So we have a Digital repository in CSIRO – now what?
Today’s Agenda

• The DAP today
• Challenging times for building new eResearch capability
• “Top down” planning activities – working with CSIRO Flagships
• Recent Global trends in eScience – a few observations from this year’s IEEE Chicago eScience conference
• Responding to the issues – CSIRO eResearch development activities 2012-2013
• Questions
What we built: Data Access Portal (DAP)

- Developed by CSIRO Information Management & Technology (IM&T) and research partners
- Secure repository
- Enables external publication but options for other access levels
- Self serve and enterprise wide
- Options for searching, retrieving and downloading
What’s next?: Further development work

Ongoing system development taking on board user feedback

• Google analytics – carrot for seeing reuse
• Sensor network auto uploads
• Interoperability. Eg. TERN
• Combine/integrate e-Publish (publications approval) and DAP
• Search by location – collections within a geographic area
• Linking to other data repositories within CSIRO, using the Cloud infrastructure currently being developed by IM&T
• Machine to machine functionality: web services, APIs
Challenging times for building new eResearch capability

• NCRIS/Super Science investments winding down – which ones are critical to our science?
• Capital is scarce – How do you prioritise your investments?
• Funding increasing use of current services as Science transitions to a higher use of Computational and Simulation Sciences
• “Free services” create unreasonable expectations of unlimited resources
• Green Computing transition costs
“Top down” planning activities – working with CSIRO Flagships

- **Senior representatives** from all 11 Flagships
- **Awareness of eResearch** as a strategic planning issue is a key challenge
- Adopt an **open access approach** to all planning activities, all team members had a clear expectation of sharing their issues and plans
- **Key support staff** need to be made **available** at any time to provide advise in order to fit in with busy schedules.
- Set clear **planning deadlines**

Two key questions:

- How will demand for our current services change?
- What new services do your need to support your science?
What do flagships have in common?

**Key challenges:**

• Supporting linkages between geographically isolated staff to foster interdisciplinary and cross sectoral approaches.
• Supporting cross-organisational collaboration globally
• Sharing Information repositories internally, with State and Federal jurisdictions as well as globally

**Key trends:**

• Increasing need for openness and transparency in use of data and decision making
• Increasing cost of data use and reuse
• Increasing need to connect models data and people
A Flagship view of the ICT landscape

- Informatics Research Domain Application
  - computational methods
  - spatial and temporal analysis
  - computer applications
  - data acquisition, storage, processing, interchange
  - Visualisation

- Flagship needs
  - NCRIS
  - Research enabler
  - Collaboration
  - Data management and sharing
  - HPC
  - Visualisation

- ICT fundamental R&D
  - Computer Vision
  - Machine Learning
  - Control Systems and signals
  - Sensor Webs

- Networks
  - Optimisation
  - Software Systems

- eResearch Cross cutting research enabling
  - email
  - file server
  - network
  - hardware

- ICT Centre
  - NICTA
  - CSIRO

CSIRO IM&T - eResearch Planning Service | John Morrissey and Liam O'Brien
## Increasing Flagship demand for our current services

<table>
<thead>
<tr>
<th>Service named</th>
<th>Suggested Service enhancements</th>
</tr>
</thead>
<tbody>
<tr>
<td>eResearch Planning Service</td>
<td>Increased engagement at the Theme Level</td>
</tr>
<tr>
<td></td>
<td>Need eResearch awareness program across organisation</td>
</tr>
<tr>
<td>Advanced Collaboration Services</td>
<td>Better external Video Conferencing services</td>
</tr>
<tr>
<td></td>
<td>Improved shared workspace (user-friendly SharePoint for outsiders/cloud storage)</td>
</tr>
<tr>
<td></td>
<td>Maximize use of NECTAR/RDSI services by creating higher awareness for researchers</td>
</tr>
<tr>
<td>Data Collection - Data Capture from Science Devices</td>
<td>Need to investigate opportunities new monitoring systems (e.g. smart traps, dense wind fields)</td>
</tr>
<tr>
<td></td>
<td>Crowd sourcing/Apps</td>
</tr>
</tbody>
</table>
## Increasing Flagship demand for our current services con’t

