CC* Networking Infrastructure:
High Performance Research Data Infrastructure at the American Museum of Natural History
AMNH CAMPUS INFRASTRUCTURE PLAN (January 2018)

Introduction

The American Museum of Natural History (AMNH) is uniquely positioned at the nexus of research, education, and public outreach. Our mission to "discover, interpret, and disseminate—through scientific research and education—knowledge about human cultures, the natural world, and the universe" defines our collective focus as an institution. For nearly 150 years, AMNH has been a recognized leader in innovative scientific research and the dissemination of that knowledge to the public and educators the world over. With the opening of the Richard Gilder Graduate School (RGGS) at AMNH in 2006, which includes a Ph.D. granting program accredited by the Board of Regents of the State of New York, AMNH is developing the next generation of scientific leaders.

In recent years, AMNH has made a series of strategic technology infrastructure investments, including:

- Upgraded enterprise core network switches, now capable of supporting 10Gbps connections with a pathway to 40Gbps in the future. An upgrade to the core network switches and routers to
- Installation of new enterprise network firewalls deployed in a high-availability configuration.
- Deployment of more than 350 wireless access points providing 802.11ac (Wave1) wireless throughout the campus, with 802.11ac Wave2 wireless access in select areas. These wireless access points are managed via a pair of central wireless controllers.
- Installation of datacenter network upgrades supporting 10Gbps and 40Gbps network connections.

In addition, AMNH is developing projects to further extend and expand our technology capabilities, including:

- The deployment of central High-Performance Computing (HPC) facilities with the goal of adding 5,000-10,000 cores of computational CPU power in addition to our existing HPC systems.
- Cybersecurity enhancements including enhanced web application firewall (WAF) capabilities and next generation host and network-based intrusion detection/ prevention.
- Expanded network and cybersecurity monitoring capabilities, including performance monitoring enhancements, and expanded log correlation and analysis capabilities.
- Datacenter facilities upgrades to support additional cooling capacity to support Museum HPC and research initiatives as well as to provide redundant cooling capability in the event of a failure of the primary HVAC system.

Cognizant that technology continues to play a key part in scientific research, education, and public outreach, the Information Technology Department at the American Museum of Natural History has developed a series of strategic plans focused on modernizing and facilitating the use of technology to further our institutional research mission. These efforts are focused on the following key areas:

1. Outreach to Constituents

The development of an effective and responsive campus infrastructure designed to support the needs of our researchers, educators, staff, and the public requires AMNH IT to actively engage and become partners with our constituents. By becoming intimately aware of their needs and active participants in their success, AMNH IT is able to focus on delivering impactful technology solutions that best suit the needs of our users.

To this end, AMNH IT actively seeks out opportunities to provide expert assistance in the areas of technology and data management to our research and education community, and in doing so better understand their needs and plan accordingly. AMNH IT regularly engages with our science and research teams through various internal meetings and working groups and has purposely stepped up its own outreach efforts in order to develop these important collaborative relationships. Most recently, AMNH IT has begun facilities preparations for a proposed 5,000 core HPC cluster as well as proof of concept tests of HPC "cloudbursting" with commercial providers such as Amazon Web Services (AWS) and Microsoft Azure. Future directions being considered include the use of distributed high throughput computing using OpenScienceGrid.

2. Expansion of Development of High Speed Networking

Scientific research at AMNH is no longer confined to the Museum campus. Our research is national and global in nature, and it is critical that our cyberinfrastructure also reaches beyond our walls. Therefore, AMNH plans to connect to our regional research and education network provider, NYSERNet, with a dedicated network connection that will link AMNH to the higher education community in New York State and other R&E networks that NYSERNet peers with such as ORION (Ontario, Canada Research & Innovation Optical Network), CA*net (Canadian National Research & Education Network), ESnet (US Department of Energy, Energy Sciences Network), MAX (Mid-Atlantic Crossroads), TWAREN (Taiwan Advanced Research & Education Network), NoX (Northern Crossroads – New England), and MAGPI (Philadelphia GigaPoP). Initially, this connection will be made at 10Gbps but is being designed to grow as demand and new applications require. NYSERNet will also serve as a gateway to Internet 2, a connection AMNH seeks to implement in order to provide access to other research and higher education facilities nationally and globally.

