Towards a Global eResearch Service:
Challenges, Benefits & Future Directions

Mike Briers & Fethi Rabhi 27 June 2007

Agenda

- Overview of SIRCA – An eService Model
- Some distinguishing “enabling” features
- Challenges
- Research impacts
- Innovation Partner example: UNSW eResearch
- Discussion

Acknowledgements:

- DEST Systemic Infrastructure Initiative
- Reuters International
About SIRCA

Independent Member-based Research Cooperative
- SIRCA is an independent not-for-profit company established in 1997 to facilitate world-class financial services research on behalf of its (now 30) ANZ University Members
- The SIRCA Board comprises a mix of senior industry and academic representatives
- SIRCA is managed by a full-time CEO (Mike Briers seconded from UNSW) and currently directly employs around 32 professional IT management & support staff
- It’s innovation centre is located in the Sydney CBD (visitors welcome!!)
- ANZ Universities pay an annual Membership fee

Industry – Academic Partnerships
- Founded on strong industry & government data, research & technology partnerships
- Facilitation of industry-linked research (e.g. National PhD program)

Emergent eResearch Service Provider Model
- Extension of ANZ Member service to international university Subscribers
- Over 350 end-users and growing rapidly

Data: Global Historical Event Stream

SIRCA specialises in the management of global historical event stream data encompassing financial market, economic and news events captured in real-time from thousands of sources globally.

Key features of the data repository include:
- It is highly granular – time stamped to the milli-second
- It is non repeatable – the time series of unique events cannot be regenerated
- It is unique and extremely valuable – it exists no where else in the world
- It is the largest of its type in the world (over 200 TB) and growing at 30% per annum (capturing up to 100,000 real-time messages per second).
- End user researchers require on-demand searchable access to the complete data repository
- End user demand is very high and growth will continue into the foreseeable future
SIRCA's eResearch Service Model is End-User Oriented

Four Key eResearch Service Goals:

- Maximise the quality of end-user experiences (better, faster)
- Increase individual and institutional service uptake (more)
- Maintain a pipeline of new service offerings (continuous innovation)
- Create new research agendas and high impact research outcomes (leadership)

Enabled by:

- Independent co-operative not-for-profit company governance & management
- Unique historical global data repositories (financial, economic and news events)
- Secure & resilient storage, test, development and production systems
- Data management and software engineering expertise
- Collaborative end-user and technologist network
- Industry and government partnerships

Enabler: An agreed strategic plan is vital

THE FRAMEWORK

- Vision
- Mission
- Values
- Strategies
- Business objectives, plans
- Key Initiatives
- Action plans, goals and resource requirements
- Metrics, measurements, rewards & recognition

Stable foundation for several years

Need to be regularly reviewed and updated to remain viable and influential
Enabler: Operations are dynamically aligned to service provision

