Cornell University Cyberinfrastructure Plan - 2015

Providing critical, highly valued and cost-effective infrastructures for information technology is a key objective in Excellence in Organizational Stewardship, a core tenet of the Cornell University Strategic Plan. As stated, information technology is one of the "enabling conditions" for the five goal areas of faculty excellence; educational excellence; excellence in research, scholarship, and creativity; excellence in public engagement; and staff excellence. The IT@Cornell Strategic Plan (2013-2017) serves to identify the IT objectives and initiatives necessary for success in the university's goal areas. This Cyberinfrastructure plan comprises the critical elements and deliverables associated specifically with research excellence.

Research: IT supporting research in all disciplines
The issues facing IT support of research at Cornell are similar to those found at other leading universities. Our challenge is to identify the infrastructure, services, and support structures that can provide the most value in the areas of greatest need to the Cornell research community. Currently Cornell Information Technologies, CIT, provides the essential foundational IT infrastructure used by all of campus. Depending upon their expertise and requirements, faculty and researchers obtain specialized support and services from centers, institutes and core facilities reporting to the Office of the Vice Provost for Research.

Sustainability. In a very dynamic environment, determining how to best provision optimal research-oriented IT services and support to the campus is an ongoing quest. Cornell's centralized research core facilities, including the Cornell Center for Advanced Computing (CAC), are required to recover a significant portion – up to 80% – of their direct costs from user fees that are approved annually by the Cornell Division of Financial Affairs. In other words we are allocating central support to the most urgent and very diverse needs of our researchers presents challenges. The rapid emergence of IT services "in the cloud" (provided over the commodity Internet or research networks by third parties) will soon become a practical alternative to on-premises IT infrastructure, and will require us to think differently about how scale can reduce overhead costs while increasing, for example, computational and data-storage capacity and access.

Objectives. Although this plan focuses on information technology, we should aspire to create a comprehensive research infrastructure at Cornell, not only addressing IT infrastructure needs but also acknowledging a broader context of policies, funding, business planning, IP issues, curation services, trust and federation, grant writing, and government mandates – going beyond big data to include management of research workflows, metadata creation, discoverability, training, new dissemination forms, publishing models, etc.

The objectives identified here serve to support the objectives and actions in the Excellence in Research, Scholarship, and Creativity goal of the Cornell University Strategic Plan.

Objective 1: Improve campus effectiveness in addressing the requirements of "Big Data" by leveraging existing mechanisms and, where appropriate, creating new means of digital scholarly content creation, reuse, access, and preservation.

More and more, research is dependent on collection and analysis of voluminous structured and unstructured data sets coming from multiple sources over periods of time. Providing IT support for big data has quickly become a significant issue at Cornell.

Initiatives
    a. Plan for and provide coordinated institutional stewardship of research data and preservation of the scholarly record.
    b. Participate in – and where possible, lead – inter-institutional projects that aggregate resources beyond the campus level to help preserve the scholarly record (e.g., DuraCloud [http://www.duracloud.org] and the Digital Preservation Network, DPN [http://www.dpn.org]).
    c. Support and augment on-campus groups, such as the Research Data Management Service Group (RDMSG), that provide support services directly to researchers.
    d. Maintain, invest in, and continue to enhance the campus data network so that the Cornell research communities in all disciplines have access to increasingly massive digital data resources.
    e. Lead and collaborate with other institutions and with regional and national organizations (e.g.,NYSERNet [http://www.nysernet.org] and Internet2 [http://www.internet2.edu]) so that researchers have the best possible access to remote data sources. Consider related issues of policy, governance, and licensing.
    f. Promote collaboration among all of Cornell's campuses to create a seamless research "experience" for all disciplines.

Objective 2: Enable easy access by researchers in all disciplines to infrastructure, tools, services, and support.

Faculty should not encounter technical and other obstacles when accessing IT tools and services necessary for their research activities. Enabling easy access to IT furthers the research enterprise.

