Scholarly Information Futures: Melbourne School of Engineering 10-year view

Purpose

In response to the Information Futures Commission, this document seeks to answer the question:

> *What scholarly information and technologies, services and infrastructure does the School need[1] to achieve its aspirations over the next decade?*

The Melbourne School of Engineering's core teaching and research heavily depends on access to Scholarly Information and Digital Technologies. The School currently enjoys a very high international reputation for its breadth and depth of research. The continuation of its high reputation heavily depends on the quality of access to these facilities. ICT is one of the most rapidly changing technologies in the world. Huge changes can happen in very short periods in the order of two to three years. The University should not lock into any particular technology by investing large amounts of money as it can hinder the University's ability to take new opportunities/directions. It is worth noting that three of the Departments in the School are leaders in Information and Communication Technologies (ICT) and they regularly contribute new innovations in this field. The University should take advantage of the School's expertise and leadership in Information and Communication Technologies when making decisions.

The School's staff and students rely extensively on Library access for books, journals and conference proceedings. In addition Internet access to various journals and conference proceedings in digital form is necessary for its rapid and convenient access. In fact, some researchers in the School are solely dependant on Internet access for such resources. It is expected that access to hard copies will continue in the next ten years especially for books and old collections. In recent years, access to important data sets is becoming increasingly important for both teaching and research. Traditionally Libraries were not well equipped to provide such resources especially when such data sets are huge and complex in nature. These data sets are usually available in electronic form and can be acquired by paying fees.

Research also excessively depends on making use of some standard tools relevant to their research filed such as Matlab. Such tools should be acquired by the University rather than individual researchers, Departments or Faculties, especially when there are a large number of users in the University. The School of Engineering publishes nearly 1,000 publications each year and these publications should be made easily available both internally and externally subject to copyright. The Researchers in the School produce large volumes of data for scientific purposes which should also be readily available internally and externally. The types of data produced include numerical data, images, audio, video, sensor data as well as interactive three dimensional data. The School's research has recently been expanding from traditional engineering to multidisciplinary research such as Life Sciences. One of the pressing needs for the School is the access to high performance computing and data storage facilities. Many of these needs will be aligned with the Faculty of Science and Faculty of Medicine. It is also envisaged that future publications will become active publications. Active

publications have all modes of embedded information including text, animated graphics, video, audio and executable programs. These embedded programs can be run using implicit input from the publication or explicit input that reader of the publication wishes to provide. Such types of publications require digital technologies. It should be noted that active publications are already available, however, at this stage, scholarly publication do not have the facilities to do this. It is expected that scholarly publications will allow such publications in the near future with the emergence of electronic publication. This will require researchers to learn how to use new tools in order produce such publications. More details are presented in the Appendix.

The following are projections for planning purposes in the next 10 years:

Schools Size:

Category	Number of people
Teaching and Research staff	130 -150
Research only staff	200 - 250
Admin staff	100 - 120
Research students	500 - 550
3+2 Degree students	3500 -4000

Types of Publications	Number of Publications
Journals and Conferences Publications	1300-1400
PhD Theses	90-110
Masters Theses including minor theses	600-800
Lecture materials	400-500

Data storage requirements heavily depend on the type of research a School or a Faculty carries. Storage requirements for large volumes of medical, satellite, 3D rendered images and high definition videos require large amounts of storage and computing facilities to process such information.

Our School has strong research interests in several areas including

- Environment Monitoring and Engineering
- Structural Analysis
- Computational Fluid Mechanics
- Bio Mechanics
- Chemical and Molecular Engineering
- Chemical Process Modelling and Simulation
- 3D modelling and Rendering
- Satellite Imaging and Map building
- Communication Systems Modelling
- Simulation and Design of large scale Communication Systems

- Wireless and Sensor Technologies
- Advanced Signal Processing and Control Systems
- Medical Image Analysis
- Semiconductor Chip Designs
- Nano Materials
- Language Processing
- Information Retrieval
- Bioinformatics
- Brain Modelling
- Bionic Eye
- Large Database Systems
- Data mining Techniques

Most of the researchers working in these areas need to accumulate data both in their raw form as well their processed form for several years and in some instances they will be stored in perpetuity.

The following table estimates the amount of data storage facilities required.

Storage Category	Amount of storage
Teaching	40 Terabytes
Research Students	25 Terabytes
Research Only	15 Terabytes
Teaching and Research	15 Terabytes
Administration related	5 Terabytes
Total	100 Terabytes
Research Related Storage	2 Petabytes
Archival Storage	2 Petabytes
Total Storage needs for the School by 2018	4 Petabytes

*Terabyte = 10^{12} bytes and Petabytes = 10^{15} bytes

It is essential that the School maintains a single repository for the storage of data to allow effective management of information. The School also will require appropriate computational resources which generally fall in the category of high performance computing.

Other issues

There is a concern expressed on the current constraints imposed on the use of VOIP.

It is now possible to access lab equipment remotely by researchers to conduct their experiments. Such facilities require high bandwidth for high definition video and audio transmission. The benefit of such facilities can substantially reduce cost to the University as by removing the need to duplicate such lab facilities.

We also envisage that each researcher (students and staff) needs a powerful computer on their desk which should be replaced or upgraded once every three years. The University also needs

to consider the possibility of rolling out high bandwidth wireless Internet access which can be accessed across the campus by all University Staff, Students and Visiting Researchers.

Appendix

Active Paper

In the future, researchers will not only share their scholarly papers but also make their experimental data and environment available for others to use. This notion of sharing can be achieved through the concept of active papers. We believe in the future, Scholarly on-line publications would be mostly "active" papers, unlike today's passive papers. In an active paper scenario, a researcher can interact with a paper; for example, by changing input values to see its effects. In order to be able to achieve this we must have the availability of appropriate data, software and environment accessible on-demand. The following scenario demonstrates benefits in the discovery process.

A researcher from University-A publishes a new paper on an algorithm for allocation of resources. She has evaluated the effectiveness of her algorithm on a set of computers and tested its application. The produced results are published as an active paper. Now let us say a researcher from University-B reads the paper published by the researcher from University-A. He then develops another algorithm and wants to evaluate his algorithm with the same environment as was performed by the researcher from the University-A for comparison purposes. In order to do comparative analysis, it is very complex (if not impossible) for the University-B researcher to develop an experimental environment similar to the environment available to the researcher at the University-A. Even when it is possible it may take several years to build such an environment and then test his algorithm. With the notion of active papers, all he needs to do is embed his algorithm into the active paper environment. This enables rapid production of new results and substantial savings in costs. In order to provide such sharing environment for active papers requires development of new service oriented ICT infrastructure.

Developing such a technology fosters new collaborations, high research throughput and reliable innovations. The University should take leadership through the Melbourne School of Engineering in developing active papers.