<table>
<thead>
<tr>
<th>Technology &quot;better &amp; faster&quot;</th>
<th>Applications &quot;more&quot;</th>
<th>Innovation &quot;pipeline&quot;</th>
<th>Research &quot;leadership&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computing, network &amp; data</td>
<td>Delivery of products</td>
<td>Collaborative R &amp; D &amp;</td>
<td>Identifying new</td>
</tr>
<tr>
<td>storage facilities</td>
<td>storage &amp; services</td>
<td>IP Management</td>
<td>opportunities for</td>
</tr>
<tr>
<td>Data &amp; software expertise</td>
<td>(on-line &amp; off-line)</td>
<td>External seed funding</td>
<td>end-user research</td>
</tr>
<tr>
<td></td>
<td>Marketing &amp; sales</td>
<td>Data sourcing &amp;</td>
<td>end-user research</td>
</tr>
<tr>
<td></td>
<td>Product support &amp;</td>
<td>acquisition</td>
<td>collaboration</td>
</tr>
<tr>
<td></td>
<td>training</td>
<td>Agile &amp; creative</td>
<td>Industry focused</td>
</tr>
<tr>
<td></td>
<td>User consultation &amp;</td>
<td>development</td>
<td>research and</td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td></td>
<td>consulting</td>
</tr>
<tr>
<td></td>
<td>Service contracting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assets</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing, network &amp; data</td>
<td>Applications</td>
<td>Motivated, creative</td>
<td>Research domain</td>
</tr>
<tr>
<td>storage facilities</td>
<td>storage &amp; services</td>
<td>staff</td>
<td>knowledge</td>
</tr>
<tr>
<td>Data &amp; software expertise</td>
<td>(on-line &amp; off-line)</td>
<td>innovation partner</td>
<td>of academic</td>
</tr>
<tr>
<td></td>
<td>Marketing &amp; sales</td>
<td>network</td>
<td>environment</td>
</tr>
<tr>
<td></td>
<td>Product support &amp;</td>
<td>Distributed (internal &amp;</td>
<td>Industry links</td>
</tr>
<tr>
<td></td>
<td>training</td>
<td>external)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>User consultation &amp;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Service contracting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>To optimise production</td>
<td>To develop end-user</td>
<td>To incubate &amp; maintain</td>
<td>To motivate leading-</td>
</tr>
<tr>
<td>systems and continuously</td>
<td>systems &amp; promote</td>
<td>a pipeline of new</td>
<td>edge research around</td>
</tr>
<tr>
<td>improve application</td>
<td>usage &amp; promote</td>
<td>applications &amp;</td>
<td>new sources of</td>
</tr>
<tr>
<td>support &amp; delivery</td>
<td>usage &amp; promote</td>
<td>products</td>
<td>data &amp; related</td>
</tr>
<tr>
<td></td>
<td>usage &amp; promote</td>
<td></td>
<td>applications &amp;</td>
</tr>
<tr>
<td></td>
<td>usage &amp; promote</td>
<td></td>
<td>products</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competencies</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems design &amp; admin</td>
<td>Marketing &amp; sales</td>
<td>Ideas generation</td>
<td>Research &amp; publication</td>
</tr>
<tr>
<td>Data management</td>
<td>Product training &amp;</td>
<td>Agile development</td>
<td>Research consulting</td>
</tr>
<tr>
<td>Software engineering</td>
<td>support</td>
<td>Fund raising</td>
<td>Relationship management</td>
</tr>
<tr>
<td>Quality assurance &amp; testing</td>
<td>Data consultation</td>
<td>Legal</td>
<td></td>
</tr>
<tr>
<td>Service level management</td>
<td>Business development</td>
<td>Commercialisation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Value is                    |                     |                       |                       |
| created by:                 | Responsive product   | Knowledge sharing,    | Industry-based research|
|                             | development & sales &| generation of         | & consulting with     |
|                             | support             | creative solutions &  | tight feedback        |
|                             |                     | rapid technology       | loop on application    |
|                             |                     | transfer               | development initiatives|

<table>
<thead>
<tr>
<th>Business Goal</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximise the quality of</td>
<td>Increase individual &amp;</td>
<td>Maintain a pipeline of</td>
<td>Create new research</td>
</tr>
<tr>
<td>end-user experiences</td>
<td>institutional service</td>
<td>new applications &amp;</td>
<td>agendas &amp; high</td>
</tr>
<tr>
<td></td>
<td>uptake</td>
<td>products</td>
<td>impact research</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>outcomes</td>
</tr>
</tbody>
</table>

Challenges: Technology Services – “better & faster”

Secure & Resilient Systems
- Security and access control - Shiboleth enabled access control required by JISC (UK)
- Resilience – fully redundant single site (ac3) - second mirror/DR site being implemented
- 24/7 Systems support – remote system monitoring and alerting
- Data latency (was T+90, now T+7, 2007 target is T+1)
- Optimising system database performance – resisted Oracle type solution
- Broad use of open source – initial resistance from Reuters

Software Engineering Processes
- Separating development, test, and production environments - how different?
- Managing multiple application development paths (formal change control & release process)
- Emergent agile (cf “waterfall”) development process – also resisted by Reuters

Quality assurance & testing
- The key to developing production grade applications – but how to resource?
- From developer-based to independent unit, integration & acceptance testing
Challenges: Application Services – “more”

Support & training
- From local to global reach (on-line, access grid, user conferences, podcasting, remote support etc)
- Staying in tune with broadening user disciplinary base – what are their needs?

Marketing, sales & legal
- “Built by academics for academics” … and continuously improved!
- Communicating new content & functionality – create some excitement!
- Leveraging academic networks (introductions, research prizes, keeping in touch)
- Managing data provider and jurisdictional legal requirements (e.g. using channel partners like JISC)
- Co-operate or compete?

