Title
University of Central Florida Campus-Wide Deep Dive

Authors
Zurawski, Jason
Miller, Ken
Addleman, Hans

Publication Date
2021-08-20

Peer reviewed

University of Central Florida Campus-Wide Deep Dive

August 31, 2021

Disclaimer

University of Central Florida Campus-Wide Deep Dive

Final Report

August 31, 2021

The Engagement and Performance Operations Center (EPOC) is supported by the National Science Foundation under Grant No. 1826994.

ESnet is funded by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. Benjamin Brown is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This is a University of California, Publication Management System report number LBNL-2001419[1].

[1] https://escholarship.org/uc/item/7fh8p0mw

Participants & Contributors

Hans Addleman, Indiana University
Dr. Scott Branting, University of Central Florida
Dr. Shafaq Chaudhry, University of Central Florida
Dr. Henry Glaspie, University of Central Florida
Dr. Fahad Khan, University of Central Florida
Jessica Madera, University of Central Florida
Dr. Glenn Martin, University of Central Florida
Ken Miller, ESnet
Dr. Eduardo Mucciolo, University of Central Florida
Joey Netterville, University of Central Florida
Dr. Otto Phanstiel, University of Central Florida
Ryan Schubert, University of Central Florida
Barry Weiss, University of Central Florida
Dr. Gregory Welch, University of Central Florida
Dr. Murat Yuksel, University of Central Florida
Jason Zurawski, ESnet

Report Editors

Hans Addleman, Indiana University: addlema@iu.edu
Ken Miller, ESnet: ken@es.net
Jason Zurawski, ESnet: zurawski@es.net

Contents

1 Executive Summary

Deep Dive Review Purpose and Process

EPOC uses the Deep Dive process to discuss and analyze current and planned science use cases and anticipated data output of a particular use case, site, or project to help inform the strategic planning of a campus or regional networking environment. This includes understanding future needs related to network operations, network capacity upgrades, and other technological service investments. A Deep Dive comprehensively surveys major research stakeholders' plans and processes in order to investigate data management requirements over the next 5–10 years. Questions crafted to explore this space include the following:

- How, and where, will new data be analyzed and used?
- How will the process of doing science change over the next 5–10 years?
- How will changes to the underlying hardware and software technologies influence scientific discovery?

Deep Dives help ensure that key stakeholders have a common understanding of the issues and the actions that a campus or regional network may need to undertake to offer solutions. The EPOC team leads the effort and relies on collaboration with the hosting site or network, and other affiliated entities that participate in the process. EPOC organizes, convenes, executes, and shares the outcomes of the review with all stakeholders.

This Review

Between December 2020 and August 2021, staff members from the Engagement and Performance Operations Center (EPOC) met with researchers and staff at University of Central Florida (UCF) for the purpose of a Campus-Wide Deep Dive into research drivers. The goal of this meeting was to help characterize the requirements for five campus research use cases and to enable cyberinfrastructure support staff to better understand the needs of the researchers they support.

This review includes case studies from the following UCF stakeholder groups:

- University of Central Florida Campus Overview
- College of Medicine Case Study
- Archeology & Anthropology Case Study
- SREAL Research Laboratory Case Study
- Department of Physics Case Study
- Networking and Wireless Systems Lab (NWSL) Case Study

Material for this event included the written documentation from each of the research areas at University of Central Florida, documentation about the current state of technology support, and a write-up of the discussion that took place via e-mail and video conferencing.

The Case Studies highlighted the ongoing challenges that University of Central Florida has in supporting a cross-section of established and emerging research use cases. Each Case Study mentioned unique challenges which were summarized into common needs.

The review produced several important findings from the case studies and subsequent virtual conversations:

- High Performance Computing HPC environments remain critical for a number of research use cases. UCF's investment into more intuitional computing will advance a number of science use cases in the coming years. Research IT staff that can function in a research facilitation role (e.g., assisting with adapting technology to scientific workflows) would greatly assist a number of groups.

- A number of researchers at UCF could benefit from access to Controlled Unclassified Information (CUI) environments.

- A number of departments and research labs maintain research computing and storage infrastructure that is independent of the centralized UCF Advanced Research Computing Center (ARCC). Due to efforts to centralize support, coupled to the age of the infrastructure, the long-term viability of these resources is in doubt. Working with entities to use long-term supported solutions for computation is recommended.

- The process to identify, procure, evaluate, and operate particular scientific software packages can be cumbersome, and should be better documented to set expectations, and streamlined to complete more efficiently.

- Computational and storage needs will increase in the coming years for most disciplines. UCF Research IT should prepare to scale the infrastructure, as well as the available staff to assist with research facilitation.

- Monitoring internal and external networking status is critical for a number of groups and use cases on campus. To ensure proper operation, UCF IT should consider methods to monitor performance of critical system infrastructure.

Lastly, EPOC and UCF identified a number of actions that will be followed up on in subsequent years:

1) UCF Research IT should investigate the creation of staff roles and expertise that can assist research groups in managing non-centralized IT resources.

2) UCF Research IT should investigate mechanisms to decommission non-centralized IT resources, and locate them centrally, by offering incentives to groups that are still running them independently.

3) UCF Research IT should work with other campus IT organizations to ensure appropriate access and procedures that enable backup use of infrastructure (e.g., storage, portable containers) to ensure that research progress is not impacted by expected or unexpected downtime.

4) UCF Research IT should consider the creation of staff roles and expertise to assist with workflow automation for scientific use cases. This would consist of, but is not limited to: containers, data movement approaches, automated backups, addressing the use of remote computing or storage resources, and accessing national resources such as XSEDE or using OSG software.

5) UCF Research IT should work to set expectations with the research community on the process and procedure to procure and evaluate software.

6) UCF Research IT and ARCC must be nimble in supporting future research computing needs, particularly by ensuring that computational power (e.g., available cycles), types (e.g., high-throughput, high-performance, cloud, GPUs, etc.), and storage quantities continue to scale with the requirements of the campus. These should be coupled with data mobility hardware and software upgrades.

7) UCF Research IT should take steps to adopt and utilize perfSONAR monitoring for internal and external network use cases.

8) UCF Research IT should investigate ways to support faculty that perform work at remote field sites through traveling infrastructure (computation and storage), or remote connectivity (e.g., satellite) options.

9) UCF Research IT should research methods to facilitate access, or the creation of, testbed networks for research use cases.

2 Deep Dive Findings

The deep dive process helps to identify important facts and opportunities from the profiled use cases. The following outlines a set of findings from the UCF Deep Dive that summarize important information gathered during the discussions surrounding case studies and the UCF campus program in general:

High Performance Computing needs & support:

- High Performance Computing HPC environments remain critical for a number of research use cases. The source for this computing comes from a number of infrastructures: privately maintained resources, UCF institutional resources, as well as regional and national collaborations such as XSEDE. UCF's investment into more intuitional computing options will help to consolidate this further.

- With the need for HPC resources comes the need for experts in conversion of workflows to use what is available. It is recommended that UCF invest in staff resources that can work with research groups to create and support workflows for the entire lifecycle (e.g., creation of the environment, porting of tools, management of data architecture), on resources that are local or national.

Support for research with compliance needs:

- A number of researchers at UCF could benefit from access to Controlled Unclassified Information (CUI) environments. UCF has recently invested in the creation of this, and should devote resources to understanding how to onboard and support the investment long term.

Supporting decentralized research cyberinfrastructure resources:

- A number of departments and research labs maintain research computing and storage infrastructure that is independent of the centralized UCF Advanced Research Computing Center (ARCC). Due to efforts to centralize support, coupled to the age of the infrastructure, the long-term viability of these resources is in doubt. Working with entities to use long-term supported solutions for computation is recommended.

Staffing and support for research cyberinfrastructure facilitation:

- Research IT staff that can function in a research facilitation role (e.g., assisting with adapting technology to scientific workflows) would greatly assist a number of groups.

- Computational and storage needs will increase in the coming years for most disciplines. UCF Research IT should prepare to scale the infrastructure, as well as the available staff to assist with research facilitation.

- Field research activities have unique technology requirements (e.g., networks to support remote transfer, or portable technology components) that require support from central IT staff.

Services & infrastructure support for research continuity:

- The ability to share fate between UCF IT organizations (e.g., backup agreements, portable infrastructure) would be desirable to ensure research continuation during scheduled or unscheduled downtime.

Services and support for scientific support:

- The process to identify, procure, evaluate, and operate particular scientific software packages can be cumbersome, and should be better documented to set expectations, and streamlined to complete more efficiently.

Network performance monitoring:

- Monitoring internal and external networking status is critical for a number of groups and use cases on campus. To ensure proper operation, UCF IT should consider methods to monitor performance of critical system infrastructure (e.g., through perfSONAR) which will allow proactive action to be taken on discovery of a problem.

Network testbeds for research:

- Network testbeds offer a unique way to facilitate research activities without interference from public user space.

- Access to national scale research testbeds (e.g., CloudLab, Emulab, GENI, etc.) is a high priority, provided that the resources remain usable and available for the target research use cases.

3 Deep Dive Action Items

EPOC and University of Central Florida recorded a set of action items that are a reflection of the Case Study reports and discussions:

- UCF Research IT will continue to seek out users that are interested in using the Controlled Unclassified Information (CUI) constructed for campus use. This infrastructure consists of dedicated computing, storage, targeted network security profiles, access control, and staff that is capable of evaluating compliance.

- UCF Research IT should investigate the creation of staff roles and expertise that can assist research groups in managing non-centralized IT resources.

- UCF Research IT should investigate mechanisms to decommission non-centralized IT resources, and locate them centrally, by offering incentives to groups that are still running them independently.

- UCF Research IT should work with other campus IT organizations to ensure appropriate access and procedures that enable backup use of infrastructure (e.g., storage, portable containers) to ensure that research progress is not impacted by known and unexpected downtime.

- UCF Research IT should consider the creation of staff roles and expertise to assist with workflow automation for scientific use cases. This would consist of, but is not limited to: containers, data movement approaches, automated backups, or addressing the use of remote resources.

- UCF Research IT should work to better set expectations with the research community on the process and procedure to procure and evaluate software.

- UCF Research IT/ARCC must be nimble in supporting future research computing needs, particularly by ensuring that computational power (e.g., available cycles), types (e.g., high-throughput, high-performance, cloud, GPUs, etc.), and storage quantities continue to scale with the requirements of the campus. These should be coupled with data mobility hardware and software upgrades.

- UCF IT should take steps to adopt and utilize perfSONAR monitoring for internal and external network use cases.

- UCF Research IT should investigate ways to support faculty that perform work at remote field sites through traveling infrastructure (computation and storage), or remote connectivity (e.g., satellite) options.

- UCF Research IT should research ways to facilitate the creation of testbed networks for research use cases.

4 Process Overview and Summary

4.1 Campus-Wide Deep Dive Background

Over the last decade, the scientific community has experienced an unprecedented shift in the way research is performed and how discoveries are made. Highly sophisticated experimental instruments are creating massive datasets for diverse scientific communities and hold the potential for new insights that will have long-lasting impacts on society. However, scientists cannot make effective use of this data if they are unable to move, store, and analyze it.

The Engagement and Performance Operations Center (EPOC) uses the Deep Dives process as an essential tool as part of a holistic approach to understand end-to-end research data use. By considering the full end-to-end research data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports five main activities

- Roadside Assistance via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- Application Deep Dives to work more closely with application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues;
- Provision of managed services via support through the Indiana University (IU) GlobalNOC and our Regional Network Partners; and
- Coordinated Training to ensure effective use of network tools and science support.

Whereas the Roadside Assistance portion of EPOC can be likened to calling someone for help when a car breaks down, the Deep Dive process offers an opportunity for broader understanding of the longer term needs of a researcher. The Deep Dive process aims to understand the full science pipeline for research teams and suggest alternative approaches for the scientists, local IT support, and national networking partners as relevant to achieve the long-term research goals via workflow analysis, storage/computational tuning, identification of network bottlenecks, etc.

The Deep Dive process is based on an almost 10-year practice used by ESnet to understand the growth requirements of Department of Energy (DOE) facilities (online at https://fasterdata.es.net/science-dmz/science-and-network-requirements-review). The EPOC team adapted this approach to work with individual science groups through a set of structured data-centric conversations and questionnaires.

4.2 Campus-Wide Deep Dive Structure

The Deep Dive process involves structured conversations between a research group and relevant IT professionals to understand at a broad level the goals of the research team and how their infrastructure needs are changing over time.

The researcher team representatives are asked to communicate and document their requirements in a case-study format that includes a data-centric narrative describing the science, instruments, and facilities currently used or anticipated for future programs; the advanced technology services needed; and how they can be used. Participants considered three timescales on the topics enumerated below: the near-term (immediately and up to two years in the future); the medium-term (two to five years in the future); and the long-term (greater than five years in the future).

The Case Study document includes:

- ***Science Background***—an overview description of the site, facility, or collaboration described in the Case Study.
- ***Collaborators***—a list or description of key collaborators for the science or facility described in the Case Study (the list need not be exhaustive).
- ***Instruments and Facilities***—a description of the network, compute, instruments, and storage resources used for the science collaboration/program/project, or a description of the resources made available to the facility users, or resources that users deploy at the facility.
- ***Process of Science***—a description of the way the instruments and facilities are used for knowledge discovery. Examples might include workflows, data analysis, data reduction, integration of experimental data with simulation data, etc.
- ***Remote Science Activities***—a description of any remote instruments or collaborations, and how this work does or may have an impact on your network traffic.
- ***Software Infrastructure***—a discussion focused on the software used in daily activities of the scientific process including tools that are used locally or remotely to manage data resources, facilitate the transfer of data sets from or to remote collaborators, or process the raw results into final and intermediate formats.
- ***Network and Data Architecture***—description of the network and/or data architecture for the science or facility. This is meant to understand how data moves in and out of the facility or laboratory focusing on local infrastructure configuration, bandwidth speed(s), hardware, etc.
- ***Cloud Services***—discussion around how cloud services may be used for data analysis, data storage, computing, or other purposes. The case studies included an open-ended section asking for any unresolved issues, comments or concerns to catch all remaining requirements that may be addressed by ESnet.

