



## What Is E-Research?

The term *e-science* has been used to describe large-scale, distributed, collaborative science enabled by the Internet and related technologies. *E-research* is a broader term that includes nonscientific research but that also refers to large-scale, distributed, national, or global collaboration in research. It typically “entails harnessing the capacity of information and communication technology (ICT) systems, particularly the power of high-capacity distributed computing, and the vast distributed storage capacity fuelled by the reducing cost of memory, to study complex problems across the research landscape.”<sup>1</sup>

Within the United States, the term *cyberinfrastructure* has been used to describe the computing and network infrastructure

founded in 1853 and began teaching its first students in 1855. Today, the university enrolls more than 40,000 students in a broad range of professional disciplines. The university has approximately 3,000 academic staff, with about one-third being research-only and the remainder being teaching and research. The university has more than 4,000 research higher-degree students. Total annual research income is in excess of \$200 million AUD, making the university the second-largest research organization within Australia after the CSIRO. It was ranked twenty-second in the world by the *Times Higher Education Supplement Study* in 2004.

The future of astrophysics research is critically dependent on cyberinfrastructure through the University of Melbourne Virtual Observatory. By linking expensive

other digital material relating to cultures of the regions surrounding Australia (and beyond). Established by the University of Melbourne, the University of Sydney, and the Australian National University, PARADISEC contains more than 1,900 records in 254 languages from 39 countries, with more than 800 hours of sound recordings. Rare field-recordings have been salvaged, material that was at risk of being lost, and have been made globally accessible for research and teaching using appropriate international digital data standards and the power of the cyberinfrastructure. Professor Nick Thieberger at the University of Melbourne requires that his research students add their field material to the archive, provide appropriate metadata, and store it in a digitally sustainable format, further building this rich



**The grid plays an important role in the development of e-research, enabling the sharing of data, computational power, and specialized equipment on a global scale.**

that enables research environments such as the “collaboratory, co-laboratory, grid community/network, virtual science community, and e-science community.”<sup>2</sup> The *grid* is a specific instance of cyberinfrastructure: a collection of distributed computing resources (data repositories, specialized scientific equipment, computing power, knowledge services) that appears to users as one virtual system.<sup>3</sup> The grid plays an important role in the development of e-research, enabling the sharing of data, computational power, and specialized equipment on a global scale.

## Why Does E-Research Matter?

E-research<sup>4</sup> provides opportunities to develop whole new areas of valuable research and to see existing research in new ways. Perhaps this is best illustrated by some examples from my own university, the University of Melbourne. An international research and teaching university, the University of Melbourne was

astronomical equipment and providing data mining and curation through grid technologies, the observatory enables astrophysicists to work in large global research teams on terabytes of data. By making this data accessible, the observatory allows some of the stored data to be used up to six times for different pieces of research, often in unexpected ways.<sup>5</sup> A current experiment being conducted by Professor Rachel Webster is one of three running in parallel internationally to examine the low-frequency radio domain for evidence of the origins of the universe (the other two are Harvard and MIT). The Australian research alone involves a six-terabyte link to a remote part of Western Australia and hundreds of terabytes of storage. This research could not be undertaken without the capabilities offered through cyberinfrastructure.

Another example is PARADISEC, a collaborative digital research resource of endangered recorded field material and

research resource.<sup>6</sup> These recordings are now accessible to researchers internationally, allowing new research to occur—research that was not even contemplated when the primary material was collected.

These two very different examples show the power of cyberinfrastructure to provide opportunities for international collaboration in research and to enable new research from existing primary research data, research that was either not possible or not envisaged when the data was initially collected.

## What Are the Opportunities and Challenges?

As more researchers see the possibility of e-research, the middleware becomes critical to an effective research process. *Middleware*, a term used in different ways by different communities, is essentially the software that “provides standard community tools and services for knowledge management, knowledge sharing,

collaboration and interoperability between applications, computing resources, institutions, and individuals across the cyberinfrastructure.”<sup>7</sup> The world is recognizing that good research requires good infrastructure, with many billions of dollars being spent on research infrastructure.<sup>8</sup> Yet despite this investment, using existing resources across the cyberinfrastructure is still time-consuming and laborious. Although the physical infrastructure is well developed, the logical and intellectual infrastructure,

including middleware, is a work-in-progress. Developing effective middleware and solving the technical and organizational issues to enable international collaboration will be critical to success.