User consultation & feedback
- Custom off-line data consultation to automated on-line service (e.g workflow mapping)
- Formal requirements specification, road map and prioritisation

Challenges: Innovation Services – “pipeline”

Innovation Centre
- Idea generation starts where? How far should the innovation network extend?
- Distance from the source of ideas – SIRCA as an independent entity
- Skunk works model? Competitive model?
- IP management & incentives
- Sourcing funding for high risk, long-term projects
- Strengthen and leverage alliances with key data providers
  - from nuisance to value added service provider!
- Commercialisation guidelines – not because we can but only if it benefits core academic e-service
- Collaboration with broader national e-research community!

Managing two cultures
- Production services versus “the sand pit” – control versus creativity
- Gearing for rapid technology transfer – tight feedback loop?
Challenges: Research Services – “leadership”

Shifting and creating new research directions/paradigms (& future “audience”)
- Motivating academics to change research direction – direct support and first-user-advantage principle
- User-builder relationship management
- Industry links – managing expectations
- Project management & research grant support

New Innovation & Research Projects:
- Real-time event stream capture engine – from T+1 to T+0!
- Historical and real-time analytics engine – e.g. compliance and risk management research
- News, economic and market data integration – e.g. combinatorial, data mining research
- Text based semantic modeling tools – e.g. sentiment analysis of news

Research impacts

Academics are becoming vastly more productive:
- Australian academics now regularly publish in the best international journals
- Over 350 ANZ academics and 60 PhD students now use the service
- Research students are more productive and are completing their degree in record time
- Over 500 Research papers have been produced using the service
- Ground breaking research currently in the pipeline limited only by imagination
Research Impacts (con't)

The financial markets research agenda has expanded:
- From traditional market design work to financial crime, corporate governance, forensic accounting, energy and environmental markets, superannuation and aging, complex systems, data mining, computational linguistics

... and become more multi-disciplinary:
- Academics and students now come from base disciplines in finance, economics, accounting, law, mathematics, physics, engineering, informatics, linguistics, computer science etc

... and a large positive impact on research training and education outcomes:
- More and better prepared PhDs
- Improved quality of undergraduate education
**Ad-hoc Data Grid Environments (ADAGE) Project**

- Involves researchers at UNSW, University of Sydney and several European Universities
- Aims of the project: develop tools and techniques for storing, managing and processing ad-hoc data grids
- Ad-hoc data grids
  - Huge repositories of heterogeneous data resources
  - Data has inconsistencies and gaps
- Addressing the needs of a wide range of user communities
- SIRCA’s is providing main case studies for this project
  - Datasets (News and Trading Transactions)
  - User communities (Industry, Reuters, Academic etc.)
- Other datasets will be used (e-health, Web logs)

**ADAGE Project Milestones**

- **Milestone 1**: Analyzing Requirements and Evaluating other Solutions
- **Milestone 2**: Development of Interactive Management Environment
- **Milestone 3**: Development of Grid-based Architecture and Platform
- **Milestone 4**: Case studies and validation
UNSW Centre for e-Research and Services Engineering CeRSE)

- Supports three main research areas
  - Services engineering: software as a service, ubiquitous service computing, service delivery and management
  - Intelligent Information Discovery and Analysis: processing large information repositories, databases, web sites, blogs etc.
  - Transforming Services Access and Sharing: integrating user productivity tools (spreadsheets for business analysts, MATLAB for mathematics etc.) with underlying software environments
- Builds on
  - Service-Oriented Computing Research Group (Faculty of Engineering)
  - FinanceIT Research Group (Faculty of Business)
  - Asia-Pacific ubiquitous Health Centre (APuHC)
  - DEST-funded ADAGE Project, Smart Services CRC

CeRSE Application areas

- CeRSE will have a strong application focus
- Partnership with established domain specific research centres (e.g. SIRCA)
- From e-research services to research e-services

<table>
<thead>
<tr>
<th>Finance</th>
<th>e-Health</th>
<th>e-Government</th>
<th>e-Learning</th>
<th>e-Science</th>
<th>Smart Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transforming Services Access and Sharing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services Engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligent Information Discovery and Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>