- ***Resource Constraints*—**non-exhaustive list of factors (external or internal) that will constrain scientific progress. This can be related to funding, personnel, technology, or process.
- ***Outstanding Issues*—**Final listing of problems, questions, concerns, or comments not addressed in the aforementioned sections.

At an in-person meeting, this document is walked through with the research team (and usually cyberinfrastructure or IT representatives for the organization or region), and an additional discussion takes place that may range beyond the scope of the original document. At the end of the interaction with the research team, the goal is to ensure that EPOC and the associated CI/IT staff have a solid understanding of the research, data movement, who's using what pieces, dependencies, and time frames involved in the Case Study, as well as additional related cyberinfrastructure needs and concerns at the organization. This enables the teams to identify possible bottlenecks or areas that may not scale in the coming years, and to pair research teams with existing resources that can be leveraged to more effectively reach their goals.

4.3 University of Central Florida Campus-Wide Deep Dive Background

Between December 2020 and August 2021, EPOC organized a Campus-Wide Deep Dive in collaboration with University of Central Florida to characterize the requirements for several key science drivers. The University of Central Florida representatives were asked to communicate and document their requirements in a case-study format. These included:

- University of Central Florida Campus Overview
- College of Medicine Case Study
- Archeology & Anthropology Case Study
- SREAL Research Laboratory Case Study
- Department of Physics Case Study
- Networking and Wireless Systems Lab (NWSL) Case Study

Virtual meetings took place throughout 2021 (see discussion in Section 6).

4.4 Organizations Involved

The Engagement and Performance Operations Center (EPOC) was established in 2018 as a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy (DOE) Office of Science (SC), the single largest supporter of basic research in the physical sciences in the United States. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 25 years.

Indiana University (IU) was founded in 1820 and is one of the state's leading research and educational institutions. Indiana University includes two main research campuses and six regional (primarily teaching) campuses. The Indiana University Office of the Vice President for Information Technology (OVPIT) and University Information Technology Services (UITS) are responsible for delivery of core information technology and cyberinfrastructure services and support.

University of Central Florida Founded in 1963 to provide talent for Central Florida and the growing U.S. space program, UCF has been making an impact on the state, the nation — and outer space — ever since. With 13 colleges and more than 230 degree programs, follow your passion at one of our campus locations designed to help you succeed. From the main campus in east Orlando to UCF Downtown and Rosen College of Hospitality Management to the Academic Health Sciences Campus, you'll have the opportunity to learn in the heart of your industry. UCF has a strong research focus, and received $204.5 M in research funding in 2019-2020.

5 University of Central Florida Case Studies

University of Central Florida presented four scientific use cases, and one campus technology overview, during this review. These are as follows:

- University of Central Florida Campus Overview
- College of Medicine Case Study
- Archeology & Anthropology Case Study
- SREAL Research Laboratory Case Study
- Department of Physics Case Study
- Networking and Wireless Systems Lab (NWSL) Case Study

Each of these Case Studies provides a glance at research activities for the University, the use of experimental methods and devices, the reliance on technology, and the scope of collaborations. It is important to note that these views are primarily limited to current needs, with only occasional views into the event horizon for specific projects and needs into the future. Estimates on data volumes, technology needs, and external drivers are discussed where relevant.

University of Central Florida is committed to supporting these use cases through technology advancements, and is actively pursuing grant solicitations. The landscape of support will change rapidly in the coming years, and these use cases will take full advantage of campus improvements as they become available.

5.1 Campus Overview

Content in this section authored by Barry Weiss, Glenn Martin, Joey Netterville, Henry Glaspie, Shafaq Chaudhry, and Fahad Khan

5.1.1 Institutional Background

As one of the largest state universities in the U.S., the University of Central Florida's network supports 13 colleges and a multitude of programs for over 70,000 users. The goal of the University's architecture is to support the needs and services of a diverse customer base of students, faculty, and employees while ensuring network stability, availability, and scalability to accommodate unanticipated needs. The University of Central Florida's network has about 1,200 devices in service using primarily Cisco as the vendor of choice for the enterprise network, and the vendor Extreme for the residential and research networks. The University of Central Florida is currently in the process of a major network redesign with an emphasis on service provider designs and services. Future designs will include a move from traditional Enterprise Campus design to routed segmentation using a mixture of L2VPNs and L3VPN MPLS architecture. Furthermore, the direction is to replace existing core infrastructure from Cisco Catalyst 6500s using VSS to Cisco ASR and NCS routers. The goal is to migrate to a completely routed underlay network.

5.1.1.1 UCF Campus Network

The main campus currently consists of 6 core "node" switches, which are 6509 V-E dual VSS chassis using Sup2T-10G (4 supervisors in total per node). The nodes located at the CSB building and the LSR building act as cores that connect all remote sites and the main campus WAN Edge network. The CSB core and LSR core also connect the two primary main campus data center networks. The remaining nodes, data centers, and WAN have one connection to the LSR and the CSB core switches. The overall core and node design adopts a CLOS style architecture, with the idea of maintaining simplicity while maintaining stability and availability. As of right now each connection to the CSB and LSR nodes uses 20 Gbps Ethernet port-channels to each node and data center.

The LSR, SSR, LCSR, NSR, and CFRP core switches also act as area aggregation nodes for multiple campus buildings. Each building has a layer 3 distribution switch that connects to the aggregation nodes via OSPF. The distribution switch at each building serves as the layer 3 gateway for a building local network and aggregates the access layer switches at the building layer as well. Almost all buildings have a 20 Gbps Ethernet port-channel from their distribution layer switches to their respective area node. Most buildings are single homed to one core node VSS pair; however, some buildings that are highly critical (e.g., Police Department, Facilities, IST, etc.) are dual homed to two geographically separated cores.

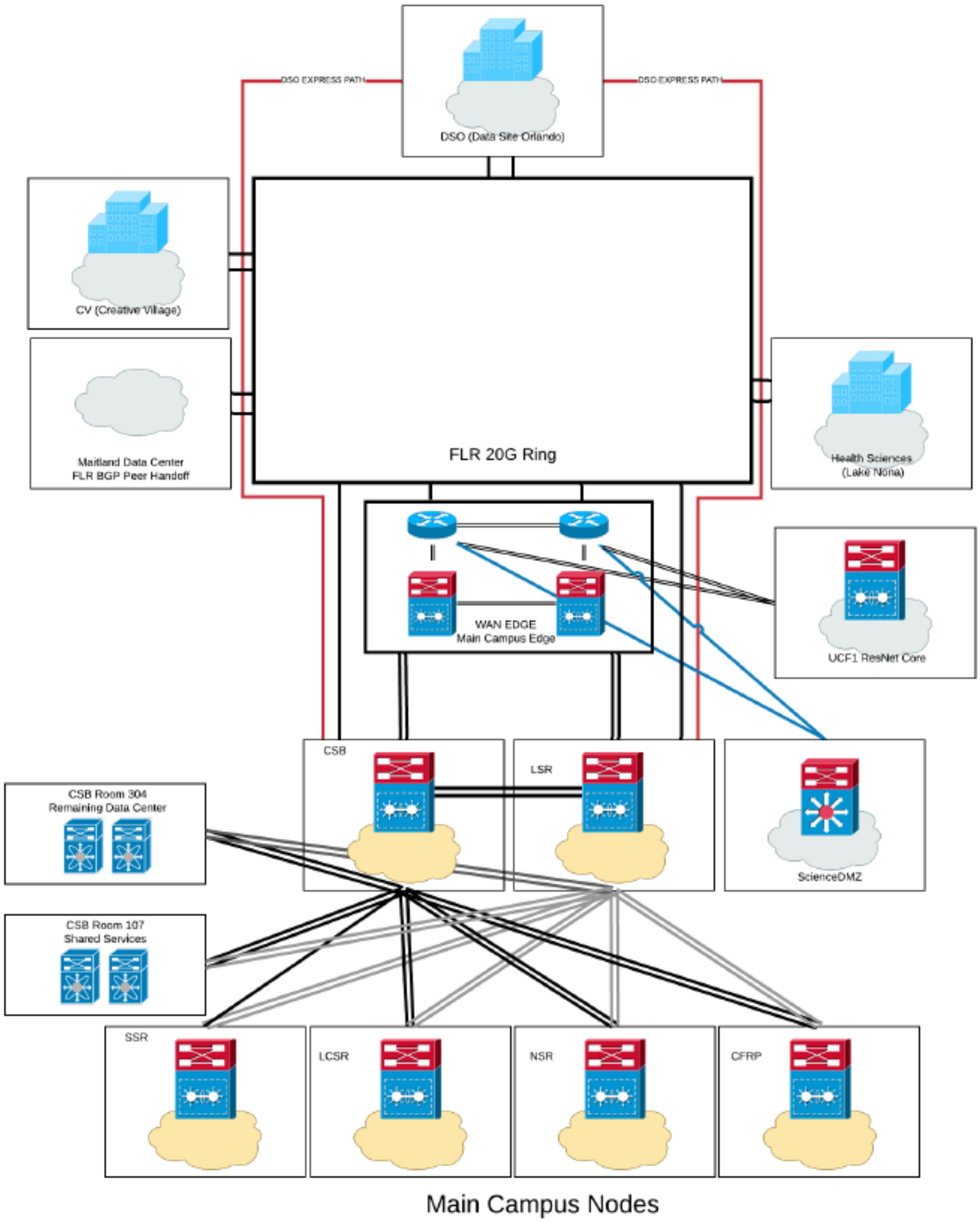

Figure 3.1 – UCF Campus Network

5.1.1.2 FLR Ring Summary

The main campus WAN Edge is made up of two Cisco Catalyst 9500 virtual stack pairs (one at CSB and one at LSR), one Cisco Catalyst 9300 that aggregates main campus guest and Eduroam wireless firewalls and controllers, and two Cisco ASR 9906 border routers. All uplinks between the Cisco Catalyst 9500s and the ASR 9906 routers are 20 Gbps Ethernet port-channels. The WAN aggregate Catalyst switches aggregate the main campus core network, the Vendor network (UCF DotCom), border routers, and firewalls. The uplinks on the Cisco ASR routers to FLR are 20 Gbps Ethernet port-channels. The ASR routers serve the Main Campus I1, I2 and commodity networks, by using full internet BGP routing tables to ensure the most efficient routing possible.

FLR (Florida Lambda Rail), an independent statewide research and education fiber optic network, is currently the sole WAN provider for UCF and operates the DWDM ring in Central Florida that connects the UCF main campus and other remote sites. The black set of lines ("Black Line) on the FLR ring depicted on the diagram above is currently 20 Gigs in capacity and is physically and logically managed via MPLS L2VPN by FLR. The Black Line currently supports the internet connectivity as well as current inter-campus traffic. The FLR "Red Line" is managed physically by FLR but UCF manages the Logical connectivity. The Red Line is also called the DSO Express circuit, that connects DSO (Data Site Orlando, UCF's offsite datacenter) directly to UCF's main campus. In the future, UCF IT Network Services plans to build a parallel ring that will be physically run by FLR but logically run by UCF. The initial proposed use for the UCF MPLS ring, also called the "Orange Line," will be for inter-campus traffic.

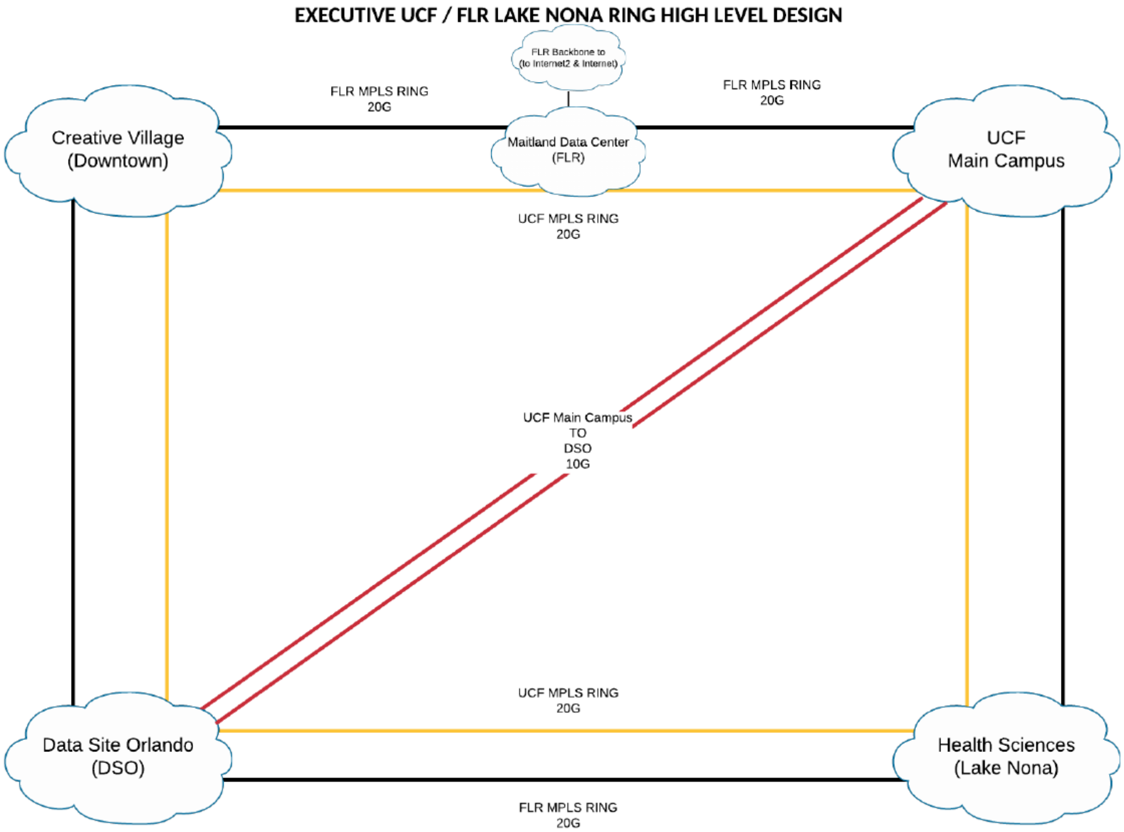

Figure 3.2 – UCF / FLR Metro Ring

5.1.1.3 Lake Nona (College of Medicine) Network

The Lake Nona campus is located 22 miles to the southwest of the main campus. Currently, the UCF network at this campus services the College of Medicine, Burnett Bio-Medical Sciences, and the Lake Nona Cancer Center. Lake Nona serves about 1,100 students, researchers, faculty, and staff. The University is currently constructing a hospital on this campus as well, but as of this writing that network will not be serviced by UCF IT Telecommunications.