A great deal of effort is being directed at middleware development to enable the progression of “production-strength” middleware tools that traverse discipline boundaries and provide broad value to research communities internationally. The aim is to enable collaboration internationally and nationally, to make existing mid-



dleware tools and services more robust and interoperable, to develop new tools where there are gaps, to make accessible large heterogenous research datasets, and to develop a knowledge grid layer to add to the existing computational or data grids—a layer that integrates knowledge management services into the grid environment.

Arguably, however, technology is the easy bit. The harder part is the human dimension. The fundamental issues and questions flowing from e-research relate to organizational issues rather than simply to matters regarding middleware development. If we broadly interpret the Joint Information Systems Committee (JISC) definition of middleware, which uses the term *middleware* to describe the process of helping institutions to connect people to resources,<sup>9</sup> organizational issues are clearly part of that middleware “glue.” The matter of connecting people (researchers) to resources is not only an international issue but also a national, regional, and local issue.

### **What Does E-Research Mean for Information Professionals?**

Linking people to resources—researchers to high-performance computers and

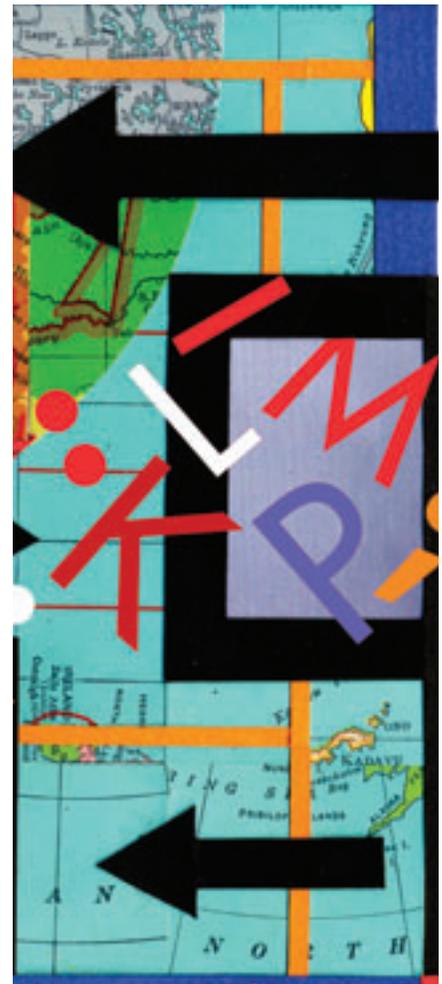
related technologies—has been the domain of the information technology professional in recent times (comparatively speaking). Interestingly, the “e-researcher” readily sees why access to information and communication technology infrastructure, and to the expertise of IT professionals, is required to facilitate e-research endeavors. Yet typically, researchers employ their own IT professionals to develop tools and capabilities (middleware) to meet their specific research project needs, even though they

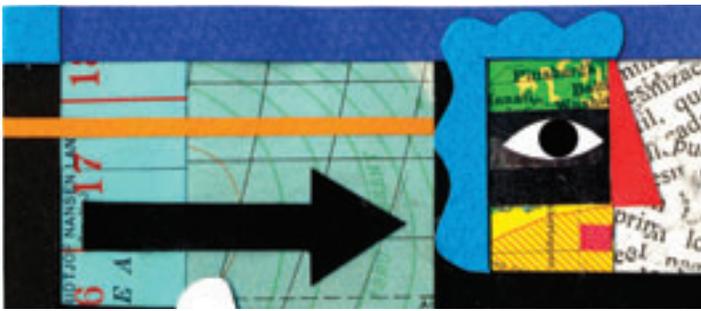
recognize that they will need access to shared infrastructure. Responsive IT units within colleges and universities have been able to develop partnerships with their research communities to provide access to relevant infrastructure—bandwidth, advanced computing infrastructure, grid facilities—and to provide consultancy, advice, and effective linkage points to access national research infrastructure.

On the other hand, linking people to resources—researchers to scholarly materials—has been the role of the librarian

for centuries. Libraries have traditionally been central to the research endeavor, managing and preserving scholarly resources, increasingly in digital form, and making these resources accessible to the researcher, often through collaboration and partnerships with other libraries. Hence, libraries have know-how not only in managing, making accessible, and preserving scholarly resources but also in forming federations and collaborations to share published scholarly work.

