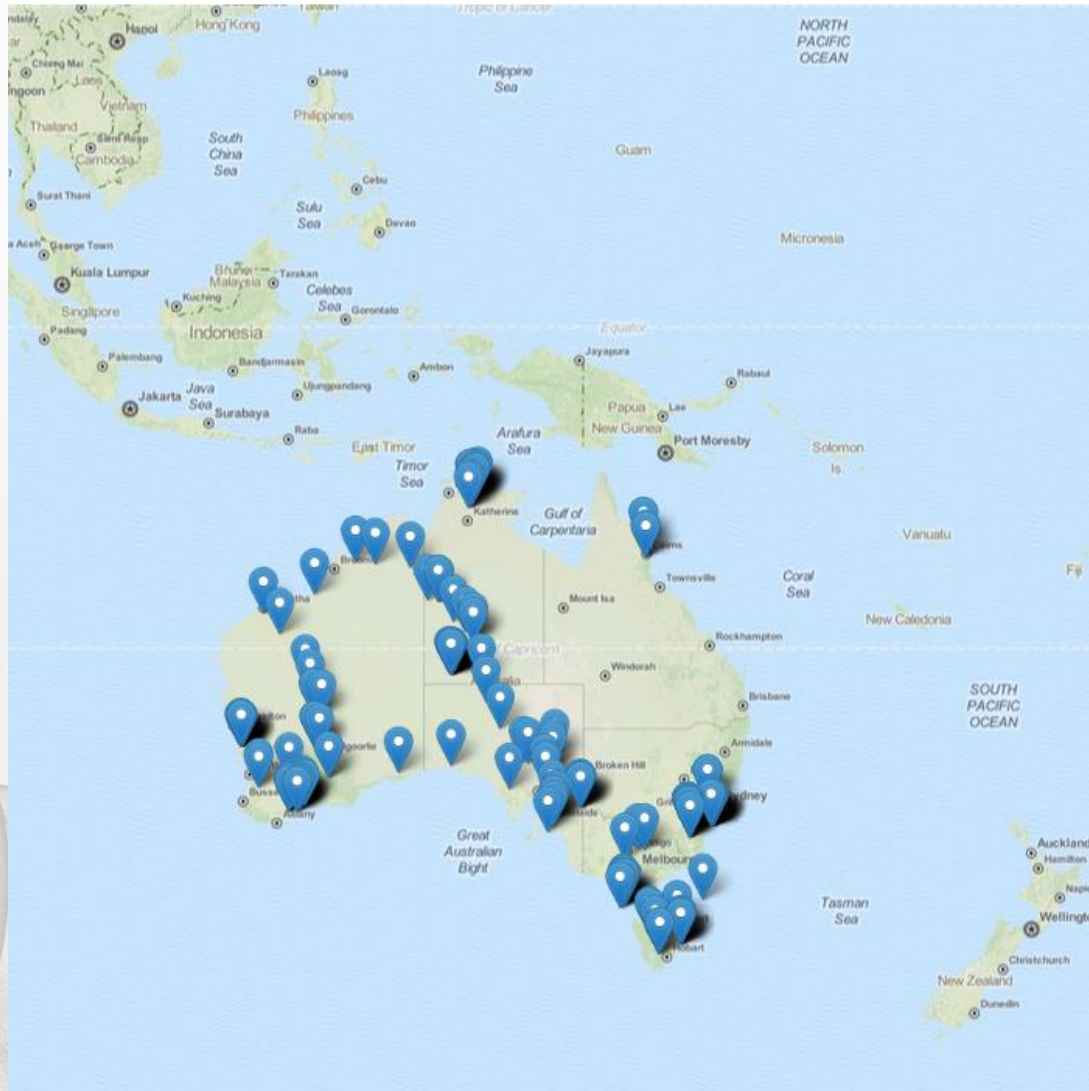
- **1 metagenomic sample
closer to 5GB**

Melanoma

- **136 whole genomes
(~60 patients)**
- **27 TB**

- **200 GB per sample**

Example – Continental Soil Survey



Collection Sites

Filter by name

Location Name	Elevation	Current Land Use
Woomera		Grazing native vegetation
Lake Hart	103	Lake - saline
Cooper Pedy	236	
Booderee	24	National park
Marla	340	Grazing native vegetation
-25.7096333333, 133.24855	501	Grazing native vegetation
Tanami	561	Grazing native vegetation
-20.1876833333, 129.159233333	451	Grazing native vegetation
WestMacdonnell	588	Grazing native vegetation
Lake Lewis	560	Grazing native vegetation
Tanami	634	Grazing native vegetation
Booderee	35	National park
Tanami	514	Residual native cover
Oberon Lake	333	Grazing native vegetation
Oberon	347	Residual native cover
-20.28805, 129.987316667	337	Residual native cover
Panton	347	Grazing native vegetation
Devonian	200	Grazing native vegetation
Derby	26	Residual native cover
Booderee		National park
Wallah Downs	7	Grazing native vegetation
Munni Munni	93	Grazing native vegetation
Hamersley	736	Residual native cover

Example – Continental Soil Survey



Bioplatforms Australia Metadata

Search Form - Search All Samples?

Find samples with All or Any of the matching fields added below:

Search Type	Field	Value
<input checked="" type="radio"/> Single <input type="radio"/> Range	Current land-use ▼	National park

Taxonomic Filters

Level	Value
Kingdom	Bacteria ▼
Phylum	— ▼
Class	— Verrucomicrobia Bacteroidetes
Order	MVS-104 OP3 BRC1
Family	Elusimicrobia OP8 OD1
Genus	Actinobacteria Chlorobi
Species	WS3 GAL15 GOUTA4 OP1 Tenericutes Lentisphaerae Themi AD3

Example – Continental Soil Survey



Order	Flavobacteriales ▾
Family	Blattabacteriaceae ▾
Genus	-- ▾
Species	-- ▾

Search

Results [Export](#)

BPA ID	Sample Context	Chemical Analysis	Amplicon	Metagenomics
102.100.100.7857	Booderee A,O Shrubland	Chemical Analysis	16S 18S ITS	No Data
102.100.100.8163	Uluru Grassland	Chemical Analysis	16S ITS	No Data
102.100.100.8519	Kakadu NP None	Chemical Analysis	16S ITS	No Data
102.100.100.9432	Namadgi NP Marsh/bog	Chemical Analysis	16S ITS	No Data
102.100.100.9452	Namadgi NP Forest	Chemical Analysis	16S ITS	No Data
102.100.100.9463	Namadgi NP Forest	Chemical Analysis	16S ITS	No Data
102.100.100.7923	Mt Lesueur NP Shrubland	Chemical Analysis	16S ITS	No Data

- **Possible – already happening here and overseas**
- **Informative – allowing targeted deeper analysis**
- **Data management will require significant thought and development – storage, tools, compute**
- **Prioritisation of digitisation an inevitable challenge**
 - Types?
 - Strategic selection?
 - Demand driven?
 - Ad hoc digitisation – but harmonised data management and access
- **Needs to be functional**

Andrew Gilbert

0410 538 648

agilbert@bioplatforms.com

www.bioplatforms.com