The Lake Nona campus connects to the FLR ring for WAN and inter-campus connectivity. It should be noted that Lake Nona has its own border routers that it uses for internet connectivity. These border routers have their own eBGP peers with FLR, which provides a full internet route table providing Lake Nona I1, I2, and commodity routes. The border routers also provide INTERNET VRF iBGP connectivity between all other campuses on the FLR Black Line ring. The two border routers are Cisco ASR 1002-HX routers, each with a 10 Gbps ethernet uplink to the WAN ring. It should be noted that this 10 Gbps uplink supports both internet and inter-campus traffic; in other words, the bandwidth is shared.

There are plans to create a new MPLS ring, also known as the "Orange Line," which would offload the intercampus traffic. As mentioned previously in this document, the Orange Line would start at 20 Gbps. As of right now Lake Nona is solely using FLR Black Line for connectivity.

5.1.2 Collaborators

5.1.2.1 Research Computing and Data Services

Research computing and data services are brought through to researchers via a concerted effort of various groups across UCF, which include Advanced Research Computing Center (ARCC[2]), Graduate and Research IT (GRIT[3]), UCF Libraries[4], Health IT[5] (at College of Medicine[6]), UCF IT[7], Office of Cyber Risk Management[8], InfoSec[9] (under UCF IT), and others. GRIT and OCRM operate under the banner of Office of VP Research. These services cover research computing (Stokes, Newton, HPC, HTC, etc.); data services (data management plans, long-term data archival, data transfer, STARS, Globus, etc.); proposal support; cloud computing (AWS, Azure); regulated research environments (Knight Shield); data collection and analysis tools (ArcGIS, SPSS, Qualtrics, REDCap, etc.); collaboration tools; and workshops for scientific computing and data management. More details can be found on the UCF Research Cyberinfrastructure website[10].

5.1.2.2 Research Facilitation at UCF

Facilitation for research computing and data services is provided through a strong partnership between Graduate and Research IT (GRIT), Advanced Research Computing Center (ARCC) and UCF Libraries. Facilitation includes guidance, consultation, liaising, or support for the following bracket of services:

1. *Technical Consulting:* Researchers are provided consulting, guidance, & liaising for research computing, data, archival, collaboration, cloud, and research software tools.
2. *Proposal Support:* Researchers can seek technical write-up assistance for proposals regarding infrastructure or service particulars; including both cloud and on-premises resources.
3. *Cloud Computing for Research:* Researchers at UCF can use AWS and Azure for scientific workloads requiring cloud computing support. The AWS environment for scientific research is sponsored by the Office of Research; OR pays for shared network resources and core services (which includes necessary guardrails for research activity). GRIT manages this environment and assists with on-boarding. Researchers are provided with budget & cost

---

[2] https://arcc.ist.ucf.edu
[3] https://it.research.ucf.edu
[4] https://library.ucf.edu
[5] https://med.ucf.edu/administrative-offices/health-information-technology/
[6] https://med.ucf.edu
[7] https://it.ucf.edu
[8] http://ocrm.research.ucf.edu
[9] https://infosec.ucf.edu/
[10] https://rci.research.ucf.edu

estimates for cloud services, consultation for architecting cloud solutions, and auditing & cloud security review for research projects. Researchers are also facilitated with assistance in deploying GPU & HPC workloads on Azure and AWS resources.

4. *High Performance Computing and Data:* Stokes & Newton, managed by ARCC, are HPC clusters subsidized by the UCF Provost and VP for Research for supporting researchers with compute-intensive workloads. ARCC assists researchers with on-boarding to these clusters and provides guidance and support for them to run their workloads in an efficient manner. Researchers can leverage Globus for large-volume and high-speed data transfers. Researchers are provided guidance for leveraging either-or-both on-premises and national research computing and data facilities. They can be connected to HPC experts for consulting purposes, both inside the University (ARCC, CECS, etc.) and externally across the United States (e.g., XSEDE, OSG, HiPerGator, etc.). ARCC & GRIT team members serve as XSEDE Campus Champions.
5. *Assessment of Research Cyberinfrastructure Needs:* These teams work with faculty one-on-one, through surveys, and through external assessments (e.g., EPOC) to determine the current state of needs and issues regarding research cyberinfrastructure at UCF. Based on these assessments and monitoring resource usage statistics, proposals are made for evolving, fixing, or introducing new services and resources for research cyberinfrastructure at UCF.
6. *Workshops for Research Computing & Data Management:* UCF provides trainings opportunities for researchers through workshops and hands-on labs. These workshops are brought about through the combined effort of multiple partners at UCF which are GRIT, UCF Libraries, ARCC. The speakers include experts from both within the University (e.g., GRIT, UCF Libraries, ARCC, UCF IT, CECS, etc.) and external to UCF (AWS, XSEDE, OSG, Globus, etc.)
7. *Other Services*
    a. Contract review and compliance coordination.
    b. Facilitating the review of any vendor agreements for research cyberinfrastructure tools; both hardware and software.

5.1.2.3 Knight Shield for Regulated Research

Through a strong collaborative partnership among the UCF Office of Research, IST (School for Modeling, Simulation, and Training), UCF Information Technology, and the UCF Information Security Office, Knight Shield is a managed environment for handling sensitive and restricted research data. Knight Shield is the only approved UCF environment in compliance with NIST SP 800-171 and Cybersecurity Maturity Model Certification (CMMC Level 3) for handling Controlled Unclassified Information (CUI) and can be used for supporting Federal Information Security Modernization Act (FISMA), and Health Insurance Portability and Accountability Act (HIPAA), Export Controlled (ITAR/EAR), and other contractually regulated research.

The Knight Shield environment currently consists of an on-premises network infrastructure and an AWS cloud environment.

The on-premises environment is separated from the academic network and provides an operational and managed environment that provides network level support (including but not limited to firewall and intrusion prevention systems), system level support (for all endpoints), computing and storage for research projects, and security incident response and monitoring.

Since research is spread throughout campus, many researchers in various locations on campus need access back to the Knight Shield environment to stay in compliance. Unlike the ScienceDMZ, which the LAN portion is physically separated and private network from the general users' network (UCF VRF), IST's network is only reachable via the layer 3 general network. To provide proper segmentation across the general network to Knight Shield, UCF makes use of either MPLS L2VPN circuits over the campus underlay network, or a direct 10 Gbps ethernet fiber to the distribution switch at the researcher's location. Once again it is worth noting that this research is not part of the ScienceDMZ and is not designed for high-speed data transfers, but rather NIST 800-171 compliant research.

The cloud environment is a secure enclave implemented in the AWS East region with FedRAMP Moderate or equivalent controls and uses a managed service provider. The architected infrastructure gives each researcher a virtual private cloud developed and managed to meet the research need and the specific contracted compliance requirements of their grant, contract, or award.

Managed services in these enclaves include:

- Assistance with System Security Plan (SSP) development;
- Virtual desktops and servers with memory and storage to meet research requirements;
- Installation of approved and licensed software;
- Secure data transfer and data sharing;
- Multi Factor authentication
- Network security monitoring and incident response; and
- 24/7 availability (excluding maintenance and planned outages, and unavoidable events)

5.1.3 Instruments and Facilities

5.1.3.1 Advanced Research Computing Center

UCF is home to the Advanced Research Computing Center (ARCC). This is capable of housing a variety of high-end computational resources, including existing infrastructure such as the Stokes High Performance Computing cluster. The ARCC is located in the Central Florida Research Park adjacent to the UCF Orlando campus. The ARCC is a full member of Florida's Sunshine State Education and Research Computing Alliance (SSERCA).

The ARCC houses two High Performance Computing ("HPC") clusters, Stokes and Newton, and 240 TB of storage. Stokes includes over 4,300 Intel Xeon® CPU cores and 56 Gb/s FDR InfiniBand interconnect between all nodes. Newton is a GPU accelerated cluster with a total of 40 NVidia Tesla V100 GPUs across 20 nodes with 100 Gb/s EDR InfiniBand

The ARCC machine room is outfitted with 3-phase, 208 volt 60 amp power, as well as industrial scale UPS modules and a separate generator from the main building for critical ARCC equipment. It also has five redundant, industrial scale air conditioning units. There is sufficient power, cooling, and rack space for the proposed equipment.

In addition, the ARCC maintains a dedicated research network and Science Demilitarized Zone (DMZ) for high-speed, low-latency data needs for research. This network has direct, unfettered access to the Internet2 and implements a frictionless 10 Gb connection from the border router to wall jack. This research network is also available for use for the current project.

5.1.3.2 Advanced Research Network (Science DMZ)

UCF's Research Network was implemented in 2013 based on funds from CC*IE grants. The primary goal of the research network is to provide high-speed access for researchers and equipment to campus resources. The Research Network also provides a direct link from the campus to Internet2 resources. As a secondary focus, the Research network provides a platform to monitor bandwidth through perfSONAR.

The UCF Research Network uses Extreme network switches at the building distribution and floor-based access layer. During initial acquisition, Extreme was implemented due to inexpensive cost for dense access ports. Additionally, at the time of purchase, Extreme was one of the few network hardware vendors that provided OpenFlow capabilities and support for SDN services. The UCF Research Network relies on Cisco routers to connect to the Enterprise edge. Cisco provides access to advanced networking features, high availability and line speed routing needed to support a service provider level of quality.

The Research Network is directly linked to the buildings housing the College of Engineering, College of Computer Science and The Modeling and Simulation program. These locations provide direct access for Research members to connect to UCF perfSONAR nodes, High Performance Computing (HPC), Software Defined Networks (SDN) and Data Transfer Nodes (DTN) within campus, as well as Internet2 resources.

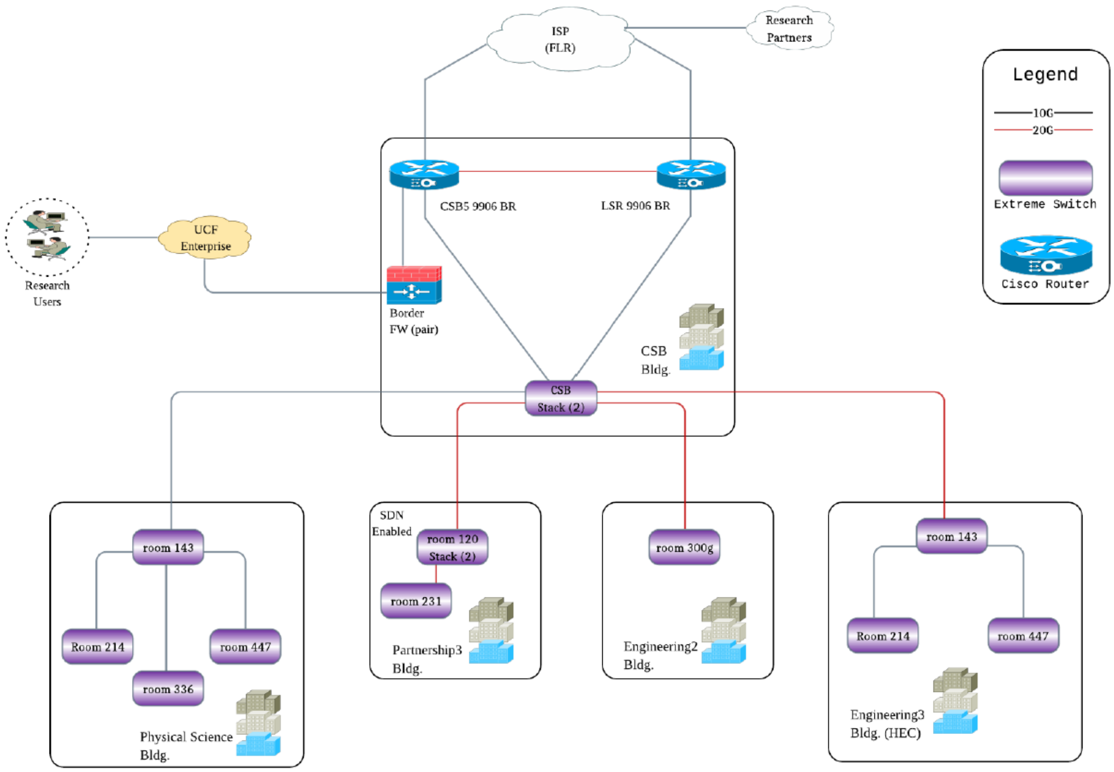

Figure 3.3 – UCF Research Network

5.1.3.3 Data Site Orlando (DSO)

As part of the University's 2017 Cloud initiative, a decision by UCFIT leadership was made to migrate our primary data center at CSB to a co-located hosting facility in south Orlando. The goal is to reduce large data centers, owned by UCF IT, from main campus and relocate them to a dedicated third-party colocation center or move them to Microsoft's Azure where applicable.