But the nature of scholarly communication is changing, with researchers wanting access to primary research data, often in digital form. No longer is scholarly communication a final discrete publication that is to be managed, made accessible, and preserved.<sup>10</sup> Libraries may even risk fading from existence if they don’t respond effectively to the changing environment. In e-research, it is the primary research data that must often be managed, made accessible, and curated. Clifford Lynch argues that the role of libraries will





**Those in the research community have not often recognized the role that librarians could play in providing specialist know-how in managing, preserving, and making accessible the research data.**

shift from primarily acquiring published scholarship to managing scholarship in collaboration with researchers who develop and use this data.<sup>11</sup>

Currently in the majority of existing e-research projects, the researchers, having the domain-specific knowledge, have sought to perform these tasks of managing and making accessible the research data. This data may be generated across multiple countries and across multiple research projects. Many are now realizing that this data is valuable beyond their initial research, which has a limited life. But who will take responsibility for the longer-term curation of and access to this data? Unlike their recognition of the need for IT know-how, those in the research

community have not often recognized the role that librarians could play in providing specialist know-how in managing, preserving, and making accessible the research data. For example, the PARADISEC project mentioned earlier is storing, preserving, and making accessible data of international significance through annual research grants. What happens to this data, and to the infrastructure required for preserving it and making it accessible, when the funds cease?

#### **Human Middleware**

In developing the capability to support e-research, clearly higher education must build partnerships at the international

and the national levels, not only within the research community but also among information professionals seeking to effectively support e-research. Although it is important that these matters be tackled at the national and international levels, there is a good case to “think globally and act locally.” Many of the questions that must be resolved at a national or international level are equally relevant within individual institutions:

- Who will be the champion for e-research support and related initiatives?
- Should we set up centralized or virtual distributed organizations to support e-research?
- How do we share expensive infrastructure and developed expertise to achieve research outcomes?
- How do we get domain-specific research communities to adopt recommended standards?
- How do we leverage what is being learned by domain-specific research communities to the broader research community?

These questions can be effectively addressed only through a strategic approach and strong partnerships between the research community and information professionals. But effective partnerships require an alignment of self-interest among the collaborators.<sup>12</sup> The champion must be able to develop this alignment of self-interest.

At the University of Melbourne, some within the researcher community recognized that collaboration in e-research needed to start at “home.” A working party of interested academics was established in partnership with the Information Division (which covers the library and the information and education technology units of the university). The resulting report provided a catalyst for further discussions with the research



**The information professionals highlighted matters such as “data mining” and “data preservation,” but those in the research communities did not recognize these areas as being beneficial to them.**

community. These discussions enabled those of us in the Information Division to better understand the researchers’ needs and, more important, informed the researchers about the value (the know-how) that the Information Division could bring to their research. Interested information professionals were brought together to discuss e-research and how everyone might work to support e-research. This session provided a great opportunity for people to learn about the breadth of expertise and resources already available to the research community from different parts of the Information Division.

The Information Division is already offering and can offer the following:

- *Technical infrastructure and services:* ICT infrastructure including HPC (high performance computing), networks, data management and storage, repository management, grids, digitization, data mining, statistical support, data preservation, authorization and authentication mechanisms, and help desks
- *Leadership and coordination:* brokering the needs of academics (the “human middleware”), for example by providing connections to needed support structures or expertise on campus; providing a “management” structure for e-research across the university; and collaborating with the other relevant bodies on campus to foster

knowledge and resource sharing across campus communities

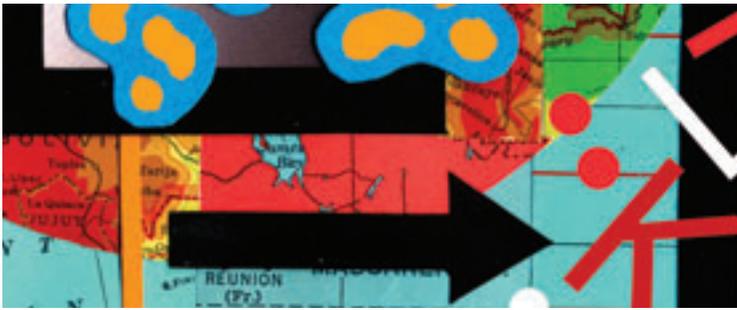
The Information Division interviewed a number of researchers to find out their needs, which were identified as the following:

- *Technical infrastructure and services:* terabytes of disk storage, network upgrades for handling large data volumes and video streaming, access grid nodes, middleware; friendly grid portal, technical support for integration and implementation of databases and software within a virtual organization environment, access to grid programming expertise, help in extending the tools, access to expertise in big databases
- *Leadership and coordination:* need for an e-research champion to bring together a virtual team across the university and beyond, to help researchers understand what is possible with the grid, to assist people in working on their own

Interestingly, the information professionals highlighted matters such as “data mining” and “data preservation,” but those in the research communities did not recognize these areas as being beneficial to them.

