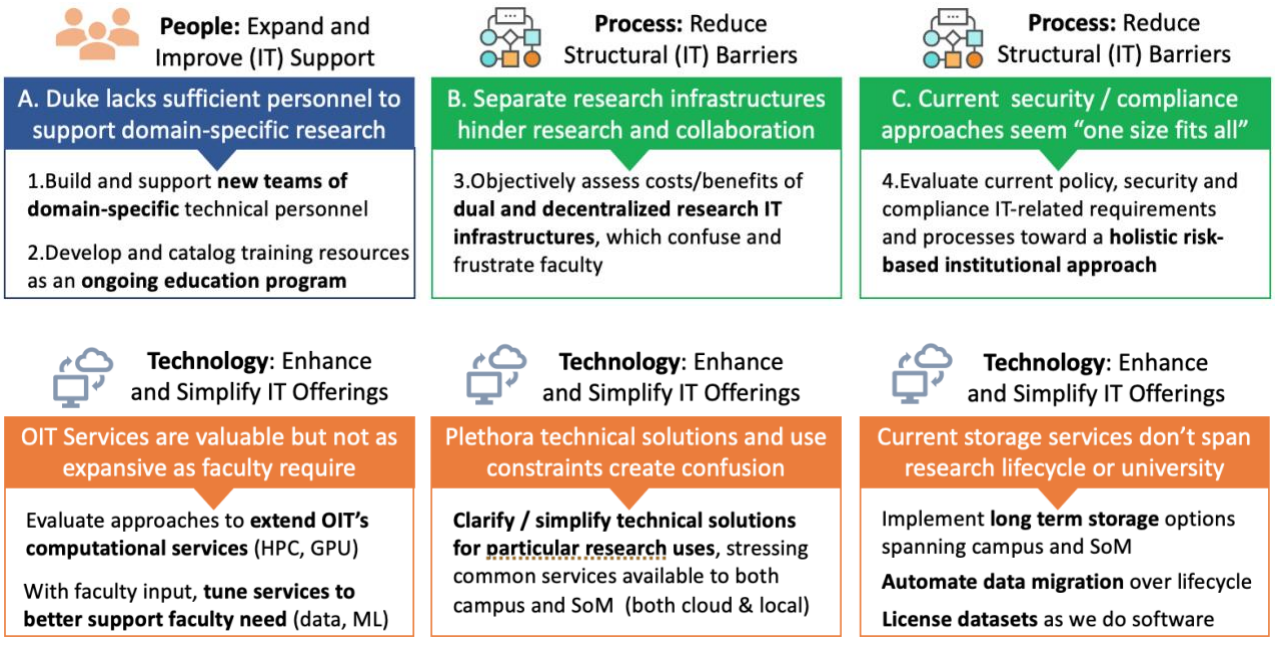
Executive Summary

Duke's Information Technology Advisory Council (ITAC) is the primary faculty advisory body responsible for guiding Information Technology (IT) decisions at Duke, principally those undertaken by the Office of Information Technology (OIT). ITAC has served this function since its founding in 1995 and as a very active Council (meeting bi-weekly), ITAC has weighed in often on matters of policy, strategy, and operations.

Beginning in February 2022, ITAC launched a comprehensive assessment of research IT needs across Duke faculties, with a goal of identifying the actual support needs and service gaps that exist today in meeting the needs of Duke's modern, computationally intensive research enterprise. The assessment was motivated by a recognition that research methods and domains had changed dramatically over the years, while IT services and support structures tended to change more incrementally. In collaboration with staff support from OIT and in consultation with cognizant deans for natural sciences, basic (life) sciences, social sciences, humanities, arts, and engineering, teams of faculty were convened to share feedback on their experiences and the ways in which support and services could be improve. These faculty teams (which also included a few teaching/research staff) were identified as a reflection of the range of scholarly work across Duke (excluding clinical research) rather than to mirror school-by-school representation.

Over the next six months, each domain faculty/research staff team met, and its major discussion points were summarized, documented, and brought forward for more detailed discussion and examination with ITAC. Final findings and recommendations were developed, and the process culminated in October 2022 with six findings and ten recommendations, categorized into the areas of People, Process, and Technology and summarized in the following graphic. Findings are contained in the six colored boxes, with the associated recommendation(s) in the white boxes below each. The report's ten *process* recommendations are designed to progress these six findings, and the basis for each are conveyed through this report.



After feedback is received from University Leaders and Deans, it is anticipated that leaders from concomitant units (Research, IT, Libraries, etc.) will collaborate to develop proposals aimed at addressing these findings from the services perspective.