DSO has a build similar to Lake Nona. It uses two Cisco ASR 1002-HX routers, each with 10 Gbps uplinks via the FLR Black Line. Unlike Lake Nona, these uplinks only support internet traffic. Each border also has a dedicated 10 Gbps uplink back to the University's main campus. This uplink is physically operated by FLR, but uses MPLS that is operated by UCF IT. This circuit is known as the "Red Line" or "DSO Express Circuit." Depending on funding the Orange Line ring will be expanded to support further inter-campus connectivity to Lake Nona and Creative Village (Downtown).

5.1.3.4 Networking for Microsoft Azure and AWS Transport

The University of Central Florida IT (UCF IT) operates an IaaS platform on Microsoft's Azure Cloud Services. In addition to public facing connectivity for Azure, which uses internet connectivity, UCF also maintains two point-to-point VPN tunnels to Microsoft Azure Central and Microsoft Azure East regions. These VPN Tunnels to Azure allow direct connectivity to Azure from the general users' network (UCF VRF). The tunnel headend devices are geographically diverse. One device is located at main campus at CSB room 304 and the other is located at Data Site Orlando (DSO). Both devices are Cisco ASR 1001-X routers with a 2.5 Gbps throughput license and are physically connected with 10 Gbps uplinks.

UCF does not currently have a direct connection or VPN to AWS, so all access to AWS is done over the Internet, which is FLR's black line.
UCF IT does not yet have a formal agreement with Google Cloud Services and, like AWS, is only accessible over the internet.

The Health Information Technology Office resides at the Lake Nona Health Sciences Campus and primarily supports College of Medicine, Burnett School of Biomedical Sciences, UCF Health and the Lake Nona Cancer Center. As the Medical Education program has expanded into the Academic Health Sciences umbrella, collaboration and support have started to extend to other colleges and departments that form part of the Academic Health Sciences at UCF.

5.1.3.5 UCF Health IT

The Heath IT department manages a data center that employs industry standard virtualization and is able to provide stand-alone virtual systems, servers, and assist with cloud solution delivery according to technical requirements provided by Principal Investigators in the departments supported. These systems reside on separate virtual networks classified according to UCF's data classification policies. Services provided include compute, storage, and back up services based on individual grants and contracts. All systems are backed up on a daily basis, and disaster recovery copies are maintained in an offsite secure location (data retention varies per compliance). Support to researchers is also provided for deployment of lab instrumentation and software, provisioning secure access to existing systems and shared cloud storage, software development and application delivery, and security compliance and risk.

5.1.4 Wireless Network

The University of Central Florida's wireless network is made up of over 4,000 wireless access points (WAP) and 45 wireless LAN controllers (WLC) that service the main campus as well as Health Sciences (Lake Nona), Creative Village (Downtown Campus), and a number of remote campuses throughout Central Florida. The wireless architecture employs a distributed WLC design, and each section of campus has its own pair of WLC that uses an N+1 redundancy configuration. There are also some WLCs that are specialized as anchors to facilitate unique access requirements, such as guest wireless and campus vendor access. All WLCs on the campus are protected by a firewall. As for authentication, depending on the wireless network that is used, the University of Central Florida can use WPA2/Enterprise, WPA2/Personal, and 802.1X (PEAP MS-CHAPv2).

Currently, Cisco is the sole vendor of the University of Central Florida's wireless network. About 13% (537 WAPs) of the WAPs on the wireless network are the Cisco 2800s and 3800s series models, which are 802.11ac wave 2 capable. About 14% (554 WAPs) of the WAPs on the wireless network are Cisco 2700 series models, which are 802.11ac wave 1 capable. The remaining 73% (~2,920) are a variety of older Cisco WAPs that are 802.11a/b/g/n capable.

The University of Central Florida operates three primary wireless networks and a handful of specialized networks for particular use cases. The primary wireless user network, which is used by all students, faculty, and staff, uses WPA2/enterprise authentication. This primary network uses the SSID UCF_WPA2 and provides internet as well as LAN (UCF VRF) access for authenticated users. UCF also uses a Guest network with a captive portal that requires an acknowledgement for access. The Guest network only provides users internet access via the Guest DMZ, and only certain layer 4 ports are made accessible on the Guest wireless network. Finally, UCF also provides Eduroam access for students and faculty.

It is worth noting that the University's current wireless network has some limitations for providing IoT access and would require the use of a Network Access Control (NAC) and adequate licensing to do this. As of the writing of this document, the primary enterprise wireless network, UCF_WPA2, can only support WPA2/Enterprise authentication and resides in the general use LAN network (UCF VRF). Specialized requirements to employ IoT would likely require multiple building network refreshes and new systems purchases. One final consideration is that many of the WAPs located at buildings like Engineering 1 and 2 use the Cisco 3500, which will reach end of support on March 31, 2021. Furthermore, there are some locations, such as the CREOL building, that use Cisco 2600 WAPs, which will reach end of life December 31, 2021.

5.2 College of Medicine Case Study

Content in this section authored by Dr. Otto Phanstiel (Otto.Phanstiel@ucf.edu) and Jessica Madera (Jessica.Madera@ucf.edu)

Dr. Phanstiel is a full professor in the Department of Medical Education at the UCF College of Medicine. His specialty is medicinal chemistry, and his research focus is developing new therapies for human diseases such as pancreatic, breast, and colorectal cancers, and viral infections. His lab is also working to develop anti-aging medicines which increase organismal lifespan.

5.2.1 Science Background

Dr. Phanstiel's research mission is to discover new drugs to treat human diseases. As such the process of science involves conducting small molecule screens using specific human cell lines to determine which small molecule has the best activity in targeting a specific biological process. This molecule then becomes a "hit" where chemistry-based operations are performed to develop new derivatives of the initial hit. These derivatives are then tested for their ability to perform in the assay. The top performing derivative then becomes the lead molecule that will be moved into animal studies using mouse models of the human disease.

This research is based in the UCF COM Med Ed department[11] but collaborates with Dr. Deborah A. Altomare, from the Burnett School of Biomedical Sciences[12]. Our team is generating new molecular structures, cell-based data, and animal data, including histological sections of mouse tissue treated with our drug combinations.

Typically, the data is placed into a flat file (e.g., formatted as a spreadsheet) and stored using local resources. However, there is broad adoption (and available software licensing) of the CDD Vault infrastructure[13]. This site is used by major pharma collaborations and corporations to store and analyze both biological and chemical data.

Origin and custody of all research data is handled by Dr. Otto Phanstiel and Dr. Deborah A. Altomare's labs.

5.2.2 Collaborators

The collaboration space for medicinal chemistry is vast, and occasionally ad-hoc. It is regular to share molecular definitions with a large number of individuals and collaboration groups, anywhere in the world. The goal of the collaboration is often to test constructions and evaluate effectiveness.

Main collaborators include:

- Deborah Altomare: University of Central Florida, Orlando, FL
- Susan Gilmour: Lankenau Institute for Medical Research, Philadelphia, PA

---

[11] https://med.ucf.edu

[12] https://med.ucf.edu/biomed/

[13] https://www.collaborativedrug.com/benefits/

- Leah Vardy: Skin Research Institute of Singapore at the Agency for Science, Technology and Research (A*STAR), Singapore
- Jean Michel Brunel: Aix-Marseille Université, Marseilles, France
- Robert Casero: John Hopkins University, Baltimore MD

Scientific data sharing is not sophisticated, and often occurs via email attachment versus data transfer.

5.2.3 Instruments and Facilities

Medicinal chemistry involves a number of hands-on components to the research process, most of which are located local to the Phanstiel laboratory environment at UCF.

Present -2 years

Primary instrumentation currently consists of:

- Chemical fume hoods
- Nuclear Magnetic Resonance (NMR) spectrometer
- Biotek plate reader
- Radioactivity counter
- High Performance Liquid Chromatograph (HPLC)

This set of local instruments are used to generate/characterize molecules, and then evaluate activity in cell culture.

The addition of a local mass spectrometer (which will be used with the HPLC) will enable the ability to run and acquire LC-MS data sets. Data sizes are manageable, and can be processed for analysis using local computer resources (e.g., single workstations). Data management is handled locally as well.

The instruments produce a number of raw data formats and files:

- NMR data is output as a .fid file (e.g., free induction decay) and can be converted into a PDF that shows the spectrum analysis. Sizes are typically 1-10MB.
- Synthetic protocols and radioactivity tables are text, and captured using word-processor native formats (e.g., KB in size).
- LC data is captured as both a trace of the fluorescence, or absorbance chromatograph of the eluted molecules. The output of this process is typically a spreadsheet that can be analyzed (e.g., KB in size)
- Most forms of biodata are tabulated as a spreadsheet that can be analyzed (e.g., KB in size)
- Molecular modeling data is typically saved as pdb (Protein Data Bank) file types. These are typically formatted text files, and can vary in size from a few PB to several MB.

Next 2-5 years

I would like to move my data to CDD Vault to obtain a more easily searchable database with comparative capabilities.

Beyond 5 years
I would like to integrate this biochemical information with genomic information obtained via my other collaborators and the new UCF genomics core.

5.2.4 Process of Science

Due to the proprietary nature of medicinal chemistry research, most instrumentation and processing is done locally. The entire workflow (hypothesis, analysis, verification, storage, etc.) is done using UCF resources within the College of Medicine. This assists in the pursuit of funding, but also patenting findings.

Present-2 years
Use of localized resources for the entire workflow should not change during this time period.

Next 2-5 years
Instrumentation and research will be accomplished locally, but during this time use the Collaborative Drug Discovery (CDD) Vault will become wider spread. This will require sharing of small amounts of data.

Beyond 5 years
The process to share and manage drug discovery information, which is by nature closely protected and guarded until patented, will evolve over this time. It is anticipated that new approaches to facilitate publication and sharing will be available.

5.2.5 Remote Science Activities

At this time there are no remote science drivers or activities on this work. Instrumentation and analysis are located within the College of Medicine. This is not expected to change substantially over time, with the exception of new mechanisms to share and publish research data.

Present-2 years
The most prevalent "remote" use case for the Phanstiel laboratory environment is lateral, e.g., movement within the UCF administrative domain between departmental resources. In particular the College of Medicine and the Department of Chemistry have instrumentation that can be used for research activities. A common use case is to connect to resources in the UCF Department of Chemistry to download NMR data sets

Next 2-5 years

It is anticipated that more publication of research data will begin to be more common, namely the use of the CDD Vault. In this use case it is anticipated that MB sized data sets will be uploaded to the remote resource.

Beyond 5 years

Medicinal chemistry research could undergo several changes involving the handling and processing of research data during this time period. It is not known what impact this will have on the use of remote resources.

5.2.6 Software Infrastructure

Dr. Phanstiel's lab also performs molecular modeling using the ChemDraw[14], Chem3D[15], PyRx[16] and Pymol[17] programs. These programs generate protein data bank (PDB) files and can be graphics intensive in terms of rendering 3D images in a rotatable, zoomable, searchable landscape context.

5.2.7 Network and Data Architecture

Local resource usage dominates, with the majority of analysis and storage being accomplished on locally connected workstations, or employee mobile devices. Use of resources within the UCF Department of Chemistry, and College of Medicine are also common.

There are no plans to deploy computational or storage resources local to the Phanstiel laboratory environment at this time.

5.2.8 Cloud Services

Cloud services are currently not a critical part of any portion of the Phanstiel laboratory environment, beyond enterprise use cases (e.g., collaboration tools) provided by UCF. Migration to cloud services will be driven by external factors, e.g., as medicinal chemistry research utilizes more resources provided by industry and collaborations, cloud resources are a natural place to exchange and manage data streams.

Present-2 years

There are no cloud use cases expected during this time frame.

Next 2-5 years

Use of the CDD Vault will be a primary remote/cloud-based use case during this time window.

Beyond 5 years

---

14 https://www.perkinelmer.com/category/chemdraw

15 http://www.cambridgesoft.com/support/ProductHomePage.aspx?KBCatID=112

16 https://pyrx.sourceforge.io

17 https://pymol.org/2/

There is a strong desire to incorporate genomics data sources with the existing chemical and biological analysis that is being performed. Doing so would require adoption of a workflow that relies on remote data sources, and potentially analysis pipelines that are provided via local or remote computing. Specifics are not known at this time.

5.2.9 Known Resource Constraints

The data set sizes for medicinal chemistry are not significant, thus there is not a critical need to increase network or storage capabilities to support this use case.

5.2.10 Parent Organization(s)

The UCF College of Medicine provides IT support as needed, which unburdens individual groups from having to maintain services individually.

5.2.11 Outstanding Issues

Single sources of failure can be challenging to address, particularly during the COVID era when more remote work is being accomplished while still using college-based resources. A recent occurrence was an outage that prevented external access to research data sets.

While addressed quickly, this period of non-availability stalled progress, and points to the need to ensure there are multiple ways to access critical campus resources in the future.

5.3 Archeology & Anthropology Case Study

Content in this section authored by Dr. Scott Branting (Scott.Branting@ucf.edu)

Dr. Branting is an archaeologist with specializations in the ancient Near East and geospatial science. He directs the Kerkenes Project[18] in central Turkey, an enormous ancient city that was built around 600 BC by the Phrygians of King Midas fame and destroyed around 547 BC during the rise of the Persian Empire under Cyrus the Great. The Kerkenes Project seeks to understand this ancient city, and aspects of other cities by comparison, through excavations, remote sensing, and advanced simulations. Dr. Branting is also involved in using satellite images to monitor cultural heritage sites from space, and has worked on archaeological projects around the world.

5.3.1 Science Background

The Branting laboratory has several research focus areas:

- Archeology
- Biological/Physical Anthropology
- 3D model reconstruction
- Photogrammetry
- Geography
- Transportation Systems

Much of this research is funded through grants awarded by the National Science Foundation (NSF), National Endowment for the Humanities (NEH), the United States State Department, the United Nations Educational, Scientific and Cultural Organization (UNESCO), the University of Chicago, and several foundational sponsors.