<table>
<thead>
<tr>
<th>Service named</th>
<th>Suggested Service enhancements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Management - Research Data Service</td>
<td>Assistance with training of Scientists in Managing Data</td>
</tr>
<tr>
<td>Computation – Accelerated Computing Service</td>
<td>Increasing demand needing further staff training</td>
</tr>
<tr>
<td></td>
<td>Transfer of a set of smaller software tools to high performance computers could open new opportunities for the science</td>
</tr>
<tr>
<td>Discovery – Advanced Visualisation Service</td>
<td>Increasing demand for Visualisation and Augmented reality technology</td>
</tr>
<tr>
<td></td>
<td>Remote visualisation – access to high-end 3D</td>
</tr>
<tr>
<td></td>
<td>Stereoscopic kits – for displaying data in true 3D</td>
</tr>
<tr>
<td>Publication – Data Access Portal</td>
<td>Reference datasets and geospatial data-server to improve staff effectiveness, and to serve up data products</td>
</tr>
<tr>
<td></td>
<td>Need assistance in working through Legal assessments for data release</td>
</tr>
<tr>
<td></td>
<td>Support to publish ‘mega’ datasets such as projections of climate change</td>
</tr>
</tbody>
</table>
Flagship demand for new services

• What about working data?
• We need to better link the DAP to external community portals
• Awareness of available eResearch services (both internal and external) a major issue
• Need to reduce the effort required to manage data? Enter data once, use often!
• Advanced ICT Training for researchers
• Linking the DAP to other tools, Spatial Services Information Stack, My Geospatial platform (not yours) and enterprise science management platforms (SAP)
• Not all eResearch tools are run corporately so we need architectural advice on linking systems better
• Help us connect to our clients to maximise impact!
Recent Global trends in eScience – a few observations from this years IEEE eScience conference - Chicago

• The reproducibility of scientific results is becoming a major issue in ensuring that science is defensible both in the research and public domains

• New tools like HUBZero seek to make non-production quality research code more accessible to other researchers by providing a web accessible generic parameter/data input interface and basic visualisation services.

• Keeping workflows is not enough, 95% of workflows in MyExperiment were not able to be rerun due to changes in ICT environments.

• Australia is keeping up with the rest of the world in eScience, in some areas we are leaders
Responding to the issues – CSIRO eResearch development activities 2012-2013

Ramping up our eResearch outreach activities

- Training opportunities for researchers
- Seminars, social media and multimedia campaigns

Organisational Metadata collector

- Making the “points of truth” metadata available to multiple systems
- Enter data once and reuse often
- Initially internal data sources with possible extension to external data sources like ResearcherID

Do we publish software in the DAP?

- What do we have to keep to make science more reproducible?
- Do you archive or keep complete systems active and available?
- Will a virtual platform be re-deployable in the future (5 years from now?)
Responding to the issues – CSIRO eResearch development activities 2012-2013 – con’t

Managing Working data using DAP technologies CSIRO will need to:

- Identify what core functionality researchers and scientists need when they want to manage their working data;
- Identify the various categories of data and which need/should be managed in a system like the DAP;
- Assess our existing tools and technologies to store and manage data already in use across CSIRO;
- Automate metadata gathering as much as possible so managing data doesn’t become a burden for researchers;
- Look at new capabilities like provenance tracking tools in R and NETCDF
- What do commercial platforms like SharePoint offer?
Responding to the issues – Linking the DAP to existing platforms

Discovery layer

Clients (i.e. Discovery Portal, Analysis Workflows)

Community Agreed Services Interface Transfer Standards (information models, file formats)

Exchange layer

Vocabulary Service (SKOS)

Structured Data (Web Feature Services)

Images (Web Map Services)

Coverage Data (Web Coverage Service/DAP)

Service Registry (RIF-CS ISO 19115)

Resources

Persistent ID Service

International/Community Standard Vocabularies

Service Catalogue (Data + Applications)
CSIRO eResearch Program Booth Demonstrations

Data Access Portal
- Interactive web based repository for CSIRO’s research data

Visualisation Service
- Advanced visualisation applications, animations and systems

Workspace Workflow Tool
- Powerful, integrated workflow development framework

CSIRO Collaboration Platform (CCP)
- Promotional movie for sophisticated collaboration platform

CSIRO GPU Cluster
- Promotional movie for Australia’s greenest supercomputer
Questions
Thank you