CSIRO today: a snapshot

Australia’s national science agency
One of the largest & most diverse in the world
6500* staff over 55 locations
Ranked in top 1% in 14 research fields
150* spin-offs based on our IP & expertise
170* active licences of CSIRO innovation
Building national prosperity and wellbeing
**eResearch Architecture Principles**

- **Support the entire science workflow:** not just some areas
- **Research data is an asset:** research data has value and should be accessible, shared and managed accordingly
- **Technology development is enhanced through user involvement:** benefit from the involvement of users in project delivery
- **Ease of use:** enhance productivity and have a common look and feel
- **IT solutions are driven and prioritised by CSIRO’s research support agenda:** enhance research support within CSIRO
- **Maximise computing power available to researchers:** “on-demand” access to computing power
- **Maximise commonality across projects:** identify a core set of services needed by all projects
- **Facilitate collaboration:** both internal and external

**eResearch Supporting Science**
eResearch Planning

• **eResearch Planning Service:**
  - Providing assistance to Flagship Directors and Business Unit Chiefs, and Commercialisation Managers to assist them in creating their eResearch Plans
  - The key issues for CSIRO Management is responding to CSIRO’s Strategic Action 2.4 which calls for to “Embed e-Research and E-Enablement strategies to increase research productivity in a data intensive and collaborative world”
  - Consolidated eResearch Plan for the organisation
  - Supporting proposals to external agencies (e.g. NeCTAR)

• **eResearch Architecture**
  - Defines the range of services, software and systems technologies and infrastructure that are available to scientists
  - Work with research leaders and scientists to define how their work can be supported within the range of services and systems and infrastructure already available or to be made available
  - Identify future requirements for services, software and systems technologies and infrastructure and plan accordingly

Collaboration

• **Two complementary Platforms developed in CSIRO:**
  - CSIRO Collaboration Platform (CCP) – people focussed + data
  - CSIRO Ultra High Resolution Visualisation System (CURVS) – data focussed + people

• **Video Conferencing**
  - Updating to Telepresence
  - CISCO’s Jabber on Desktop

• **Collaboration Tools and Infrastructure**
  - Confluence
  - SharePoint
  - MeetingPlace
  - Australian Access Federation (AAF) (Identity Provider and Service Provider)
Data Collection

• **Electronic Lab Notebooks (ELN)**
  - Completed an evaluation phase which addressed issues such as:
    - Articulation by the business of what material is intended to be managed via an ELN, versus what should be managed elsewhere;
    - Identification of policy, process and governance changes across various domains that would need to be made to support the implementation of an ELN
  - Still working to identify the most tactical and practical solution for an electronic laboratory notebook

• **Data Capture from Science Devices**
  - Supporting the capture and management of data from a variety of science devices
  - Some projects underway and several projects completed

Data Management – Research Data Service

**Targeted Outcomes**

- Enterprise-level data management
- Self-management of data
- Decrease data management costs
- Increase reuse of CSIRO research data
- Increase citation of CSIRO research data
- Increase quantity of data records harvested to ANDS/RDA
- Extensible for various domains (e.g. Microscopy, Sensors, etc.)
Computation

- **World Class Computational Resources**
  - HPC Centres
    - Several HPC Centres prioritising work in climate, earth systems and environmental research as well as radio astronomy, geosciences and life sciences
  - Multi-institutional Facilities
    - NCI Specialised Facility for Bioinformatics and the EMBL EBI Data Mirror in Brisbane
    - MASSIVE NCI Specialised Facility for Imaging & Visualisation in Melbourne
  - Several mid-size HPC systems
    - “Burnet” conventional commodity cluster
    - “Cherax” shared memory server, which uses NUMA (non-uniform memory architecture) technology to provide 512 cores and 4TB memory within a single system image
    - ASC Data Store HSM system, which currently provides two disk and two tape tiers with heavy weighting towards high capacity tape to optimise security, cost and energy efficiency; these will be complemented by adding SDD and MAID (massive array of idle disk) tiers
    - GPU cluster, for applications that can utilise accelerator hardware – currently NVidia “Fermi” processors.