There are a number of active and future projects:

- Simulation and modelling of city destruction using data collected on site in Turkey
- Study destruction of heritage and cultural sites
- Analysis of glass artifacts from an archeological site[19]
- Wide-scale pedestrian transportation in terms of city operation (e.g., escape routes)
- Automated detection of new buildings, agricultural fields, and human activity[20]
- Satellite imagery, with a focus on collecting data from around the globe in relation to cultural heritage[21]

5.3.2 Collaborators

The Branting laboratory features collaborators that are located around the world. These include research universities and governmental agencies.

---

18 https://sciences.ucf.edu/anthropology/kerkenes/

19 Collaboration with Dr. Sudipta Seal, https://mse.ucf.edu/person/sudipta-seal/

20 Collaboration with Dr. Mubarak Shah, https://www.crcv.ucf.edu/person/mubarak-shah/

21 Collaboration with Dr. Adrienne Dove, https://sciences.ucf.edu/physics/person/adrienne-dove/

- Universities:
    - USA
    - University of Chicago
    - Boston University
    - Arizona State University
- Australia
    - University of Sydney
- Netherlands
    - Leiden University
- Turkey
    - Istanbul University
    - Istanbul Technical University
    - Mustafa Kemal University
    - Abdullah Gül University
    - Middle East Technical University
    - British Institute at Ankara
    - American Research Institute in Turkey
- Government:
    - United States State Department
    - Argonne National Laboratory (ANL)
    - Oak Ridge National Laboratory (ORNL)
    - National Geospatial-Intelligence Agency
- At UCF:
    - Dr. Sudipta Seal – Materials Science and Engineering
    - Dr. Mubarak Shah – Center for Research in Computer Vision
    - Dr. Adrienne (Addie) Dove – Physics, Astronomy
    - Dr. Lori Walters[22] – History, Institute of Simulation and Training (IST)
    - Dr. Joseph Kider[23] – Institute of Simulation and Training (IST)

5.3.3 Instruments and Facilities

The Branting laboratory relies on a number of instruments to complete the process of science. These include a number of commercial products, as well as newly developed technologies created by collaborators (hardware and software based). Some of these include:

- Magnetometers
- Electrical resistance devices
- Hot air balloons & drone technology (to carry sensors and cameras, that assist with site surveys)
- Large scale GPS (million and half points for GPS)
- HoloLens (for use in an augmented reality project funded by NEH)

[22] https://womenfaculty.afia.ucf.edu/profile/lori-walters/

[23] https://www.ist.ucf.edu/People/Joseph-Kider

5.3.4 Process of Science

There are two scientific workflows that will be profiled:

5.3.4.1 Simulation and Modelling of City Destruction

The Branting laboratory participates in research that is attempting to simulate and model city destruction using data collected at an archeological site in Turkey. During this research, the summer months are used as the data collection phase, while the rest of the year models are generated based on collected data.

- ***Data Collection:***
  - Data is collected in summer and then it is brought back to USA via flights;
  - Data is transferred in from the field back to US by the team carrying USBs on flights
  - This strategy includes team members taking separate flights with the same replicated data to ensure data redundancy in case of some failure
  - About 1 TB of data is collected per summer
- ***Remote Sites:***
  - Archeological site in a rural area in Turkey; Site at about 4 hours from the nearest major urban locality
- ***Types of Data:***
  - Generating point clouds; geospatial datasets; radar, GPS, images, wind, solar radiation; drone photographs
- ***Model Generation:***
  - Computers slow in Turkey, therefore, for processing data is brought back to US for processing
  - Data is processed in Fall and Spring to create models
  - Then based on feedback from these models, new data is gathered next summer
  - This is inconvenient as ideally real time processing is needed for developing models so that excavation can be redirected within the same data collection phase; the summer being the data collection phase while fall & spring being the data processing phase is not ideal at all
  - Makeshift Beowulf clusters were once planned for this project
  - Cloud computing support is needed for this project
- ***Data Batch Size for Running Models:***
  - While data collected is 1 about TB per summer but the batch size to run a model can be done with even a week's worth of collected data
- ***Example:***
  - Ancient gate was excavated; 1000 drone photos and 2000 hand photos; photos data size in hundreds of gigabytes; A 400x400 meters 3D model created from this data; A weeks data chunk to run and update the model

5.3.4.2 Study of the Destruction of Heritage and Cultural Sites

The Branting laboratory is involved in a project, sponsored by US Department of State and UNESCO, that is studying destruction of heritage and cultural sites by the Islamic State.

- ***Data Collection:***
    - Use of declassified data that is made available from US Department of State resources
    - Data must be downloaded to local resources at UCF for analysis.
- ***Remote Sites:***
    - Primarily from the Middle East and Egypt.
- ***Types of Data:***
    - Data can be very large; satellite data; lots of images
    - Data is not classified but is sensitive (ethical & policy restrictions/implications)
- ***Data Sensitivity:***
    - Even though the data may not be classified (e.g., it consists primarily of maps and photos for cultural heritage in locations that include Syria), there are sensitivities to be aware of. In some cases, making this information public means that anyone can exploit the information, thus there is extreme care placed on data security
- ***Collaboration & Funding:***
    - Funding is approximately $3 million USD
        - Joint project with Dr. Mubarak Shah for structural findings and the development of algorithms
        - Branting lab is focused on data curation and SME of the research.
    - There are issues of handling this policy & access and of coordination
    -
- ***Known Areas of Friction:***
    - There are known issues of handling policy & access with the sensitive data, and how it can be coordinated.

5.3.4.3 Satellite Based Study of Heritage Sites

The Branting laboratory is planning for future work that focuses on leveraging satellite data collected from around the globe to support the study of cultural heritage.

- ***Data Collection:***
    - The goal is to collect remote sensing data of cultural sites from satellite resources (directly by communicating with base-stations, and indirectly through collaborating agencies).
- ***Remote Sites:***

    - UCF is parting with Arkansas University which has access to a base station. This functions by communicating with the satellite that is deployed in the ionosphere.
    - The “downlink” is located near the Arkansas University campus, and data would flow from there to UCF resources.
- ***Data Management:***
    - Space required to support data storage could be large (e.g., several TB a day).
    - Initially data will be saved at UCF, but will need to be shared with other collaborators
    - Digitalglobe[24] services can be used to store, and then allow others to access and download data.
- ***Collaboration & Funding:***
    - Project is in collaboration with Dr. Adrienne Dove, UCF Physics Department & Arkansas University
    - Project proposal is in progress, with a discussion expected by Fall 2021.

5.3.5 Remote Science Activities
There are two main categories of remote activity in this research:
- Data collected remotely via field research
- Data gathered from remote instruments and resources

Field research is an important part of the Branting laboratory, and comprises several months of activity on a yearly basis. Due to the remote nature (and lack of telecommunications support in certain regions) it is not feasible to develop a fully automated workflow for real-time data analysis. This leads to a slower process of science, where all remotely gathered data cannot be processed until it is physically received back at UCF after a field exercise.

Using data from remote instruments (e.g., satellites) or repositories is also a regular occurrence. In these scenarios there are well-defined access methods: e.g., interfaces provided by collaborators, websites/databases that can be queried). Once data is brought to UCF for analysis, it will reside there till the process of science completes.

5.3.6 Software Infrastructure
A number of software products are used to assist with the research projects in the Branting laboratory. These include tools that assist in the interpretation and creation of maps and models, processing survey and photo data, or just interpreting output from instrumentation:
- ArcGIS (Satellite license)[25]

---

[24] https://discover.digitalglobe.com
[25] https://www.arcgis.com/index.html

- QGIS[26]
- Custom software for instruments e.g., magnetometers

5.3.7 Network and Data Architecture

The network architecture for UCF is described in Section 3.1.5. This section will highlight some of the challenges for supporting remote field work and processing of sensitive data.

5.3.7.1 Simulation and Modelling of City Destruction

During field study, the Branting laboratory team deploys to a remote region in Turkey, far from telecommunications infrastructure to support any form of analysis. Therefore, all research must be done locally using a smaller set of available tools (e.g., instruments that gather data must output to local storage on laptops and external drivers). Without network connectivity, the data is physically transported back to UCF before analysis.

- ***Data Size:***
    - Data collected from agroecological sites using a variety of instruments
    - 30TB of field data has been collected across 28 years
        - Some is redundant; therefore, it might end be being 20TB
        - 1TB of data added per field trip
    - 1 field study per year (in the summer)
- ***Data size of output models (GIS):***
    - Several GBs
    - Sparse Mesh
    - Produced after field work completes; desire to have this created during work to guide progress
- ***Storage:***
    - Data stored in local servers (moved from Chicago to Orlando Data Center)
    - Orlando Data Center is maxed out
    - Data after capacity is being retained in USBs (not ideal)

Due to local capabilities, networking would need to be provided by satellite connectivity, and portable storage, processing, transmission, and analysis framework. If this capability were available, it would be possible to construct a workflow to provide near-real time capabilities to influence research progress.

5.3.7.2 Study of the Destruction of Heritage and Cultural Sites

The acquisition of data used in this research is done via interacting with remote repositories hosted by collaborators and government agencies. The sensitivity of

[26] https://qgis.org/en/site/

the data can be challenging to work with, which motivates the need to develop a more scalable solution:

- Central location that meets appropriate security control thresholds that can be used to download and store sensitive data
- Mechanism to enforce access to particular research groups
- Appropriate process to review access and handling of data over time.

5.3.8 Cloud Services

Cloud-based resources are being explored to support a number of use cases. Sharing data with collaborators requires a solution that will facilitate seamless data sharing. Additionally, the use of storage and processing resources that would simplify the workflow for the field work (e.g., ability to upload/process during the field work, instead of after) would be desirable.

5.3.9 Known Resource Constraints

These will be addressed by project area:

5.3.9.1 Simulation and Modelling of City Destruction

A large team collaborates on this research project, and they are highly distributed around the world (e.g., USA, Australia, Netherlands, Turkey). Therefore, centralized & synchronized storage, connectivity/speed for data transfer, real-time computing, and policy/access issues remain a challenge to be implemented:

- ***Compute***:
    - At the current time, there are no dedicated computational resources available. Shared resources exist, but converting the workflow to use them has not been prioritized.
    - Time to science is drastically affected when in the field. A mechanism to either travel with reduced computational resources, of facilitate a remote workflow, is desirable
- ***Network***:
    - Site in rural part of Turkey; Site is about 4 hours away from any major urban locality;
    - DSL speeds are available now, but would be unable to keep pace with GB of data produced daily.
    - Satellite phones have been historically too expensive
- ***Data Transfer***:
    - Ideally data should be transferred through a medium other than physically carrying USBs
    - Data should be transferred during the stay in Turkey so that researchers can take live feedback from models form more informed excavating
- ***Data Storage***:
    - Orlando (UCF) Data Center is maxed out, 25TB limit enforced for research group.

Some data is still being kept on portable drives (USBs) because of this issue
    - The Data Center is not up to scale for this project
- ***Data synchronization***:
    - Data synchronization between collaborators is not accomplished in real-time, which is an impediment to progress.
    - Synchronized database needed;
    - A team in Australia is trying to help with this
- ***Policy/Access issues***:
    - At UCF level the access had been handled; but access becomes an issue at a global level because of so many collaborators

5.3.9.2 Study of the Destruction of Heritage and Cultural Sites

Data management is a primary issue, due to the nature of the research:

- Data is not classified but is sensitive (ethical & policy restrictions/implications)
- Students should only have access to data related to their work
- There are issues of handling this policy & access and of coordination
- A solution that offers the ability to:
    - Manage the policy on a per-data and per-user basis
    - Review by trained team on a regular basis

5.3.9.3 Satellite Based Study of Heritage Sites

This project is still under evaluation, but anticipated problems include:

- Data sizes over time
- Ability to store/process efficiently
- Ability to share with collaborators

5.3.10 Outstanding Issues

There are several barriers to research progress that can be addressed:

- Computational resources available to UCF faculty
    - ARCC resources not directly connected to Orlando Data Center. This impacts data mobility and storage
    - No method to handle sensitive data
- Data transfer performance in/out of UCF
- Computational availability for near-realtime use cases
- Data synchronization and sharing with collaborators
- Standard way to define and manage access policies for data within and external to UCF

5.4 SREAL Research Laboratory Case Study

Content in this section authored by Dr. Gregory Welch (welch@ucf.edu) & Ryan Schubert (rschuber@ist.ucf.edu).

Gregory Welch has appointments in the College of Nursing, the Department of Computer Science, the Institute for Simulation & Training, and is a Co-Director of the Synthetic Reality Laboratory (SREAL) at the University of Central Florida's Institute for Simulation & Training. He conducts research in areas including virtual and augmented reality, human-computer interaction, human motion tracking, and human surrogates for training and practice, with a focus on applications such as healthcare and defense.

Ryan Schubert is the SREAL Lab Manager and Special Projects Coordinator, and supports a variety of research in that capacity. His specific research interests include computer graphics, virtual reality, augmented reality (including spatial augmented reality), and software to drive augmented virtual environments.

5.4.1 Science Background

SREAL is a research lab at the University of Central Florida. The SREAL team consists of faculty researchers, affiliated faculty members, software developers, PhD students, artists (modelers/animators), interactors (digital puppeteers mostly associated with the TeachLivE project), undergraduate research assistants and a network of campus and institute collaborators, both faculty members and students. SREAL is part of several larger UCF entities, most notably the Institute for Simulation & Training (IST), which houses it, and the Department of Computer Science.

The SREAL laboratory carries out fundamental research in areas including:

- Human Computer Interaction (HCI)
- Human Interaction with Social Robots
- Virtual, Mixed, and Augmented Reality (VR, MR, and AR)
- Embodied Virtual Agents
- Physical Virtual Avatars
- Digital Avatar Technology
- Spatial User Interfaces
- VR Technology (e.g., Head-Mounted Displays and Tracking)

Application areas include healthcare, education, defense, and law enforcement. Examples of organizations supporting SREAL research include the National Science Foundation, the Office of Naval Research (ONR), and the Department of Defense (DoD).