There was a close correlation, however, with Markus Buchhorn’s list<sup>13</sup> of researchers’ e-research wants:

- Access to storage and computational resources
- Access to computational software and services
- Videoconferencing and collaboration tools
- User-friendly, application-specific, Web-based portals
- Shared access to large data repositories for searching, replication, and updating



**We must bring our know-how forward and actively engage in strengthening our partnerships with each other and with the researchers within our own institutions.**

- Assistance with organizing and managing research data sets
- Collaborative steering of remote research experiments and the ability to collaborate in international projects

The notable exception is the University of Melbourne research community's recognition of the importance of a champion at an institutional level to assist in achieving the outcomes sought. Surprisingly, this need was echoed in the Information Division's assessment of likely ways that the division could assist—a sure sign that there is a positive way forward, one that recognizes the need for a champion to align self-interest among the collaborators. The Information Division is

currently in the process of creating such a role.

### Conclusion

Research is changing dramatically. It is becoming more multidisciplinary, more collaborative, more global, and more dependent on the capabilities offered through advanced networks and large data storage. These changes provide new opportunities and challenges for information professionals within higher education: we must bring our know-how forward and actively engage in strengthening our partnerships with each other and with the researchers within our own institutions if we are to continue to be a relevant and important part of the re-

search endeavors of our institutions. E-research challenges us to think globally and act locally! *e*

### Notes

1. Australian Research Council, "ARCE-research Support: Invitation for Funding Proposals under ARC Special Research Initiatives for Funding to Commence in 2005," <[http://www.arc.gov.au/pdf/Invitation\\_for\\_Funding\\_Proposals\\_ER05\\_060105.pdf](http://www.arc.gov.au/pdf/Invitation_for_Funding_Proposals_ER05_060105.pdf)>, p. 2.
2. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, January 2003, <[http://www.communitytechnology.org/nsf\\_ci\\_report/report.pdf](http://www.communitytechnology.org/nsf_ci_report/report.pdf)>, p. 6.
3. John O'Callaghan, "APAC's Advanced Computing and Grid Infrastructure for eResearch," presentation to the Australian Synchrotron Meeting, October 25, 2004, <[http://www.apac.edu.au/communication\\_media/publications/synchrotron\\_talk\\_251004.pdf](http://www.apac.edu.au/communication_media/publications/synchrotron_talk_251004.pdf)>.
4. Throughout this article, I will use the term *e-research* as an inclusive term.
5. Rachel Webster and David Barnes, "Future of Astronomy: Enormous Datasets, Massive Computing, Innovative Instrumentation," presentation at the APSR (Australian Partnership for Sustainable Repositories) meeting at the University of Melbourne, March 2005.
6. Nick Thieberger, "PARADISEC," presentation at the APSR meeting at the University of Melbourne, March 2005.
7. Distributed Systems Technology Centre, *E-Research Middleware: The Missing Link in Australia's E-Research Agenda*, discussion whitepaper on e-research middleware for submission to the Commonwealth of Australia DEST National Research Infrastructure Taskforce, March 2004.
8. Australia alone has invested more than \$1 billion. See Evan Arthur, "Government Priorities for Research and Innovation Policy," presentation at the Australian Middleware Forum & CAMP Workshop, Griffith University, Queensland, December 7–11, 2004.
9. JISC Core Middleware Programme, <[http://www.jisc.ac.uk/index.cfm?name=middleware\\_team](http://www.jisc.ac.uk/index.cfm?name=middleware_team)>.
10. For a full description of the changing roles of the research library, see Wendy Pradt Lougee, *Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age* (Washington, D.C.: Council on Library and Information Resources, 2002), <<http://www.clir.org/pubs/reports/pub108/contents.html>>.
11. Lynch quoted in Diane Goldenberg-Hart, "Libraries and Changing Research Practices: A Report of the ARL/CNI Forum on E-Research and Cyberinfrastructure," *ARL Bimonthly Report*, no. 237 (December 2004), <<http://www.arl.org/newsltr/237/cyberinfra.html>>. See also Brad Wheeler, quoted in *ibid.*
12. Daniel Atkins, quoted in *ibid.*
13. Markus Buchhorn, "Research Data Meets 'The Real World,'" (2004), <<http://www.aarnet.edu.au/engineering/middleware/archive/middle/2004/forum/Buchhorn.pps>>.