To make full use of these connections, the software and hardware of the AMNH network that serve our research and education users will need to be upgraded to drive faster connections with network partners around the world. In 2015, the network core was upgraded to support 10Gbps connections to various network IDFs throughout the campus with the capability to scale to 40-100Gbps if needed. Building on this work, AMNH will deploy new distribution and access layer switches capable of driving 10Gbps to desktops and lab systems that require this throughput to support optimized connectivity. Additionally, AMNH seeks to expand high-speed wireless into labs, classrooms, other research and education facilities to support the many users who use wireless as their primary connection method.

While these upgrades will improve the overall user experience and connectivity capabilities for our research and education community, the ability to facilitate large scale data transfers will require the use of specialized networking methods designed specifically for this purpose. Therefore, AMNH also seeks to build an enhanced throughput capacity via a “Science-DMZ”, a purpose-built high-speed network designed to facilitate high-volume bulk data transfer, remote experiment control, and data visualization. The Museum’s Science DMZ, to be called AMNH-SciNET, will be deployed alongside the existing campus network and will interconnect the scientific research departments, labs, imaging and visualization facilities (including the AMNH Microscopy and Imaging Facility (MIF) and the Hayden Planetarium), High Performance Computing (HPC) systems and visualization clusters. To facilitate bulk file transfers between AMNH and collaborating institutions, a series of Data Transfer Nodes (DTNs) will be deployed within the AMNH-SciNet, using the FIONA architecture developed at UCSD.

AMNH IT has also begun the implementation of IPv6 throughout the Museum’s technology infrastructure. To that end, we have obtained an IPv6 allocation from ARIN (2620:0:2840::/48) which will be deployed in a phased project beginning in 2018. Starting with the core network infrastructure (border routers, core network switches, firewalls, DNS servers, etc.), IPv6 addressing will be extended to advanced systems (HPC clusters, research specific networks, Science-DMZs, R&E WAN connections), before being deployed in a dual-stack IPv4/IPv6 configuration on departmental research systems to ensure a seamless transition from IPv4 to IPv6. Once IPv6 has been firmly established on research-associated systems and networks, it will be extended throughout the campus in 2019-2020.

In order to ensure that network performance is maintained throughout the campus and that the efficacy of upgrades and enhancements can be measured, AMNH will integrate the PerfSONAR infrastructure into both the AMNH-SciNET and the campus network to measure end-to end performance monitoring.