5.4.2 Collaborators

Collaboration includes a number of UCF faculty and facilities, as well as external partners. The following are some examples.

- Universities:
  - Stanford University
  - University of Florida

- Corporate:
    - SoarTech[27]
    - SA Photonics[28]
- Government:
    - Army Research Laboratory (ARL): Dr. Celso de Melo
- At UCF:
    - Prof. Gerd Bruder[29]
    - Dr. Joseph LaViola[30], Dr. Roger Azevedo[31], Dr. Carolina Cruz-Neira[32], Dr. Mindi Anderson[33], Dr. Frank Guido-Sanz[34], and Dr. Laura Gonzalez[35]

5.4.3 Instruments and Facilities

SREAL has a number of instruments and devices located in the research facility. A breakdown appears in the subsequent sections.

5.4.3.1 Facilities

- HuSIS (highly instrumented CAVE-like [Cave Automatic Virtual Environment] interaction space) and associated human subject user study infrastructure (isolation booths, etc.)
- Reconfigurable open lab space (can be temporarily partitioned for development, demonstration, or experiment setups including specialized hardware, tracked spaces, supporting computing devices, and workstations)

5.4.3.2 Instruments

- AR and VR headsets
- Worn biometric sensors (e.g., wristwatch-like devices)
- IoT devices
- Networked mobile devices (e.g., cellphones)
- Security cameras
- Tracking systems
- Camera-based motion capture systems

5.4.3.3 Specific Devices

- Oculus Rift & Oculus Rift S
- Oculus Quest
- HTC Vive & HTC Vive Pro

[27] https://soartech.com
[28] https://www.saphotonics.com
[29] https://sreal.ucf.edu/people/bruder/
[30] https://www.cs.ucf.edu/person/joseph-j-laviola/
[31] https://ccie.ucf.edu/profile/roger-azevedo/
[32] https://www.cs.ucf.edu/person/carolina-cruz-neira/
[33] https://nursing.ucf.edu/people/mindi-anderson/
[34] https://nursing.ucf.edu/people/frank-guido-sanz/
[35] https://nursing.ucf.edu/people/laura-gonzalez/

- Magic Leap One
- Microsoft HoloLens & Microsoft HoloLens 2
- SA Photonics SA-92/S
- InterSense InertiaCube4
- Stereo Labs ZED Mini
- Arduino
- Raspberry Pi
- Intel NUC

5.4.4 Process of Science

The SREAL facilities and instruments are used primary for three purposes:
- Human-subject research studies
- Development/construction/testing of research studies or device prototypes
- Tours and/or demonstrations of research

HuSIS is a dedicated space for studying human interactions[36] [37], and is based on CAVE technology. Typically, this can be 1-3 participants of a study (larger groups are possible during tours). In addition to the main HuSIS, a transportable version (e.g., THuSIS) can be used for natural-habitat experiments to contrast against those that would be conducted at SREAL.

The devices that are created and profiled within SREAL produce a wide variety of data, including:
- Audio
- Video
- Biometrics
- Encoded behavior
- Head/body tracking data
- Eye tracking data

Additionally, human subjects may be involved in producing and curating:
- Self-reported questionnaire responses
- Studies on task proficiency/speed

Due to some privacy implications, IRBs are used for data that is shared. Most research data is collected and operated within SREAL, although a number of activities feature external data sources, for example:
- A collaboration with UCF Police where data was collected in the UCF Student Union
- Data collected from lobby of a UCF campus building (e.g., Partnership-3) over a period of multiple months; behavior coding was done in real-time with a human in-loop for doing the coding

---

[36] https://www.cs.unc.edu/~welch/media/pdf/Schubert2016aa.pdf

[37] https://ieeexplore.ieee.org/document/7750515

Specific facility usage is dynamic and opportunistic based on changing funding research areas, and specific grants, as well as shorter-term needs to support student projects, specific research prototypes, and specific user studies.

5.4.5 Remote Science Activities

Most data SREAL uses is generated locally to the laboratory environment, or within campus (UCF) collaborations. Some instances require the use of external data sets, or the sharing of local data sets with external collaborators. Planned research projects would include collecting certain external experimental data at UCF, and then sharing that data externally.

5.4.6 Software Infrastructure

Software used by SREAL includes specific tools that interact with instrumentation, statistical and modeling software, human-subject focused tools (e.g., surveys) and tools used to storge, mange, and curate locally-developed tools. Some examples include:

- OptiTrack Motive[38] (optical tracking and motion capture)
- Integrated environments for data and code, e.g., Unity[39] (custom scripts as well as third party plugins)
- Shared repositories (e.g., SVN, Git)
- Qualtrics[40], Google Forms[41], etc. for surveys/questionnaires
- Amazon's Mechanical Turk[42]
- MATLAB[43], SPSS[44], Excel, etc.
- Customized C/C++ code to collect and manipulate experimental data, and to control experimental prototypes.

5.4.7 Network and Data Architecture

The network architecture for UCF is described in Section 3.1.5. The SREAL research routinely involves networking. Several example use-cases include:

- Peer-to-peer live data connections (e.g., remote/networked control of virtual characters)
    - This use case requires low latency and high bandwidth to support a more immersive experience
- Local wireless networks/LANs where one or more node also needs WAN access (e.g., for externally hosted survey)

---

[38] https://optitrack.com/software/motive/
[39] https://unity.com
[40] https://www.qualtrics.com
[41] https://www.google.com/forms/about/
[42] https://www.mturk.com
[43] https://www.mathworks.com/products/matlab.html
[44] https://www.ibm.com/products/spss-statistics

    - Security requirements to protect the devices, yet allow them to interact with the external world, are required
- Network attached storage
    - To facilitate data gathering and sharing within the environment

5.4.8 Cloud Services
Cloud computing and storage are not being actively employed at this time. There are several cloud-use cases that are being considered:
- Cloud services such as the ability to process NLP (natural language processing)
- Certain IoT type devices interface to cloud services automatically (e.g., Amazon Echo), and thus require external access methods

5.4.9 Known Resource Constraints
At this time, the following resource considerations should be considered:
- HPC is not an active part of the process of science. To support some 3D models and data processing, local CPU and GPU resources used.
- Data transfer external to UCF is not a current use case. Efficient data sharing within the SREAL environment (e.g., to support VR use cases) is required.
- External collaboration data sharing needs are relatively infrequent, but when they arise, they are often ad hoc in nature.
- Secure data sharing (e.g., ITAR controls) is occasionally needed on specific grants and collaboration groups.

5.4.10 Outstanding Issues
There are several desirable features that would assist with the research process.

1. Setting up computational environments can be a slow and problematic process.
    a. Causes include access policy issues, hardware differences, lack of common software configuration, or use of non-production software.
    b. These factors limit the team's ability to react to immediate needs and opportunities. This can be paralyzing for research.
2. Affiliated with the first item, the turn-around time for changes to SREAL resources controlled by campus or IST staff can be very slow. This is exacerbated by the lack of a common operating environment, and non-dedicated resources for the job.
3. Because the environment is managed centrally by UCF, it is uncommon to allow "local admin" on lab machines. Hardware/software research development frequently involves performing tasks which require admin privileges (installing third party software, specific hardware drivers, low-level hardware access for connected devices, etc.)
4. The aforementioned policy issues hinder usage of heterogeneous devices.
5. The use of IoT devices on a public network poses implementation challenges for SREAL and UCF.

    a. UCF wants to maintain a secure environment that can be monitored and defended.
    b. SREAL needs an easy to use and rapidly reconfigurable environment that allows for the seamless addition of devices and ways to monitor their behavior.
6. Access policies for networks (e.g., "plug and play") have become more complex over time. This is exacerbated by funding requirements (e.g., DoD) that emphasize security.
7. There is a need for controlled environments, but the controls must be implemented in such a way to not hinder usability and performance. This relates to devices, software, user behavior, and data.

5.5 Department of Physics Case Study

Content in this section authored by Dr. Eduardo Mucciolo (Eduardo.Mucciolo@ucf.edu).

Dr. Mucciolo is the previous Department of Physics Chair at the University of Central Florida. His work combines analytical modeling and computer simulations. Some projects are developed in collaboration with colleagues at UCF and elsewhere and include:

- *electronic transport in nanostructures, molecules, and low-dimensional materials*
- *quantum entanglement in many-body systems*
- *physics-inspired algorithms and methods for solving hard computational problems*
- *applications of tensor networks in physics and in computer science*

5.5.1 Science Background

Research within the Mucciolo lab deals with theoretical topics in two distinct areas:

- quantum information processing
- condensed matter physics

Both focus areas involve analytical and numerical modeling and calculations utilizing computational resources. Recent efforts have taken place to develop tensor network methodologies to compute properties of quantum many-body systems (often relevant to quantum computation and to condensed matter physics), as well as developing and employing techniques to compute properties of electronic systems at nanometer scales. While the goal of the former is to understand fundamental new phenomena in quantum many-body systems, for the latter the focus has been more applied; for instance, to help design new interconnects in microchips.

Collaboration within UCF is limited to a few computer science and experimental physics-focused colleagues, but does include several external partners (e.g., Boston University, and some others listed below) for both areas of research.

Data set size is not large when compared to other areas of physics (e.g., experimental work in High Energy or Astrophysics areas). The lab considers itself "small data", in the sense that data is very expensive to generate and not much of it is required to accomplish research goals. 10 years of research data could fit easily into commodity storage (e.g., < 1TB). Data is created by running numerical calculations and simulations. Because problems of exponential complexity are typically modeled, if something can't be figured out with 10 or 100 GB, the next step would be to generate and store tens or hundreds of terabytes, which is unrealistic. Thus, the group focuses on quality and efficiency, and not on data quantity, because that does not typically bring any substantial advantage.

As a result of the small data sets, transfer and sharing has historically not been a factor. Transfers can be accomplished using relatively small connections (e.g., 1Gbps) and using standard tools (cloud storage, email, etc.).

The main research product produced by the lab is not the data itself, but the process to generate them (e.g., the creation of computer codes). The data analysis that is employed is rather simple from a complexity perspective, consisting mainly of computing statistical distributions and moments of these distributions for a variety of values of model parameters. Often, all the information can be distilled into a few scattered data plots or graphs; sophisticated plotting tools are not required in most cases.

5.5.2 Collaborators

Current collaborators include:

- limited collaboration within the Physics Department at UCF
- limited collaboration with the Computer Science Department at UCF
- Physics Department at Boston University, Boston MA
- Université de Sherbrooke, Québec, Canada
- University of Massachusetts Boston, Boston MA
- Northeastern University, Boston MA
- Universidade Federal Fluminense (UFF), Rio de Janeiro, Brazil

Collaboration sizes can be individuals, or small groups (less than 6 people).

It is not common to share raw data for this type of work. When data is generated at UCF (or other facilities), it is analyzed and the results of the analysis are typically shared with all parties. The valuable output is not the data, but the methods and computer codes that are employed to generate the data. Information about the methods (notes, papers, presentations) is mostly shared via email or Dropbox. Data analysis results can be shared via Skype, email, Dropbox.

5.5.3 Instruments and Facilities

The Mucciolo lab has access to a Linux cluster which consists of:

- 17 dual-core (RAM 2-8 GB) nodes
- 16 quad-core (RAM 4 GB) nodes
- 1TB in storage capacity

Age of this resource has impacted operational status, and nearly half of the resources are not functional. Despite this, the cluster is still used for some analytical work, along with Stokes at the UCF ARCC. The former is preferred given there is no queue to launch jobs, where the latter sometimes has a wait time. Certain software packages can only run Stokes due to license requirements, as well requiring larger amounts of memory.

The group has some emerging use of remote resources at Northeastern University that is facilitated via a DOE-sponsored project.

Upgrading the local computation within the lab is not feasible for a number of reasons. Funding to support the initial capital investment is possible (e.g., the aforementioned grant that is jointly researched with Northeastern University could facilitate purchase of some resources), but long-term support of the hardware and operating systems is not available via UCF IT. Additionally, a prior grant facilitated some use of cloud computing (e.g., Amazon Web Services), which worked well. It is possible this could be leveraged in the future, if funding was available.

The lab often pursues many funding sources, but has refrained from requesting funds to purchase more computing power for a few reasons:

1) there are no resources on campus to support non-centrally located scientific computing hardware or Linux computer systems;
2) Physics funding sources/program managers tend to think that computing resources are better used at the university or national center level;
3) The lab has made a conscious choice to be involved with projects that require computing resources that match current capabilities. E.g., leveraging higher capacities resources implies spending more time on development of theoretical models, computational methods, and software, rather than focusing on extensive number crunching and data processing.

5.5.4 Process of Science

The primary output of research for the Mucciolo lab is the creation of numerical simulations and calculations on computational resources (both at UCF, and located at collaborating locations). A typical workflow is:

- design the project
- discuss methodologies
- Program/create simulations
- Execute the simulations
- analyze the data against the expected outcomes
- turn the results into a publication

Data analysis needs tend to be simple. In models with no disorder, or that are not generated out of a statistical ensemble, the numerical calculations are just one single shot with each set of values for the input parameters providing a result (e.g., a number or a curve on a plot). The data analysis performed by checking the numerical values or the curves describe experimental data or fit into a prediction based on more general considerations. Often the numerical results hint to what experimentalists should try to check or reproduce in their labs.

When models are generated from ensembles, data sets are generated where a distribution of values of a certain quantity are computed for members of this ensemble (say, from hundreds to tens of thousands). Then distributions are plotted and moments such as average and variance are computed. The results are checked to see if they fit into an analytical prediction, explain experimental data, or provide clues for future experiments.

Most of the research done at the Mucciolo lab emphasizes concepts, even when backed by numerical and computer calculations or simulations. The numbers per se are often not important, unless they provide some general or universal law, such as a scaling power. Only recently have there been attempts to create numerical calculations where the end result is a number that is more descriptive of a physical property that can be directly measured (e.g., the electrical resistance of a nanometer metallic interconnect).