Initiatives
    a. Make all appropriate Cornell systems and services accessible to our own researchers and those from other institutions with whom we wish to collaborate by providing trusted federated identity access and management services (such as those supported by the InCommon Federation).
    b. Simplify and streamline software distribution and licensing processes so that researchers get the most value from these licenses while spending the least possible time managing them.
    c. Explore emerging opportunities to maximize the benefits of research-related IT investments through cloud services (including but not limited to a growing suite of services brokered by Internet2's Net+ division), Red Cloud (Cornell's cloud computing platform), and other appropriate means (such as bursting to public cloud providers or participating in academic and public agency cloud federations).
    d. Extend Administrative Streamlining Program (ASP) efficiencies from administrative areas to address similar needs in research domains. (This includes Infrastructure Virtualization, IT Service Management, and perhaps Application Streamlining.)

Objective 3: Improve administrative support systems and services for researchers, particularly as they relate to the efficient and effective management of grants.

Research administration is often seen as an obstacle to carrying out research activities. Improved systems and services in this area can be of significant benefit to researchers.

Initiatives
a. Implement Kuali Coeus, a comprehensive research administrative system.
b. Plan for and deliver improvements to other research administrative systems that are beyond the functional scope of Kuali Coeus.
c. Optimize the administrative processes that researchers must use and make those processes as consistent as possible across the university.
d. Explore ways in which appropriate IT expertise can be of more help with research grant preparation.

Objective 4: Develop a coherent cyberinfrastructure for research according to the recommendations of the NSF Advisory Committee for Cyberinfrastructure.

Cyberinfrastructure is the totality of networks, computers, instruments, services, support structure, and community that faculty and researchers require. The 2011 report by the NSF Advisory Committee for Cyberinfrastructure (ACCI) Task force on Campus Bridging makes the recommendation to university leaders that, "Every institution of higher education should have a strategic plan […] for the establishment of a coherent cyberinfrastructure."

Initiatives
a. Leverage foundational infrastructure and services provided by CIT such as networking, backups, and InCommon Federation support.
b. Provide Cornell faculty, researchers and their collaborators leading edge core facilities offering essential services and resources that benefit from economies of scale and scope to optimize research funding.
c. Pursue opportunities to lead and partner with other institutions in the delivery, support, education and training efforts of national cyberinfrastructure resources.
d. Leverage national cyberinfrastructure resources such as those offered by federal funding agencies like NSF, NIH and DOE.
e. Ensure infrastructure and services are in place to enable seamless bridging of access to and from Cornell and other external cyberinfrastructure resources.
f. Track and adopt industry trends such as the adoption of public clouds when it provides a strategic cost and/or performance advantage.
g. Leverage the value of and participate in leadership of national and regional research networks and collocation resources. (NYSERnet & Internet2)

The Cornell University Center for Advanced Computing (http://www.cac.cornell.edu) plays a significant role in these initiatives while ensuring there is communication and coordination between the Office of the CIO and Vice President for Information Technologies and the Office of the Vice Provost for Research to deliver resources and services in the most cost effective way.

Located on the Ithaca, NY campus of Cornell University, CAC serves the computational and data analysis needs of the Cornell community and their collaborators from science and engineering and the humanities. Examples are astronomers searching for pulsars, biologists analyzing gene sequences, cardiothoracic

surgeons reporting and researching treatment outcomes, physicists analyzing high energy collisions, social scientists mining census data, and psychologists studying language acquisition in children.

CAC serves the national community as a partner on the NSF-funded Extreme Science and Engineering Discovery Environment (XSEDE). The current CAC director is the XSEDE Architecture and Design Coordinator. CAC staff consultants develop and deliver XSEDE training programs and software distributions. CAC's director is also an active member and past Chair of the Coalition for Academic Scientific Computation.

CAC provides high-performance computing and cloud computing services configured with a variety of software including Red Hat Linux, CentOS, Eucalyptus, Hadoop, SQL Server, and MySQL. A large-capacity DataDirect Networks storage system and Dell EqualLogic SANs are available for standard and archival storage. CAC configured and operates a public cloud computing service called Red Cloud, http://www.cac.cornell.edu/redcloud, that offers unique features desired by researchers, i.e., dedicated cores/RAM and no charge for network traffic.