3. Cybersecurity and Privacy

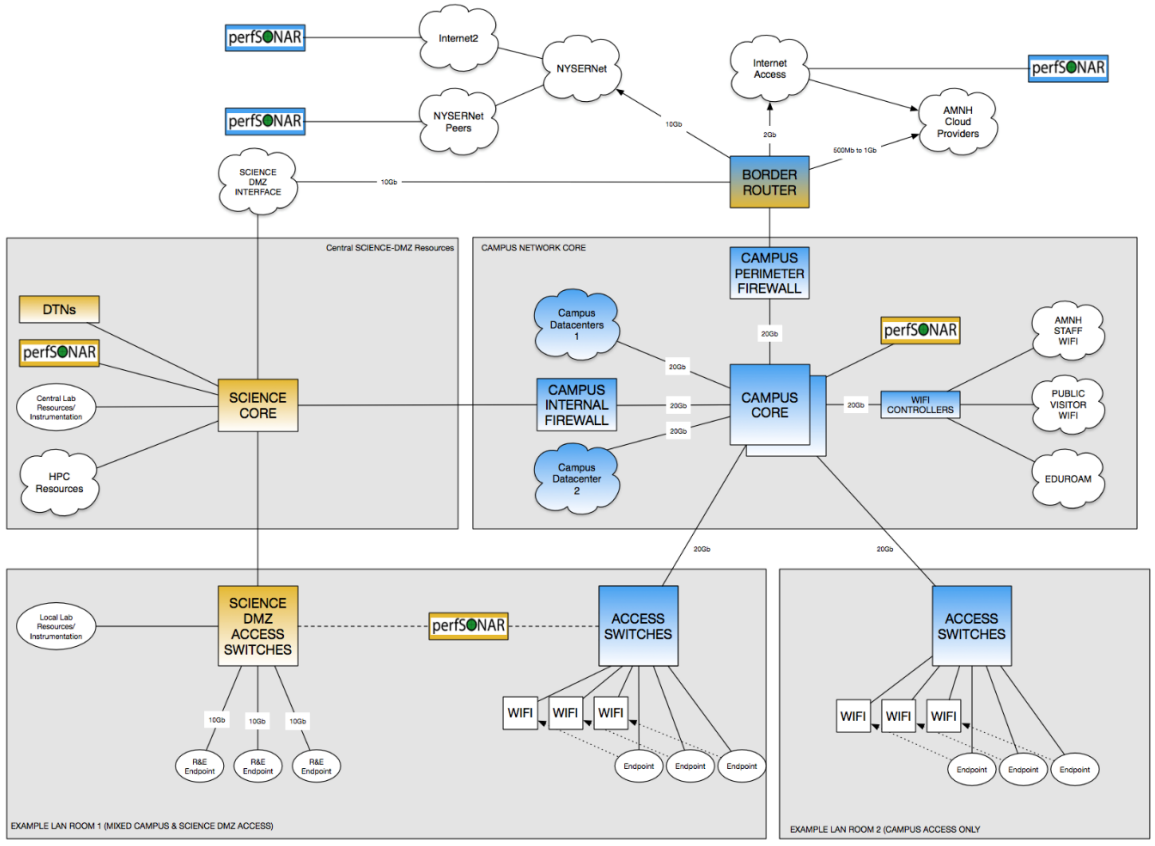
A key component supporting high-speed computing throughout the infrastructure is data security. Our aim is to provide sustainable data security that accounts for the varied risks and sensitivities of a given system or data while minimizing the impacts to those data flows as much as possible. AMNH has already made strides in segmenting its infrastructure into logical groupings based on common needs and security

requirements. This segmentation will allow us to apply the appropriate levels of security to those assets as needed. This means we can apply stringent security controls on systems housing sensitive data (whether research data, administrative data, or otherwise) while maintaining appropriate levels of openness on less sensitive segments of the network.

As shown in the network design drawing below, AMNH has deployed a perimeter firewall to protect the enterprise network from external attack and control outbound data flows to the Internet. AMNH has also installed an internal campus firewall to protect “high security” subnets (AMNH servers, HR, payroll, finance, cardholder data environments, campus security and safety systems, building management systems, etc.). Science and Education subnets are situated between the perimeter firewall and the internal high security firewall. Security for those subnets is provided via core router access control lists (ACLs). This allows for data flows between the science and education subnets to occur without the overhead of a firewall. Other security best practices (intrusion detection/prevention, host-based security, continual monitoring, blacklists, etc.) are used to enhance the overall security of the enterprise network.

In the case of the planned Science DMZ, AMNH is designing it using the best practices outlined by ESNet. Specifically, router ACLs will be used to limit access, log and Netflow aggregation and correlation, utilization monitoring, host port control, and host and network intrusion detection will be used to provide the necessary security on Science DMZ connected systems without a firewall impeding data flows.

The proposed network design is as follows:



Items shown in blue are existing devices. Items in gold are proposed.

4. Expansion of High Performance Computing (HPC) Capabilities for Researchers

Often regarded as a key enabler of scientific advances in the modern age, AMNH will be expanding upon our existing HPC efforts. AMNH IT has taken an active role as a partner with our research community in the development and deployment of the most recent HPC clusters at the Museum. In addition to working on the technical aspects of the Museum's clusters, AMNH has formed an HPC working group made up of senior members of the IT and science research teams, as well as the CIO and Provost.
Currently, AMNH maintains several smaller clusters including:

- "COSMO"- 280 Compute CPU
- "ARES"- 100 Compute CPU
- "FREYJA"- 382 Compute CPU
- "SGE"- 36 Compute CPU (SPARC)

These resources are largely allocated to specific researchers, projects, or labs due to the limited resources currently on hand, however the goal of the Museum is to develop a generalized HPC capability that can support all AMNH researchers and students. Additionally, through the planned deployment of our Science-DMZ and associated connections to regional, national, and global research networks, AMNH will seek to further expand our HPC capabilities by collaborating with XSEDE and Open Science Grid, enabling our researchers to have greater and more seamless access to available HPC assets around the world while contributing our spare cycles and expertise back to the community.

