Federal Support for Institutional Cyberinfrastructure: Challenges and Opportunities

Prepared by Robert J. Robbins

Introduction

Today biomedical research requires access to a substantial array of wet-lab equipment and facilities, ranging from the exotic (Fourier transform ion cyclotron resonance mass spectrometers) to the mundane (glass washers and autoclaves).

In a large research institution, access to these instruments is often provided through different mechanisms (e.g., individual lab purchases, shared resources, etc.), which can be conceptually arranged into four logical layers:

- *research superstructure* — where specific research is funded and accomplished,
- *top-level infrastructure* — shared special purpose equipment, such DNA sequencers or mass spectrometers,
- *mid-level infrastructure* — shared general purpose equipment that is needed for many different kinds of research, and
- *deep infrastructure* — buildings and other large-scale facilities involving construction, remodeling, and fixed equipment.

These levels are presented in Figure 1.

Equipment might be made available in any level in the figure. For example,

- simple centrifuges and other routine laboratory gear might occur as in-lab equipment associated with an individual research project,
- mass spectrometers, confocal microscopes, flow cytometers, or Solexa sequencers might occur as specialized equipment within a shared user core, or within a particular research laboratory,
- autoclaves or dishwashers might be present as general equipment either in a user core, or in a separate common core, and
- specialized building construction or large heating and cooling equipment would be provided as deep institutional infrastructure.

Similarly, biomedical research requires access to a substantial array of IT equipment and facilities, ranging from the exotic (high-performance supercomputers) to the mundane (email and desk-top computers).

As with laboratory equipment, access to these IT resources is often provided through different mechanisms, which can be conceptually arranged into the same four logical layers as with wet-lab support. Information technology might be made available in any level. For example,

- desktop computers and routine analytical software might occur as in-lab equipment associated with an individual research project,

