Workforce Capability Development for eResearch: what skills do we really need?

Lesley Wyborn, Christiana Eze, Nicola Connor and Tony Marks
Outline

1) Setting the context: Why did we do it
2) How did we do it
3) Comparison of eResearch Scientific vs Technical Teams
4) Education and training implications for eResearch
5) Key learnings
Background to the GA eResearch Pilot Project

• In 2011 GA ran an eResearch pilot to asses if there would be business benefits to GA in applying advanced ICT technologies to enhance scientific outcomes

• Done through 3 partnership agreements between GA and
  1. CSIRO to assess an agency-wide eResearch infrastructure (Rob Woodcock)
  2. NCI to trial the use of HPC/HPD (Ben Evans)
  3. ANDS to expose the GA catalogue as a web service (Cynthia Love & Julia Martin)

• Built an internal data access portal with three projects participating:
  1. Geophysics Virtual Laboratory (WFS, WCS, HPC, cloud)
  2. Landsat Satellite delivery as web coverage services (HPD)
  3. Petroleum and GA Catalogues (metadata discovery via CS/W & OAI-PMH)

• eResearch Business Case, then Roadmap that was based on 4 Topics:
  1. High Performance Computing (including cloud)
  2. High Performance Data
  3. Integrative infrastructure
  4. Human Resources
The HR component

• HR component had 2 elements:
  1. Conduct an audit of e-Research Readiness of GA skills
  2. Conduct a gap analysis of GA IM and IT skills and resources

• During the pilot this was difficult to do as:
  1. You had to finalise the 3 projects before you knew who was needed!!!
  2. The elephant eviscerator problem soon emerged

• Without doubt, this was the most difficult and hence underdone component of the pilot project
After completion, General Manager and HR team thought that the eResearch Road Map could potentially be indicating a significant skills change.

HR wanted input to the GA ICT Workforce Plan, based on capability gaps identified during the GA/CSIRO eResearch collaboration project.

Key questions for the new GA 2012-2014 People Strategy:
- Do other projects have the necessary scientific and technical skills?
- Do we need to change some of our Learning and Development programs?
- Who do we really need to target in our Graduate Recruitment Program?
Setting the Scene: we are at a great time of change

• Unless we recognise this we are doomed
• Tall ships failed because they could not get any bigger
• We cannot handle the ‘oceans of data’ we need to recognise we can no longer ‘hand craft’
• In many cases, we are currently asking ‘carpenters’ to become ‘mechanics’ as per during the change from Sail to Steam
WARNING: the study sample is NOT statistically viable

- Twelve scientists participated from GA and 5 technical staff from CSIRO
- This is accepted as not necessarily being a statistically viable sample, but yet the results from both teams were internally consistent
- The team leader profiles will NOT be discussed further: that is just not scientifically valid. Their opinions were NOT included in the results
The capability analysis – how it was done

- Questions were designed by HR and a template was used to generate and document the data gathered
- Part 1 - Each participant was asked to define what they felt were the core elements of eResearch
- Part 2 - Each participant was asked questions relating to:
  1. Their **Qualifications** and how they have applied them at work (knowledge application)
  2. Their **Job Experiences** (that are directly relevant to eResearch functions/achieving results)
  3. Their **Skills/knowledge** (that they apply in eResearch functions)
  4. The **Behavioral attributes** identified as key for eResearch
  5. The **Organisational Support** they felt essential for eResearch
Part 1 - What is this eResearch Stuff?