- **Accelerated Computing Service**
  - Accelerating computational science codes by enabling them to run on a broad range of high performance and high throughput computing resources

---

Discovery

- **eTools and Workflow**
  - eTools/Science Apps (Phase 1 complete)
  - Geoportal
  - Pasture Growth Rate Model (Phase 1 complete)

- **Advanced Visualisation Service**
  - Remote visualisation
  - High-Resolution Tiled Displays
  - Visualisation upgrades for Standard PCs
  - Stereoscopic Display Systems
  - Visualisation Software

- **Research Data Service**
  - Data discovery, reuse and management
**Discovery – eTools/Science Apps**

**Potential Benefits**
- CSIRO’s research algorithms are readily discoverable and executable
- There is increased re-use of CSIRO research algorithms
- There is an improvement in the productivity of CSIRO scientists.
- There is increased citation of CSIRO research models and algorithms
- Opens up the possibly for pay-for-service-based access to CSIRO research models and algorithms for external users
- Creates the opportunity for cross-domain collaboration and the emergence of new research areas

**Discovery – GeoPortal**

**Potential Benefits**
- CSIRO’s geospatial research is readily discoverable and reusable
- There is increased reuse of CSIRO’s geospatial research algorithms and models
- There is an improvement in the productivity of CSIRO geospatial scientists.
- There is increased citation of CSIRO geospatial research
- Opens up the possibly for pay-for-service-based access to CSIRO geospatial research models and algorithms for external users
- Creates the opportunity for cross-domain collaboration and the emergence of new research areas
Discovery – Pasture Growth Rate Model

The PGR model

\[ \text{APAR} = \frac{\text{PAR}}{\text{TERRA and AQUA}} \]

\[ \text{LUE} = \text{optimum for best fit to field data based on minimising residuals, formalised into a spatial map only for WA} \]

\[ \text{MI} = \text{index of growth response to soil water availability (Nix, 1981)} \]

\[ \text{TI} = \text{index of growth response to temperature (Nix, 1981)} \]

\[ \text{APARu} (\text{MJ/m}^2/\text{day}) = \text{APAR} \times \text{MI} \times \text{TI} \]

\[ \text{PGRT (kg/ha.day)} = \text{APARu} \times \text{LUE} \]

\[ \text{PGRA (kg/ha.day)} = \text{PGRT} \times 0.67 \]

\[ \text{GAPT (kg/ha.annum)} = \sum_{\text{Jan-Dec}} (\text{PGRT} \times \text{days in month}) \]

Reuseable Algorithms:

1. The Fraction of Absorbed Photosynthetically Active Radiation (fPAR) calculation derived from the Terra and AQUA MODIS NDVI imagery
2. The Photosynthetically Active Radiation (PAR) calculation from the daily solar radiation index
3. The Moisture Index (MI) calculation of plant growth response to soil water availability
4. The Temperature Index (TI) calculation of growth response to temperature

Weekly Climate Information 5km²

\[ \text{APAR} (\text{MJ/m}^2/\text{day}) = \frac{\text{PAR}}{\text{TERRA and AQUA}} \]

Spatial soil water holding capacity map of region

In detail only for WA and SA

Reusable Algorithms:

1. The Fraction of Absorbed Photosynthetically Active Radiation (fPAR) calculation derived from the Terra and AQUA MODIS NDVI imagery
2. The Photosynthetically Active Radiation (PAR) calculation from the daily solar radiation index
3. The Moisture Index (MI) calculation of plant growth response to soil water availability
4. The Temperature Index (TI) calculation of growth response to temperature

Publication

• **ePublish**

  Link publications to data and vice-versa

  ePublish to be made available to external users

• **Research Data Service – Data Access Portal**

  Data Access Portal

  Link publications to data and vice-versa

  ePublish to be made available to external users
eResearch Supports the entire Science Workflow

CSIRO eResearch Architecture

Information Management & Technology
Dr Liam O’Brien
Chief Software Architect
eResearch Program

Phone: +61 2 6214 2962
Email: Liam.OBrien@csiro.au

Questions......