

CyberInfraStructure

MICHAEL RIDLEY

Cyber what? Cyberinfrastructure (CI) is a new concept that describes the highly sophisticated information technology infrastructure needed to support advanced, groundbreaking research. The components of CI are such things as advanced networks, massively distributed computers, enormous storage, visualization tools, remote sensors, and collaboration tools. And people. Librarians. This isn't just a better internet and its got nothing to do with web 2.0. Cyberinfrastructure is about problem solving and nation building in the virtual realm.

In the early days, cyberinfrastructure was called "e-science." This was unfortunate since it is about much more than science, and arguably the biggest impact of CI will be in the humanities. In Canada we tried to call it "networked-enabled environments" (ouch!). The European Union, the Australians, and the Americans have a significant head start in creating a national CI. Aside from Alberta, which has started to build a provincial CI, Canada is largely missing in action on this one. And it is something we need be concerned about.

What are the components of CI? For the most part CI is comprised of fairly cool but admittedly geeky stuff: very fast computers (grids, clusters, supercomputers), high-speed and highcapacity networks (fibre optic lightpaths, for example), lots and lots of data storage (petabytes, maybe exabytes), remote sensors and instrumentation (for

environmental monitoring, climate data, cosmology), visualization and interpretation tools (that analyze and present data in different ways), and advanced collaboration tools (that enable people to engage and contribute from around the world).

One of the technology layers of CI is often called middleware. Middleware is the tools and services that connect all the other layers (hardware, software, networks, people). Much of the innovation and insight comes from the middleware. Without trying to push this metaphor too far, it seems to me that libraries and librarians are the ultimate middleware. And that is why I see them as the missing ingredients in the evolution of CI.

Another aspect of the middleware concept can be seen in the research organizations that have been created to leverage the capabilities of cyberinfrastructure. These are Virtual Research Organizations (VROs) that have no physical manifestation but rather are massively distributed organizations linked by technology. VROs are fueled by information, particularly the specialized information set in the context of the problem being solved. These VROs are not being well served by research libraries although the members of these organizations are typically university faculty and graduate students. Instead they are developing their own capacities and expertise. Librarians and their invaluable expertise need to find their way into these VROs.

A catalyst for building CI is the availability of data on the network. Lots of data – massive amounts of data. The Large Hadron Collider at CERN didn't cause a black hole when it started up but it did initiate a cataclysmic flow of data – more than 15 million gigabytes a year. How will we understand this much data? We won't. CI tools will monitor, analyze, and recommend answers or interpretations. The operative phrase is "correlation is enough." It will be sufficient proof to see in the data patterns evidence to suggest an answer. As a result, getting data modeling and data interpretation right will be a critical function. This suggests that librarians must branch out from their focus on texts and documents (an honourable pursuit but no longer fully sufficient) to specialize in the arcane arts of data, numbers, statistics, numerical transformations, and just plain complex math. Numbers not letters.

One of the reasons Canada is struggling with CI is the lack of national focus. In the U.S. the National Science Foundation has taken the lead, and there are similar organizations in the United Kingdom, the European Union, and Australia. In Canada there is no equivalent central agency. In typical Canadian fashion, we will have to build a coalition of organizations to advance this agenda. Groups like CANARIE (the national advanced network organization), CUCCIO (the Canadian university CIOs), CRKN (the Canadian Research Knowledge Network), CARL (the research libraries), and Compute Canada (the high-performance computing group) will have to find the common ground and shared determination to build a strategy and coordinate the implementation. The failure of these groups to cooperate will profoundly hurt Canada's capacity to participate in the 21st century world of research and development.

Cyberinfrastructure may be just a new buzz word to you now. However, it represents a significant challenge to our country and an important opportunity for our profession. Librarians are the missing components of the CI model. Perhaps the Ontario Library Association can, as it has done so many times before with other initiatives, take a leading role in this innovation.

[Michael Ridley](#)

is the Chief Information Officer (CIO) and Chief Librarian at the University of Guelph.