Both teams defined the core elements of eResearch as:

1. Being directly connected to electronic data and enabling scientists to work directly on the observational data, rather than on syntheses of sub-sampled data
2. Enabling probabilistic analysis with multiple scenarios being investigated and uncertainties being quantified
3. Being characterized by technical innovation and undertaken by those willing to work on cutting edge of problem solving
4. Requiring computers to enable it and humans to drive it
Part 2: Academic Qualifications

GA Science Team

- Ranged from BSc to PhD
- With qualifications in mathematics, applied mathematics, and geophysics with some computational science, modelling and numerical analysis

CSIRO Technical Team

- Ranged from BSc to PhD
- With qualifications in computer science, programming, mathematics, physics and geophysics
- That is, most had some scientific qualifications, but a much stronger focus on computer science/software engineering than in the GA science teams
Part 2: Job Experiences/Skills Knowledge

GA Science Team
- Computational programming skills
- Mathematical skills, statistics
- Spatial skills
- Data analysis, data curation and stewardship
- Transdisciplinary science, thinking big scale

CSIRO Technical Team
- Information systems design and engineering
- Geophysics: understanding the underlying problems GA is addressing
- Ability to weave multiple disparate web services into a coherent application
- Spatial database design and development of spatial information systems
- Engineering, both process and project management
- Extensive experience in developing eResearch tools, portals and technologies
Part 2: Behavioral attributes

GA Science Team

- Intuitive
- Logical
- Non-linear thinker
- Risk taker/willingness to try new things
- Early adopter

CSIRO Technical Team

- Analytical skills
- Logical: problem solving
- Emotional intelligence
- Ability to build teams/teamwork
- All participants listed the ability to communicate and actively listen was the most important attribute for an eResearch expert
Part 2: Organisational support

The identified key enablers by the teams for institutionalising eResearch within the workforce were:

1. Organizational agility
2. An approach to fostering early adopters (aka CSIRO)
3. A specific eResearch enabling team to support the scientists
4. Recognition of the high skills levels of software developers
5. Recognition of the ‘hybrids’
6. CSIRO quote “Our team is built on ex-scientists or reformed software engineers (now scientists) who try to bridge that gap each day”

<table>
<thead>
<tr>
<th>Science</th>
<th>GA Team</th>
<th>CSIRO</th>
<th>IT, IM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>GA</td>
<td>CSIRO</td>
<td>IT, IM</td>
</tr>
</tbody>
</table>
Finally, our findings have significant implications for the education of tomorrow’s scientists and science policy funding. Scientists will need to be completely computationally and mathematically literate, and by 2020 it will simply not be possible to do science without such literacy. This therefore has important implications for education policy right now.
A trip down memory lane...

- **1980’s**
- **1990’s**
- **2000’s**
- **2012**
- **2020**

**Categories:**
- Pure Science
- Hybrid
- Pure IT, IM
Implications for Education Policy

- **1980’s**: Predominantly Pure Science
- **1990’s**: Hybrid
- **2000’s**: Hybrid
- **2012** and **2020**: Predominantly Pure IT

The chart illustrates the evolution from Pure Science to Pure IT over time.
The Education needs

• We need academic courses in:
  – Data science (management, curation, etc)
  – Computational X
  – X-informatics

• In Australia, at best these are at post graduate level

• We need these at undergraduate level urgently

• Note in some universities in the USA, courses in data science are now at undergraduate level and are compulsory
Key Learnings

• The profiling of both teams provided valuable evidence-based information and understandings regarding the skills, qualifications and characteristics required for eResearch projects

• These understandings are already being used inform recruitment, re-assignment and Learning & Development strategic decisions

• eResearch requires a culture of multi-disciplinary teams: system designers, scientists, computer scientists and information management specialists - all working collaboratively together

• The eResearch workforce needs for the future require a scientific workforce that has greater capacity in mathematics, numerical modeling, statistics, computational skills, software engineering and spatial skills, as well as a capability for integration of data across multiple domains (i.e. big picture thinkers)
Thank You and Any Questions?

Lesley Wyborn

Phone: +61 2 6249 9111
Web: www.ga.gov.au
Email: lesley.wyborn@ga.gov.au
Address: Cnr Jerrabomberra Avenue and Hindmarsh Drive, Symonston ACT 2609
Postal Address: GPO Box 378, Canberra ACT 2601