Duke University Research IT Needs Assessment Summary Report and Recommendations

Introduction

Duke's Information Technology Advisory Council (ITAC) in conjunction with staff support from the Office of Information Technology (OIT) has undertaken a comprehensive assessment of the research computing support and service requirements for campus researchers. Deans with responsibility over natural sciences, basic (life) sciences, social sciences, humanities, arts, and engineering have each selected groups of faculty to represent their own information technology (IT) needs, as well as those of their colleagues. These faculty members provided feedback on how to improve technology offerings and services to better support research needs, as well as the increasingly complex compliance and regulatory environment around the production, use, and preservation of research data. See Appendix A for a full list of the participating 39 faculty/research staff.

From February to August 2022, an OIT project team held listening sessions with seven groups of faculty and researchers in

- Natural Sciences
- Social Sciences
- Basic Sciences
- Engineering
- Humanities/Arts (Three Working Groups)
  Digital Environments, Digital Representation, Digital Pedagogy

After each listening session, a summary was prepared by the faculty participants (Appendix B) and shared with ITAC for discussion and elaboration on common needs (Appendix C). Two "poster" sessions were held in June and August to distill faculty feedback (Appendix D) into the top priorities and needs (Appendix E). Through these appendices, the interested reader can be more fully immersed in the process as it progressed.

Findings and Concerns

All seven sessions began with a focus on technology concerns, and while the conversations highlighted improvements that could be made to technology-related services, each of the groups also noted the need for specialized support in the different disciplines and training, workshops and consulting for use of existing and new techniques or technologies (often referred to as "domain expertise"). As one faculty member noted, "A lot of things that look like computational problems are not, they should be done by experts in the field."

Faculty representatives noted concerns about who is paying for what services and technologies—Big Duke or a department, or is it left up to the faculty member or a grant? In addition to costs and access to domain-specific support, faculty are also concerned about the ways in which organizational structures and policies may impact their conduct of research, such as processes associated with research administrative tasks such as completing data management plans or grant proposal development, submission and reviews (especially those involving sensitive data). This also encompassed concerns about the confusion or frustration that can be faced when navigating our complex ecosystem of inter-related, or sometimes siloed, offices and processes.

This document presents the summary observations of the working groups along with general recommendations to progress each observation (finding) in the categories of:

> **People** - Identify skill sets and support staff investments needed to support research IT needs.
> **Processes** – Evaluate existing or implement new processes to align administrative tasks with research needs. This category also encompasses observations related to IT policies or organization structures.
> **Technology** - Prioritize the expansion of an existing or addition of a new technology services offering to help meet research needs.

Summary observations and recommendations below include links to the fully detailed findings (deeper insights from each group), which appear in Appendix F. Those details are an essential aspect of the process and integral to this shortened report but are shared via an appendix so as not to overwhelm or exhaust the reader.

Collectively, findings and recommendations reflect a need for new models and sustained support, not simply a one-time infusion of funds or point-in-time set of changes.

People Findings and Recommendations

Consistent across all the seven faculty working groups was the need for domain specific expertise. The OIT engagement team facilitating the research review had anticipated this unmet need from the humanities, arts and social sciences, but also noted the strong expression of need for research personnel support that was communicated by engineering, natural sciences and basic sciences. Most units praised their local IT support and general OIT services, but also communicated the need for specialized support personnel who could more fully engage the central IT structures, technical offerings, and policies, as well as offer training assistance. A few exemplar comments related to personnel appear below; a complete list appears in Appendix F (page 24).

- "The biggest need is for human infrastructure." (Hannah Jacobs in ITAC discussion with Humanities/Arts)
- There is a need for more specialized/dedicated local support personnel who can quickly and efficiently address problems as they arise, without the need to bounce a ticket back and forth with others. Current ticketing process and search for expert help is arduous and inefficient. (Basic Sciences Faculty Working Group)
- People are important and some degree of domain knowledge is also really important—what physical chemists need, or what particle physicists need, or what genomicists need—they're not really the same thing. Compared to some of our peers we are under-resourced in that kind of domain computational expertise. The funding models of our peers are all over the map, from fully allocated to heavily charged out, but just in pure body count we have fewer of those people than many of our peers do right now. (ITAC discussion with Natural Sciences)
- An important thread is that Social Sciences (and others) have issues around involving novice graduate and undergraduate student researchers in projects. This is may not be an OIT issue per se, but it's an important layer of our digital ecosystem that can get lost. (Paul Jaskot in ITAC discussion with Social Sciences)

Finding A: The focus of OIT's "Research Computing" function as the prevailing embodiment of IT support for research established a robust and responsive infrastructure comprising networking, computation, and storage supporting general needs. However, it did not build a base level of personnel support for research endeavors that rely on data collection, analysis and storage or the application of technology to specific scientific domains. By leaving the human component to the lab, department, and school level, the result has been inconsistency in specialized research support as conveyed by all faculty working groups, even those with substantial extramural funding. See Appendix F (pages 24-27) for detailed insights reflecting all seven faculty working groups and in support of these recommendations.