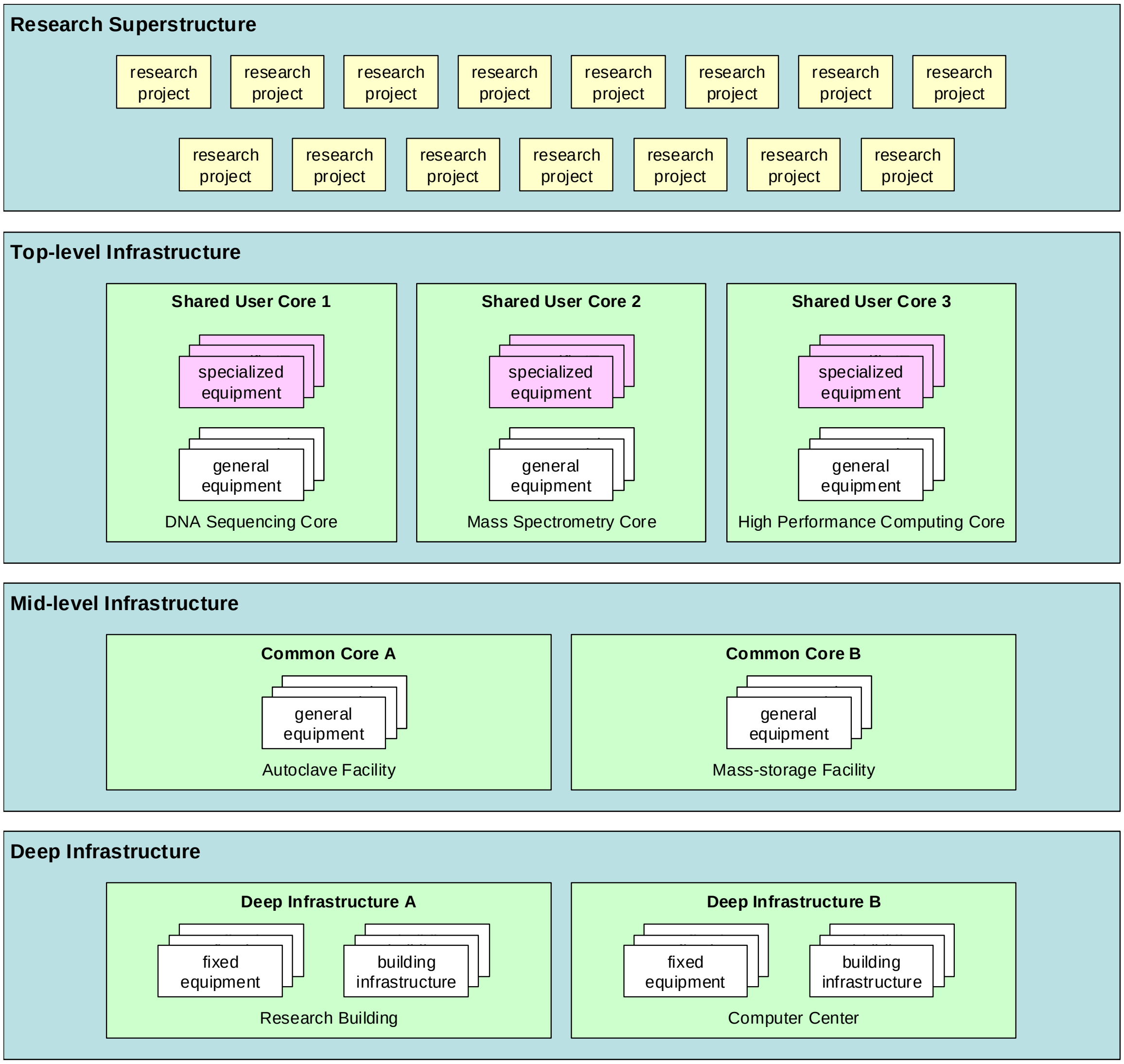

Figure 1. The relationships among research projects, shared core facilities, and institutional infrastructure in a typical, large research organization. Researchers within individual research projects would depend upon access to dedicated instruments within their labs and to shared instruments outside their labs. Specialized instruments are often provided within dedicated core facilities that might be organized around a particular class of instruments (e.g., a mass spectrometer shared resource) or around fields of study (e.g., a genomics shared resource that might offer access to sequencers, synthesizers, expression-arrays, etc). Researchers may also rely upon shared general-purpose equipment that is either provided as part of a specialized core or in a dedicated general purpose core, such as glassware or a cold-room facility. Underneath all of this, the deep institutional infrastructure provides the large-scale shared equipment, such HVAC systems, needed for all activities in the organization.

- high-performance supercomputers or expression-array analysis tools or Solexa data-rectification tools might occur as specialized equipment or services within a shared user core,
- simple mass storage or generic data-base support or campus networking or cybersecurity support might occur as general equipment either in a user core, or as separate common core services, and
- dedicated, specialized computer support such as specialized data centers or in-wall fiber networking might be provided as deep institutional infrastructure.

SUPPORT FOR RESEARCH AND INFRASTRUCTURE

Generally, activities in these various levels receive support, or are eligible to receive support, by mechanisms designed for that particular level. For example, the research superstructure level is supported by a variety of mechanisms designed with the individual investigator and investigator-initiated research in mind. Various construction support mechanisms offer assistance in the development of deep scientific infrastructure. Shared instrumentation programs provide support for top-level infrastructure. And, some mechanisms do exist to provide support for mid-level infrastructure but this is the level that frequently encounters the most difficulty in securing external funding.

Most funding agencies or programs have a mission to support a particular area of research. For example, the biology directorate at NSF supports non-medical life-science research in the United States and, more specifically, genetic research is supported by the genes and genomes systems program in the division of molecular and cellular biosciences. The division of biological infrastructure provides support for shared specialized equipment through its instrumentation program.

The NIH is similarly divided into sub organizations that cater to the needs of different scientific communities and it also has a unit — The National Center for Research Resources — that provides support for infrastructure activities.

A major challenge in infrastructure support derives from the goal of most units in most funding agencies to provide support, either at a superstructure or in infrastructure level, *only* to members of their research community. By its nature, infrastructure can support more than one activity and therefore many funding agencies require applicants to provide evidence that externally funded shared infrastructure will be used only by members of that agency's funding community. If this is not the case, the agency may agree only to provide support that is proportional to the usage by “their” community.

Because special-purpose instruments are by definition specialized, it is often the case that a shared special-purpose instrument will in fact only be used to support the appropriate community. However, general-purpose equipment may be used to support a much wider range of activities and this can make it difficult for institutions to obtain external support for mid-level infrastructure. As we will see below, this problem is especially acute for mid-level cyberinfrastructure.

Although building construction might be seen as the most general purpose of all infrastructure activities, many agencies do have programs for providing some amount of construction support, but usually this is constrained by the stated in designed purposes of the building.