5. Large Scale Storage Capacity for Scientific and Archival Data

Data storage, archiving, data mining, and dissemination of the collected data are significant challenges faced by AMNH, as well as other education and research facilities around the world. While AMNH maintains a central SAN of over 250TB on the campus, shared across various departments and disciplines, a great deal more data is being stored by individual researchers within their own labs. Decentralized storage makes it extremely difficult for data to be effectively catalogued and shared within the AMNH and with the Museum's research and science institutional partners. Additionally, we expect the need for data storage to grow into petabytes of data within the next 5-7 years. AMNH's ever-growing collection of scientific and observational data requires a system for cataloguing, preservation, and dissemination for use both internally and with the national and global research and scientific community. To address this issue, AMNH is seeking to increase the amount of centralized storage available for our research community, both through the acquisition of additional storage capabilities as well as the migration of appropriate storage loads to the cloud. AMNH has also begun to study and catalogue the data currently in its collection with the intent on unifying it through the use of a Digital Asset Management (DAM) System.

6. Identity and Access Management

In mid-2016, AMNH joined InCommon. Through InCommon, AMNH now provides secure two-factor authentication using DUO Mobile for all users. In the summer of 2016, AMNH deployed certificate services via the InCommon partnership and is now moving towards the use of InCommon Federation and Assurance to provide seamless single-sign on (SSO) capabilities for our researchers traveling to other institutions and for outside collaborators hosted at the Museum.

AMNH is also planning to provide access to Eduroam, a secure, world-wide roaming access service developed for the international research and education community. Eduroam allows students, researchers and staff from participating institutions to obtain Internet connectivity across campus and when visiting other participating institutions by simply opening their laptop. This deployment is planned for 2018.

7. Educational Outreach and Workforce Development

AMNH is committed to developing the next generation of scientists and science educators through the continued support of innovative education programs, including RGGS Ph.D. programs, MAT programs, and initiatives such as the BridgeUp STEM, a portfolio of programs that is focused on the intersection of computer science and science education for students and women. Because of the growing need for computational resources in modern research, the ability to provide these resources to our students and adult learners will become a focal point for the institution. By gaining an understanding of the proper and effective use of computational technology in a research setting and knowledge on where and how to access the large and dispersed scientific datasets necessary for their research, students will become the empowered and effective in researchers of the future.

8. Involvement in Regional and National Cyberinfrastructure Programs

As a world-class research and educational institution, it is critical for AMNH to stay on top of emerging trends in research computing and the cyberinfrastructure supporting these endeavors. AMNH IT staff will continue to take part in regional research and education network meetings and working groups facilitated by NYSERNet and expand our involvement at a national and international level with the Internet2, InCommon, Open Science Grid, Educause, and other programs and consortia focused on developing a sustainable network infrastructure supporting science. To this end, in the summer of 2017, AMNH took part in the “The First National Research Platform Workshop:Toward A National Big Data Superhighway” conference in Bozeman, MN and is currently developing a plan for further engagement with the national R&E community in 2018 and beyond.

9. Sustainibility

AMNH realizes that the long term success of any technology initiative lies in the ability to integrate those technoligies into the Museum culture and existing infrastructure. Technology advancements should be complimentary rather than conflicting. To that end, AMNH IT strives to implement best of breed technologies in a manner that allows those systems to be supported with minimal distruption to the existing operations, utilizing existing management systems, techniques, and staff where available. AMNH IT also seeks a consensus from the Museum community before undertaking any major technology initiative to ensure that users and administrators alike understand its impacts, ongoing costs, and management requirements. Additionally, AMNH IT is fully engaged with the Museum’s forthcoming expansion of the campus through the addtion of the Richard Gilder Center for Science, Education, and Innovation which will house significant portions of the AMNH collections, research and education departments, and the Museum Library. All designs undertaken as part of any infrastructure upgrade are designed to ensure that the can be extended into the new facility upon completion. This ensures that projects are sustainible at all levels of the Museum and investements can be leveraged for years to come.