These conceptual exercises are extremely important for physics and other sciences. Most notable awards in condensed matter physics are given to people who came up with new concepts, and not to those who computed something with high precision.

5.5.5 Remote Science Activities

The primary uses of remote resources are related to sporadic off-site computational use:

- Amazon Web Services (e.g., cloud computing) for a prior grant
- Northeastern University computing cluster.

Use of these external resources is typically only done when the local resources (with the Mucciolo lab or at the UCF ARCC) are not feasible.

5.5.6 Software Infrastructure

With very few exceptions, the lab writes all the computer codes that are used for research purposes. These codes are heavy on linear algebra, and for that well-known and tested libraries are used, such as LAPACK[45]. High performance is the most important aspect, and freely available compilers for languages such as Fortran, C++, and Python are primarily used.

For plotting and some basic data analysis, freely available options are preferred, such as gnuplot[46], matplotlib[47], grace[48], etc.

All data transfer and remote access are done via scp and ssh, respectively. Data, notes, plots, are typically shared via Dropbox.

Commercial/proprietary software (e.g., Mathematica, Matlab) are typically not used. In rare instances, such as recent work where electronic structure calculations were required, they may use specialty software such as Quantum Espresso or VASP[49]. For VASP, an institutional license was already available via Stokes at UCF.

5.5.7 Network and Data Architecture

The network architecture for UCF is described in Section 3.1.5. No additional resources are required due to the minimal data movement requirements.

---

45 http://www.netlib.org/lapack/
46 http://www.gnuplot.info
47 https://matplotlib.org
48 https://plasma-gate.weizmann.ac.il/Grace/
49 https://www.vasp.at

5.5.8 Cloud Services

Enterprise cloud services, such as Dropbox for sharing notes, data, codes, and results are used frequently.

In the recent past, AWS was used for an investigation of tensor network contractions (a non-hierarchical computing problem) in the cloud (a hierarchical system, thus the challenging aspect of the project). Funds from an NSF grant were used to pay for it, as well as time freely granted by AWS.

The Mucciolo lab does not have plans to use cloud computing again primarily due to cost. The experience was not bad directly, e.g., availability of resources and ease of use for the environment were both positive. The costs were mainly in student time to develop and test, before being able to produce output. For this reason, the cloud is still not a primary use case that will be used.

5.5.9 Known Resource Constraints

The main constraints are:

- The Mucciolo lab local Linux cluster is very old and dying
- There is no available Linux support on campus for local computing resources
- The computing resources provided by the university are good but oversubscribed

These problems are not new for those that require computation, and are hard to solve. For the research performed in the Mucciolo lab, computing is a tool for conceptual understanding, as well as for making accurate predictions of phenomena. For the latter, more computing power is required.

5.5.10 Parent Organization(s)

IT at UCF is currently highly centralized. In prior years, IT support was mainly done at the department level (12 years ago), or the College level (9 years ago), to now being completely commanded by a central university entity. As a result of these shifts, meaningful support to scientific computing was lost. The same technicians who support enterprise IT are employed (to a lesser extent) in supporting scientific IT needs. Most have little or no knowledge of scientific computing, which impacts their ability to support fine grained departmental needs. Researchers are mainly left to fend for themselves, and this has impacted enthusiasm for applying for/operating computing infrastructure outside of the central facility. Central IT does maintain a focus on cybersecurity and compliance at the top level, but there is not much consideration for what people actually need to get research done.

Repairing the relationship between IT and faculty should be a priority, particularly in understanding needs and providing resources to address them.

5.5.11 Outstanding Issues

Beyond the aforementioned support issues for research IT, there are not any problems related to data transfer/sharing or the lack of access to modern collaboration and productivity tools.

5.6 Networking and Wireless Systems Lab (NWSL)

Content in this section authored by Dr. Murat Yuksel (murat.yuksel@ucf.edu).

Dr. Murat Yuksel leads the Networking and Wireless Systems Laboratory (NWSL). This research touches various topics in computer communication networks encompassing both experimental and theoretical aspects. NWSL gives particular focus on concepts that keep data networks up and running with high efficiency as well as design and development of wireless systems capable of standing in the new modern mobile world. Some of the research topics include:

- *Networked, wireless and computer systems*
- *Optical wireless*
- *Spectrum sharing*
- *Cloud networking*
- *Network economics*
- *Network architectures*

5.6.1 Science Background

NWSL's research efforts touch various topics in computer communication networks encompassing both experimental and theoretical aspects. NWSL gives particular focus on concepts that keep data networks up and running with high efficiency as well as design and development of wireless systems capable of standing in the new modern mobile world. Several projects are interdisciplinary spanning areas such as behavioral science, math, economics, and physics. These projects involve use and generation of various datasets that can be in the forms of bare/un-curated data, anonymized data, and application domain data. Traffic traces, network monitoring logs and network use patterns are some examples of research efforts. These experimentations and theoretical models generate various datasets that can be used by other researchers as well.

5.6.2 Collaborators

NWSL's collaborators include a variety of locations and science backgrounds. These include:

- North Carolina State University: Engineering
- Florida International University: Engineering, Computer Science
- Virginia Tech: Computer Science, Engineering
- University of Nevada, Reno: Math, Computer Science, Behavioral Science
- New Jersey Institute of Technology: Computing, Network Science
- Rensselaer Polytechnic Institute: Engineering, Computer Science
- University of Alabama: Engineering
- University of Central Florida: Public Administration, Transportation Engineering, CREOL, Psychology
- Virginia Commonwealth University: Computer Science, Engineering
- Missouri University of Science and Technology: Engineering
- UC Riverside: Computer Science, Engineering
- University of Illinois at Chicago: Computer Science, Engineering

Various datasets and research facilities are shared with these collaborators. In most cases, this is sharing of research results or curated datasets that are used for research analysis and design.

5.6.3 Instruments and Facilities

Computation and networking capabilities are the most common resources for NWSL. These include:

- Running extensive (discrete event) simulations that use large memory and computation power. The best environment for these is compute clusters with strong compute nodes and lots of storage space.
- Network emulations that involve many virtual machines (VMs). The best setup for these is distributed computing clusters with strong compute nodes where we can run multiple VMs. A common file system is necessary, to where many VMs can write their results, resulting in tens of thousands of folders and files.
- Network hardware that is customizable and can be configured in a variety of ways, e.g., software-defined networking (SDN) routers, high-performance switches, software-defined radios (SDRs), and mobile devices.
- Embedded system platforms that can connect to the Internet and be part of large networking experiments.
- Licensed radio/cellular bands for testing and experimentation of wireless systems.

Since some of the above resources are not available at times, we also use the following facilities and testbeds that are funded by federal projects:

- GENI
- CloudLab
- EmuLab

5.6.4 Process of Science

A rough outline of a common NWSL workflow is as follows:

1. The process starts with a design idea to solve a research problem.
2. The NWSL will:
    a. design an experiment
        i. identifying metrics to look at
        ii. identify parameters to change in order to observe the trends in the identified metrics
    b. A common theme is: "what is the simplest experiment that can be designed to prove the new idea is better than others?"
3. With this mindset, the NWSL will identify facilities and testbeds that will be the most suitable for the experiment
    a. If NWSL can leverage existing resources for the experiment, this approach will be taken.

    b. If not, NWSL will purchase materials and supplies to set up the experiment.
4. NWSL will:
    a. Run
    b. Test
    c. Collect results from the experiment
    d. Very frequently, the experiment will be run again (with different parameters) after analyzing the results and observing what is missing.
5. NWSL will gather insights from the repeated experiments and disseminate the results to the research community
    a. In some cases, NWSL will provide the results as a curated dataset

5.6.5 Remote Science Activities
The following remote facilities are primarily used:
- GENI
- CloudLab
- EmuLab

Other federally funded facilities are being explored for future use cases (e.g., FABRIC). In addition to those, NWSL regularly uses research computing clusters within the UCF campus.

NWSL will expand the use of remote resources to emerging PAWR testbeds:
- COSMOS – located at NYU and Rutgers
- AERPAW – located at NCSU

5.6.6 Software Infrastructure
NWSL uses the followings for data management:
- Networked storage services: OneDrive, Dropbox
- AWS
- Github
- Overleaf

5.6.7 Network and Data Architecture
The network architecture for UCF is described in Section 3.1.5. No additional resources are required due to the minimal data movement requirements.

5.6.8 Cloud Services
NWSL regularly uses cloud computing environments such as AWS, EC2, and Google Cloud. Thus far, NWSL has utilized no-cost services from these environments, but this has limited what work can be accomplished in developing outcomes.

3.6.9 Known Resource Constraints
NWSL needs the capability of having separate/isolated traffic on the campus network and/or beyond. Several of the experimental ideas require separate

treatment of the traffic to attain real-time control of the experiment parameters and application setup.

NWSL has a strong research need to acquire licensed cellular bands for radio spectrum sensing and wireless system testing.

NWSL has a strong research need for computing and storage capability to provide data repositories (with on-demand computation) to the community.

5.6.10 Parent Organization(s)

NWSL does not require additional support from UCF at this time to assist with research activities.

5.6.11 Outstanding Issues

No outstanding issues are relevant to NWSL.

6 Discussion Summary

During discussion, with the University of Central Florida, the following points were emphasized:

- *Investing in more powerful compute and storage:* The requirements for storage and computation at UCF will grow in the coming years, and span to more use cases beyond traditional users (e.g., physics, computer science, engineering). Investing in upgraded infrastructure is recommended:
    - Research storage allocations that can scale to PB ranges
    - Availability of high-performance (e.g., massively parallel) and high-throughout (e.g., grid/distributed computing) resources
    - Availability of GPU resources, to facilitate other use cases (e.g., ML, AI, NLP).
    - Knowledgeable staff to convert and manage scientific workflows
    - Integrations with on-site instruments
    - Workflow optimized data transfer hardware and software to facilitate sharing with collaborators
    - Federated ID to streamline user management, and also facilitate integration with other regional or national resources.
    - Availability of infrastructure geared toward Controlled Unclassified Information (CUI).

- *Increased Support for research with compliance needs (classified/sensitive data):* A number of researchers could benefit from access to special infrastructure that is capable of working with Controlled Unclassified Information (CUI), e.g., security controls prescribed by NIST SP 800-171 Rev. 2. Resources exist in limited quantities on campus for select groups; investing in a centrally managed and expandable enclave for the entire campus would be a worthwhile investment considering the use cases and expected growth in the coming years. UCF is currently investing in a central managed and expandable enclave for the entire campus and should continue to increase its support for that considering the use cases and expected growth in the coming years.

- *Cyberinfrastructure for resource-constrained field-research:* Some research activities performed at UCF are performed in the field (e.g., at remote locations without dedicated research networking or computation). For those groups that require technology assistance in the field, there are multiple options to consider:
    - Using satellite connectivity (e.g., investing in equipment and purchasing service) to allow traveling groups a mechanism to enable remote connectivity. Workflows can be created that allow for semi-regular uploads of data via a laptop and satellite modem back to UCF. Once data is received, an automated workflow to perform any processing tasks can be initiated.

    - A ‘traveling’ set of computational and storage capabilities can be created for transport to remote field locations. This could take the form of a padded shipping container that contains storage and processing that can be used when on site. While not as powerful as HPC/HTC resources, it would eliminate the need for network connectivity.

- *Support for decentralized resources:* A number of departments and research labs still maintain research computing and storage infrastructure that is independent of the centralized UCF ARCC. Efforts to centralize management over the years have resulted in a human resources and knowledge gap in maintaining these aging components. This poses a significant risk for the departments: trying to keep the infrastructure alive takes considerable time, and not being proficient in IT means the work is accomplished poorly (e.g., it becomes a security risk over time). UCF Research IT should consider mechanisms to either support this infrastructure. Mechanisms include working with UCF to leverage investment in research grant "overhead" to fund IT investments (e.g., storage, computation, or networking), offering to take it over, or decommissioning it while giving the research groups dedicated resources within the UCF ARCC in exchange.

- *Automate scientific workflows:* UCF research IT should consider investments into human capital resources that have knowledge of workflow automation for scientific use cases. These staff would assist researchers in creating software infrastructure that better uses UCF ARCC resources (e.g., containers, data movement approaches, and automated backups). This would unburden researchers from needing to manage low level details of research infrastructure, and would also facilitate a more secure mechanism for systems and software.

- *Backup storage mechanism:* Reliable backup mechanisms (e.g., storage, or just the ability to run containerized resources at a different location) is highly desirable in the event of a system maintenance or unexpected failure. UCF Research IT could work with independent department or laboratory IT staff to create backup operational planning mechanisms.

- *Scientific software support*: The process to identify, procure, evaluate, and operate particular scientific software packages can be cumbersome. This leads to reduced productivity as research groups wait for approval, or may otherwise be unaware of the time expectations. Level setting and documenting the process will lead to better outcomes for both IT support and the research community.

- *Network performance monitoring:* Monitoring internal and external networking status is critical for a number of groups and use cases on campus. E.g., low latency matters in supporting video and virtual reality, other use cases require high-bandwidth access to external sites (e.g., off-site data transfer and computing). To ensure proper operation, UCF IT should consider methods to monitor critical system requirements (e.g., through perfSONAR) so action can be taken on discovery of the problem.

- *Support for research with specialized networking needs:* Creation of network testbeds for research activities has been identified as a potential need. Doing so has the following benefits:
    - Segregate research (e.g., IoT devices) away from general purpose networks to protect them from malicious intent
    - Allow greater control (e.g., by research teams) over the network testbed to simulate conditions as needed
    - Allow greater monitoring capabilities
    - Prevent general purpose users from infringing on research networks

- *Access to national research testbeds:* Access to national scale research testbeds (e.g., CloudLab, Emulab, GENI, and FABRIC) is a high priority, provided that the resources remain usable and available for the target research use cases. Often national-scale testbeds are not responsive to researcher needs, which forces local groups to investigate ways to recreate the resources locally.