THE CHALLENGE OF CYBERINFRASTRUCTURE

Most computers can, at some level, be used to address any computational task. For that reason, obtaining support for computational instruments can be especially challenging, especially with regard to documenting the fact that the shared computational instrument will be used only by the appropriate research community. With top-level instruments, such as shared high-performance compute clusters, this is often done by stipulating that the particular high-performance cluster will be used to meet the needs of a particular research program. Similarly, on occasion institutions have been able to acquire support for even more general-purpose computational instrumentation, such as large-scale mass storage, provided that the mass storage is intended to be used for a particular purpose, such as managing the high volumes of data being generated by, for example, a large-scale sequencing facility.

At the moment, however, some of the biggest institutional challenges in supporting biological and biomedical research involve the need to acquire truly large-scale (and very generalized) equipment in the area of mid-level infrastructure. There are no federally funded infrastructure programs that specifically support mid-level cyberinfra-structure and many programs seem instead to explicitly exclude such needs from their programs.

Mid-level Cyberinfrastructure Needs are Great

The Bio-IT World Expo is an international meeting with the goal of providing "the perfect venue to share information and discuss enabling technologies" that support biological and biomedical research. Chris Dagdigian, a founding partner of BioTeam[1], has been invited several times to give keynote presentations at Bio-IT meetings, where his goal has been to describe *Trends from the Trenches* — descriptions of emerging and pressing IT needs in biomedical research.

In his 2009 presentation[2], Dagdigian identified several areas of IT that present great opportunities or major challenges (or both): (1) the deployment of virtual server farms, (2) developing a federated approach to institutional storage and backups, (3) building "green IT" systems, (4) scaling out, using multiple, identical systems to accomplish cost-effectiveness, (5) utility (cloud) computing, and (6) better approaches to data protection (e.g., dual parity RAID as a requirement). Every one of these issues involves improvements in mid-level cyberinfrastructure. Nowhere in his presentation did Dagdigian even suggest that he was seeing access to high-performance computing "instruments" as rate-limiting for biomedical research. Furthermore, asserted that building data-storage islands to support specific research needs (i.e., storage as an "instrument" in top-level infrastructure) was a problem, not a solution (slide 59):

> Storage "islands" have always been an issue in our field

---

1 BioTeam is a consulting firm that provides IT solutions for biomedical research. See http://www.bioteam.net

2 http://blog.bioteam.net/wp-content/uploads/2009/04/bioitworld-2009-keynote-cdagdigian.pdf

- Made worse by lab-local large storage and (future) cloud storage
- Bad for scientists (lots of rsync & wasted productivity)
- Bad for IT (duplicated content multiplies backup & operational hassles)

Based on our experience at the Fred Hutchinson Cancer Research Center and on interactions with IT leaders at many other institutions, I believe that Dagdigian's analysis is spot on. I would go on to add that he missed a few other areas that pose current challenges at many institutions, such as the need to upgrade basic network infrastructure to support much higher bandwidth and the need to make substantial improvements in cybersecurity to protect information assets.

According to the BioTeam analysis and many research CIOs, the biggest institutional needs in biomedical cyberinfrastructure are all in the area of mid-level cyberinfrastructure — the area where federal support is virtually non-existent.

ARRA Support Falls Short

As part of the government's American Recovery and Reinvestment Act (ARRA), NSF developed two major competitions to provide infrastructure support: (1) Academic Research Infrastructure – Recovery and Reinvestment (ARI-R^2) and (2) Major Research Instrumentation – Recovery and Reinvestment (MRI-R^2). Neither of these offer support for mid-level cyberinfrastructure.

According to the slides[3] from a 28 May NSF webcast to describe these programs, the ARI-R^2 program specifically excludes computers ("No computers or data storage systems" – slide 9) and the MRI-R^2 program rules out general-purpose equipment ("The MRI program will not support requests for general purpose equipment, including general purpose computers." – slide 18).

NIH has also issued RFAs for ARRA-supported infrastructure and these also show a similar tendency to avoid supporting mid-level cyberinfrastructure. The NIH shared-instrumentation grant (SIG) and high-end-instrumentation grant (S10) programs rule out general-purpose equipment. The NIH Extramural Research Facilities Improvement Program (C06) supports efforts to make major alterations and renovations to existing buildings, but rules out non-fixed equipment. The NIH Core Facility Renovation, Repair, and Improvement (G20) announcement contains wording that could be interpreted as allowing support for mid-level cyberinfrastructure, but a personal communication from the program officer indicates that such a proposal would not be viewed favorably if too high (more than 30%) a proportion of the budget went for equipment.