Intra-campus network and related resources
The Cornell campus wired network has gone through several upgrades with the more recent Network Connectivity Project (NCP) designed to bring 1G and select 10G data speeds to the edge ports provide enhanced campus-wide WiFi, VoIP services, and a standardized the cable plant within the buildings, http://www.it.cornell.edu/about/projects/ncp/. The uplinks from building switches are upgraded as needed to support the consolidated data flow into 9 node routers that connect to a redundant backbone core. The backbone supports 40G of cross-campus traffic and is the next target for upgrade as usage increases.

The campus WiFi growth remains exponential in number of devices carried by students, in the faculty use in the classroom, and in research use for robots, Google glass, or gathering mobile data. The 4500+ access points connect to 10 controllers and access the redundant backbone via the 9 node routers. A project to upgrade more than 1000 access points is scheduled to complete by June 2015.

The backbone core has two ISP connections to two separate core routers to maintain network connectivity. They were recently upgraded to 10G burstable to support the growing Internet traffic.

The core backbone is extended via fiber to NYC for a 10th node router to support the Cornell Tech Campus, Cornell offices in NYC and our peering connection with the Weill Cornell Medical College.

Cornell Information Technologies is responsible for the operation and maintenance of the campus network and is responsible for the design and support of new network initiatives such as the open science network.

The College of Engineering/Computing and Information Science IT with help from CIT have installed a Software-Defined Network (SDN) in Gates Hall that enables researchers to access the campus network and the SDN without impacting the production network. SDNs use a concept called open flow to control where data is sent providing benefits such as control of bandwidth and better quality of service. Research groups developing this control software need access to an SDN for experimentation. Cornell will continue to investigate how SDN can provide benefits over time.

Regional and National CI Investment

Cornell recently upgraded to a 10G Internet2 connection through the NYSERnet regional R&E network to accommodate the research traffic that used to travel the NLR backbone. This year an instaGENI node was installed in one of the core datacenters to connect Cornell to the national GENI network. Cornell is also a member of and uses Eduroam as the secure Wi-Fi network on campus and for ease of travel for students and faculty.

IPv6 Deployment

Cornell has worked with our ISP providers to announce IPv6 addresses so traffic can flow though either pipe. IPv6 has been inserted into our core network and node routers and down into the two core datacenters. This year we extended into our ESX infrastructure and into our VM guests so we can investigate and build network monitoring, management and security practices for IPv6. We also have an IPv6 DNS zone and respond with IPv6 AAAA records. Planned work is to test IPv6 on our test wireless network and on our load balancers and servers.

Trust and Identity: InCommon Federation / TIER

Cornell is a member of the Internet2 TIER initiative [InCommon Federation](#) and participates in InCommon steering, which is a group of 400+ Higher Education institutions (including Cornell, Columbia, Stanford, Ohio State and many others) and Service Providers (including Microsoft, EBSCO, OCLC, and JSTOR) that trust each other's authentication systems (NetIDs). Cornell also has its own federation that includes the Ithaca campus NetIDs, Weill Cornell CWIDs, and soon the Qatar and NYC Tech campus userIDs. CIT maintains an Identity Provider that can be used with local or vendor applications that would like to do Single Sign On (SSO) with a NetID. This method of authentication is currently used mainly for integrating with vendor applications such as Remedy, Box.com, Qualtrics, Illiad, EBSCO, JSTOR, and Edublogs. See also Shibboleth (https://www.it.cornell.edu/services/shibboleth/about.cfm).

Summary

This cyberinfrastructure plan will guide investments, priorities, and resource allocation in support of Cornell University's research mission, and is an important element of the overall IT@Cornell Strategic Plan. The objectives and initiatives described here will be tracked over time, with regular reporting on progress, challenges, and the successful achievement of deliverables.