- *Maintaining institutional knowledge related to national and regional CI capabilities:* The first step to using national resources (e.g., XSEDE, OSG Software, etc.) is being aware that they exist, and the steps required to apply for and adapt them to a scientific workflow. It is recommended that UCF Research IT dedicate staff resources to documenting and keeping pace with these national resources as a benefit to the UCF research community.

6.1 University of Central Florida Campus Overview

This case study reviews the work of Barry Weiss, Glenn Martin, Joey Netterville, Henry Glaspie, Shafaq Chaudhry, and Fahad Khan

Technology Summary:

The UCF network supports 13 colleges and a multitude of programs for over 70,000 users. To accomplish this, there are roughly 1,200 network devices in service, using primarily Cisco as the vendor of choice for the enterprise network, and the vendor Extreme for the residential and research networks. The network is currently undergoing a redesign with an emphasis on service provider designs and services. Future designs will include a move from traditional Enterprise Campus design to routed segmentation using a mixture of L2VPNs and L3VPN MPLS architecture. FLR (Florida Lambda Rail), an independent statewide research and education fiber optic network, is currently the sole WAN provider for UCF and operates the DWDM ring in Central Florida that connects the UCF main campus and other remote sites. UCF's Research Network was implemented in 2013 based on funds from CC*IE grants. The primary goal of the research network is to provide high-speed access for researchers and equipment to campus resources. The Research Network also provides a direct link from the campus to Internet2 resources.

Under the VP for Research, there exists a Research IT team that provides research facilitation and cyberinfrastructure support for researchers. This team works with researchers and provides a number of services such as consultation on storage, computation, and networking needs along with identifying local and remote resources that may be useful for the process of science.

Through a strong collaborative partnership among the UCF Office of Research, IST (School for Modeling, Simulation, and Training), UCF Information Technology, and the UCF Information Security Office, a managed environment for handling sensitive and restricted research data is available. "Knight Shield" offers compliance with NIST SP 800-171 and Cybersecurity Maturity Model Certification (CMMC Level 3) for handling Controlled Unclassified Information (CUI) and can be used for supporting Federal Information Security Modernization Act (FISMA), and Health Insurance Portability and Accountability Act (HIPAA), Export Controlled (ITAR/EAR), and other contractually regulated research. The Knight Shield environment currently consists of an on-premises network infrastructure and an AWS cloud environment.

UCF is home to the Advanced Research Computing Center (ARCC). This is capable of housing a variety of high-end computational resources, including existing infrastructure such as the Stokes High Performance Computing cluster. The ARCC is located in the Central Florida Research Park adjacent to the UCF Orlando campus. The ARCC is a full member of Florida's Sunshine State Education and Research Computing Alliance (SSERCA).

6.2 College of Medicine Case Study

This case study reviews the work of Dr. Otto Phanstiel, a professor in the Department of Medical Education at the UCF College of Medicine (COM).

Science Summary:

Dr. Phanstiel's specialty is medicinal chemistry, and his research focus is developing new therapies for human diseases. The overall research mission is to discover new drugs to treat human diseases. Resources from the UCF COM are used primarily, along with some at the UCF Burnett School of Biomedical Sciences via local collaborations.

At the current time, there are no significant challenges related to data sharing or management (e.g., data set sizes are small, and use commodity methods such as email for dissemination). In the coming years, this is expected to change as the collaboration space increases, the machinery generates more fine-grained data sets that require more computation and storage, and the use of collaboration tools (e.g., central repositories to facilitate sharing) are more widely adopted.

Discussion Summary, Findings, & Actions:

- Dr. Phanstiel's research requires the use of Controlled Unclassified Information (CUI), e.g., security controls prescribed by NIST SP 800-171 Rev. 2[50]. Some resources are available at the UCF COM, but not within the broader UCF research computing infrastructure. Working to establish these would provide a critical backup for COM researchers, as well as making new capabilities for others in the broader UCF community to use. This may include bespoke storage, computing, networking, and staff that are versed in information compliance.
- Working closely with the UCF COM, or other quasi-independent UCF departments and schools that have their own IT support infrastructure, would help to create backup infrastructure that can be used in the event of scheduled maintenance or system downtime.
- UCF should invest more human capital into the role of IT support, particularly to work closely with researchers that are looking to adopt new technology, automate processes, or operate local hardware/software.
- The process to identify, procure, evaluate, and operate particular scientific software packages can be cumbersome. This leads to reduced productivity as research groups wait for approval, or may otherwise be unaware of the time expectations. Level setting and documenting the process will lead to better outcomes for both IT support and the research community.

[50] https://csrc.nist.gov/publications/detail/sp/800-171/rev-2/final

6.3 Archeology & Anthropology Case Study

Dr. Scott Branting is an UCF archaeologist, with specializations in studying the ancient Near East regions and use of geospatial science approaches.

Science Summary:

Dr. Branting directs the Kerkenes Project in central Turkey, an ancient city dating to 600 BC, with research goals of understand through excavations, remote sensing, and advanced simulations. Additional work also centers on the use of satellite imagery to monitor cultural heritage sites from space.

Due to the remote nature of the research, field work is critical and common. Collection of data (images, video, etc.) is typically done on site for short periods of time, but the lack of an efficient mechanism to transfer the remotely gathered data, and perform near-real time analysis using computational resources, limits progress. Data volumes will continue to increase as the digital record of the field work grows, as well as the need to share with collaborators (domestic and international).

Additionally, some research data borders on requiring Controlled Unclassified Information (CUI) security controls. There are not adequate facilities to handle this information at the current time.

Discussion Summary, Findings, & Actions:

- Due to the remote field work aspect of Dr. Branting's work, access to reliable network connectivity is not common. For this reason, use of satellite internet connectivity should be evaluated to judge if the performance and cost can be justified. Prices for equipment and service have been reduced over prior generations, with new options emerging now and in the coming years[51].
- Dr. Branting's work requires a certain amount of real-time capability to be effective when they are operating at a remote field site. There are two ways this can be addressed:
    - Using the aforementioned satellite connectivity, a workflow can be created that allows for semi-regular uploads of data via a laptop and satellite modem back to UCF. Once data is received, an automated workflow to process the raw images into 3D models can be initiated. Dr. Branting and team can then use the same satellite connectivity to retrieve the produced models when they are complete.
    - A 'traveling' set of computational and storage capabilities can be created for transport to remote field locations. This could take the form of a padded shipping container that contains storage and processing that can be used to produce 3D models from research images. While not as powerful as HPC/HTC resources, it would eliminate the need for network connectivity.

[51] https://www.satelliteinternet.com/providers/starlink/

- UCF research IT should consider investments into human capital resources that have knowledge of workflow automation (as alluded to in the prior recommendation). These staff would assist researchers in creating software infrastructure that better uses UCF ARCC resources (e.g., containers, data movement approaches, automated backups, etc.).
- Dr. Branting's research could benefit from access to Controlled Unclassified Information (CUI), e.g., security controls prescribed by NIST SP 800-171 Rev. 2[52]. At the current time, there are no resources at UCF (beyond specialized to departments like COM) that are capable of this distinction.
- The requirements for storage and computation at UCF will grow in the coming years, and span to more use cases beyond traditional users (e.g., physics, computer science, engineering). Investing in upgraded infrastructure is recommended:
    - Research storage allocations that can scale to PB ranges
    - Availability of high-performance (e.g., massively parallel) and high-throughout (e.g., grid/distributed computing) resources
    - Knowledgeable staff to convert and manage scientific workflows
    - Integrations with on-site instruments
    - Data transfer hardware and software to facilitate sharing with collaborators
    - Federated ID to streamline user management, and also facilitate integration with other regional or national resources.

52 https://csrc.nist.gov/publications/detail/sp/800-171/rev-2/final

6.4 SREAL Research Laboratory Case Study

Gregory Welch & Ryan Schubert supports a variety of research in the Synthetic Reality Lab (SREAL) at the University of Central Florida's Institute for Simulation and Training (IST)

Science Summary:

Research at SREAL includes virtual and augmented reality, human-computer interaction, computer graphics, human motion tracking, as well as human surrogates for training and practice with a focus on applications such as healthcare and defense.

SREAL research activities are mostly confined to the local, and do not have many use cases that require data mobility. This being said, data generation can be large (e.g., storage of video and images), and there are several low-latency requirements when dealing with augmented reality environments. Leveraging UCF computing and storage environments is preferred, although local resources are typically used. A particular area of friction is the management of multiple computational environments to support specific software needs. Use of approaches to reduce system management time (e.g., containerization) could help, but would require access to IT staff that can assist, and thus far virtual machine alternatives (for example) have not proven to be viable in most instances.

Some aspects of the research require more control over a network environment than is currently possible. For example, when researching Internet of Things (IoT) devices, there is a balance that must be found between ease of deployment and operation of the devices within a network environment, vs. protection and availability of the resource. For this reason, the use of dedicated testbeds that are functionality similar, yet administratively different, than the public internet should be considered.

Discussion Summary, Findings, & Actions:

- UCF research IT should consider investments into human capital resources that have knowledge of workflow automation (as alluded to in the prior discussions). These staff would assist researchers in creating software infrastructure that better uses UCF ARCC resources (e.g., containers, data movement approaches, automated backups, etc.). This would unburden researchers from needing to manage low level details of research infrastructure, and would also facilitate a more secure mechanism for systems and software.
- The creation of a wireless 'testbed' for use of IoT devices should be investigated. This network would have several benefits:
    - Segregate IoT devices away from general purpose networks to protect them from malicious intent
    - Allow greater control (e.g., by SREAL team) over the network testbed to simulate research conditions as needed

    - Allow greater monitoring capabilities
    - Prevent general purpose users from infringing on research networks
- Low latency/high availability requirements (e.g., to support video or virtual reality) is critical to SREAL. To ensure proper operation, UCF IT should consider methods to monitor critical system requirements (e.g., through perfSONAR) so action can be taken on discovery of the problem.
- The requirements for storage and computation at UCF will grow in the coming years, and span to more use cases beyond traditional users (e.g., physics, computer science, engineering). Investing in upgraded infrastructure is recommended:
    - Research storage allocations that can scale to PB ranges
    - Availability of high-performance (e.g., massively parallel) and high-throughout (e.g., grid/distributed computing) resources
    - Availability of GPU resources, to facilitate other use cases (e.g., ML, AI, NLP).
    - Knowledgeable staff to convert and manage scientific workflows
    - Integrations with on-site instruments
    - Data transfer hardware and software to facilitate sharing with collaborators
    - Federated ID to streamline user management, and also facilitate integration with other regional or national resources.
    - Availability of infrastructure geared toward Controlled Unclassified Information (CUI).

6.5 Department of Physics Case Study

Dr. Eduardo Mucciolo is the previous Chair of the Department of Physics at the University of Central Florida. His work combines analytical modeling and computer simulations.

Science Summary:

Some projects are developed in collaboration with colleagues at UCF and elsewhere and include:

- electronic transport in nanostructures, molecules, and low-dimensional materials
- quantum entanglement in many-body systems
- physics-inspired algorithms and methods for solving hard computational problems
- applications of tensor networks in physics and in computer science

Both focus areas involve analytical and numerical modeling and calculations utilizing computational resources.

Discussion Summary, Findings, & Actions:

- There are a number of non-centrally controlled computational resources that exist on campus owned and operated by research groups. The centralization of IT support has resulted in a lack of staff available to assist in the long-term maintenance of these resources, which places extra burden on researchers. This leads to an overall reduction in capability, and is a security and productivity risk for UCF researchers. UCF research IT should consider ways to increase staff to address management of these resources, or consider migrating them to a central location and giving 'credits' to researchers who do so to ensure they have enough computational cycles to complete their research effectively.
- UCF research IT should consider investments into human capital resources that have knowledge of workflow automation (as alluded to in the prior discussions). These staff would assist researchers in creating software infrastructure that better uses UCF ARCC resources (e.g., containers, data movement approaches, automated backups, etc.). This would unburden researchers from needing to manage low level details of research infrastructure, and would also facilitate a more secure mechanism for systems and software.
- The process to identify, procure, evaluate, and operate particular scientific software packages can be cumbersome. This leads to reduced productivity as research groups wait for approval, or may otherwise be unaware of the time expectations. Level setting and documenting the process will lead to better outcomes for both IT support and the research community.

6.6 Networking and Wireless Systems Lab (NWSL) Case Study

Dr. Murat Yuksel leads the Networking and Wireless Systems Laboratory (NWSL), and the case study presented research related to the use and simulation of computer communication networks encompassing both experimental and theoretical aspects.

Science Summary:

A primary use case for the NWSL research is the ability to utilize testbeds for network simulation purposes. The majority of this work is done external to UCF, using national-scale resources (e.g., CloudLab[53], Emulab[54], GENI[55] , etc.). Some use cases can be accomplished using on-campus resources, and more may be explored in the future as they become available.

Some national resources (e.g., DARPA's Colosseum[56]) are not as user friendly, and have been challenging to make contact with. Similarly, as national-scale efforts shift resources or near end of life (e.g., GENI) the ability to run complex simulations becomes harder. In situations where a national resource fits a research need, but is not usable, it is desirable to explore creation of the functionality locally if possible.

The group will continue to investigate ways to scale using national scale efforts when they become available. FABRIC[57] is one example that will be explored in coming years.

Discussion Summary, Findings, & Actions:

- The use cases presented for NWSL do not require special support from UCF at this time, as most of the access to external resources follows R&E networking paths. It is worth investigating if QoS guarantees would improve latency expectations for sensitive experiments within or external to campus.
- NWSL is interested in acquiring licensed cellular bands for radio spectrum sensing and wireless system testing.

---

[53] https://www.cloudlab.us

[54] https://www.emulab.net/portal/frontpage.php

[55] https://www.geni.net

[56] https://archive.darpa.mil/sc2/the-colosseum/

[57] https://fabric-testbed.net

Appendix A - FLR Regional Networking Diagram