Mid-level Cyberinfrastructure is TOO Useful

A particular problem in finding external support for mid-level cyberinfrastructure is that such infrastructure is *too* useful. Good mid-level cyberinfrastructure can be used for almost any purpose. For example, a truly federated approach to large-scale mass storage could be used to meet the needs of almost all individual research programs, of any high data volume shared instruments, and of most administrative departments. If usage purity (that is, usage only by members of the particular research community supported by the

3 http://www.tvworldwide.com/events/nsf/090528/globe_show/ppt/ari_mri_slides_for_webcast.ppt

funding agency or program) is a requirement for external support of cyberinfrastructure, then it is easy to see why such support is so hard to find.

Opportunities

The fundamental assertion of this position paper is that, for most biomedical research institutions, external support for mid-level cyberinfrastructure is desperately needed and scarcely available.

If we assume that this is true, then meeting those needs offers a tremendous opportunity for federal spending to simultaneously stimulate the economy and deliver critically important and valuable new infrastructure.

How could this be done? Providing such support will be difficult if not impossible, so long as a requirement for usage purity remains in place. Many funding agencies and programs believe, with good reason, that it would be inappropriate for them to support activities in other areas, making it nearly impossible for them to support mid-level cyberinfrastructure.

This has not always been the case, however. In the 1980s, before the Internet had emerged as a popular phenomenon, NSF supported (1) the transition of ARPANET into NSFNET, (2) the development of mid-level regional networks, and (3) the access by research institutions to NSFNET via regional networks such as NYSERNET or SURANET. In those days, when NSF helped an institution to become "wired" by connecting to a regional network, NSF did not insist that only NSF funded computational researchers be allowed to use this new infrastructure. In fact, the opposite was the case. NSF insisted that such a shared resource be made available to all academic departments at the institution.

Toward the end of the 1980s, NSF led a second transition, moving NSFnet from a university-only phenomenon into the public-private partnership that ultimately became the modern Internet phenomenon. This government-funded infrastructure-development activity, carried out without regard for the research focus of the users of the infrastructure, transformed academia and transformed the world.

The opportunity is ripe for another such transformational investment. Information technology is beginning to mature into a multifaceted infrastructure that supports, in some way, nearly all academic activities universities and research organizations. Most of these organizations now have a cyberinfrastructure that consists of a multitude of islands, operating in varying degrees of isolation. This less-than-optimal situation is limiting the benefits that academic organizations can get from their IT infrastructure.

In the 1980s and 1990s, NSF transformed academia and the world with its wide-ranging and forward-thinking support for large-scale interconnectivity. Now in the 21st century, another transformational opportunity awaits. If academic organizations have the wherewithal to make substantial and coordinated upgrades to their mid-level cyberinfrastructure, the effect upon these institutions' operating efficiency could be astounding. Some institutions, such as Clemson University in South Carolina, have already determined that major investments in their mid-level cyberinfrastructure are critical for their long-term strategic competitiveness and success. To that end, Clemson has recruited new IT leadership and has given them access to significant resources with which to build the new IT infrastructure. To date, the Clemson program has been making good progress, especially in areas such as large-scale federated mass storage.

The current stimulus dollars have already been allocated, with almost no detectable support for mid-level cyberinfrastructure. If some funding agencies could make even modest adjustments in the allocation of these funds to include some support for mid-level cyberinfrastructure, the benefits would be large. If, in the next year or two, there happen to be another major infusion of stimulus funding, it would be incredibly valuable if NSF were given the opportunity to demonstrate that it could once again transform academia by transforming academic cyberinfrastructure.

The potential benefits are great, and not just in academia. By first generalizing ARPANET into NSFnet, and then on into the Internet, NSF had a hugely stimulating effect in the US economy in total, not just in academia. If NSF were given the opportunity to inject substantial federal funding into mid-level cyberinfrastructure across academia, the benefits would serve as a model for the economy as a whole. Productivity at academic institutions could be expected to spike, driven in part by the breakdown in the local IT fiefdoms that would occur with the availability of funding only for large-scale, coordinated IT infrastructure.