> *Recommendation 1: In consultation with faculty, Duke should devise a service that builds and supports teams of domain-specific technical personnel who are well versed in both on-site and cloud options. The personnel should be equipped to assist with specialized needs in the different domains, including compute, storage, data collection and analysis, and new applications of technologies (e.g., applying machine learning, creating digital environments, and building and maintaining large, shared data resources). Evaluation of support structures should extend beyond department vs. central IT resources, also encompassing prospective engagement from areas like Libraries or other programs. The effort should start with pilots and expand with experience to take advantage of early lessons, include faculty feedback, and have central assistance in recruiting personnel.*

> *Recommendation 2: Develop and catalog training resources to support an ongoing education program targeting faculty and graduate students but also serving undergraduates who participate in research endeavors. Such a program should apply established approaches (e.g. data analysis using common applications such as R and emerging machine learning) and should leverage the domain-specific technology support resources, as well as existing co-curricular programs such as via Libraries (e.g., CVDS), Center for Computational Thinking (CCT), OIT (e.g., Co-Lab Roots), and various curricular programs offered by academic departments, and others.*

Personnel-related findings of interest to one or more groups, but not reflecting broad need, included:

- Centralized assistance for machine learning and deep learning applications like image analysis are needed (Natural Sciences)
- Include consultative support for grant preparation related to these Data Use Agreement requirements (Social Sciences)
- Re-evaluate basic technology support and research support levels and allocated resources in the humanities and arts and benchmark against programs at other leading universities. (Humanities/Arts)

Process and Organizational Findings and Recommendations

Duke University is a highly rated and well-respected research university, built upon the foundations of a liberal arts college and enriched by the incredible strength of the professional schools, including the Schools of Medicine (SoM) and Nursing and notable others, as well as the deeply technical disciplines represented by Engineering, Natural Sciences and Social Sciences. However, the impact of Duke multi-disciplinary academic and research potential is severely blunted by different or siloed organizational structures, inconsistent or competing strategic directions, and conflicting policies that restrict or make difficult the collaboration that undergirds all scientific discovery. Throughout all groups we heard these concerns noted, but most vocally from Basic Sciences and Natural Sciences. Sample comments related to Process and Organization appear below; a full list appears in Appendix F (page 27).

- Seams are beginning to show in the overall research IT service model. There is a growing lack of understanding between OIT and SoM IT providers as to what services are provided by whom and how to acquire them. (Natural Sciences Faculty Working Group)
- Better mechanisms are needed for communicating IT users' needs/requirements to the policy makers and financial managers. ITAC is an excellent mechanism for this on the University side, but there is need for a comparable avenue in SoM. (Basic Sciences Faculty Working Group)
- We need a better connection of local support personnel (with domain knowledge) to the larger IT support system, be it DHTS or OIT. Today we have a system where we file online tickets and although everybody has the best intention, sometimes it's hard to find the person who's qualified and has the security clearance to actually implement the solution and oftentimes it comes back to us. It's not super-efficient when you have to interact with 6 people, and they are doing their best with everybody trying to locate the right person. I would love a simple system where we know directly who knows enough to solve the problem and maybe could have the security clearance to fix my problem. (ITAC Discussion with Basic Sciences)

Finding B: Separate IT infrastructures supporting research have operated for more than two decades (OIT-Campus Schools, DHTS-Schools of Medicine and Nursing), along with distinct lab and departmental IT personnel. Despite each organization striving to deliver great service, and while OIT and DHTS coordinate on many items, the impact for faculty collaborating across these complex infrastructures for research has become increasingly problematic. For example, decisions of SoM research policy and support are well understood to be the purview of the School, however their impact can reverberate across campus in terms of research administration and policies as well as directly impact faculty in campus schools such as Pratt and Trinity. See Appendix F (pages 27-30), for detailed insights which were most vocally expressed by Natural Sciences and Basic Sciences, but also echoed by other faculty working groups.

Recommendation 3: Conduct an external review with peer expert leaders (from academic medical centers that are tightly coupled with their universities) to objectively assess costs/benefits of dual and decentralized research infrastructures (OIT/DHTS/Schools/Institutes); present for review to University leadership including the Provost, SoM Dean and other University and DUHS leaders. Regardless of the outcome of the review, seek to discontinue or drastically limit the practice of SoM policies and technologies being imposed on the campus without input from faculty.

Finding C: Disparate and sometimes conflicting compliance and security measures (Campus and SoM) and a seeming lack of gradations of risk for various data types and uses can inhibit intra- and inter-institutional research collaboration. While researchers acknowledge and accept the necessity to comply

with requirements to protect sensitive and restricted data, they desire approaches that better define options and balance the risk of actual data loss with the perceived adverse impact to Duke's research endeavor itself. See Appendix F (pages 31-33), as conveyed to varying degrees by the majority of faculty working groups.

> *Recommendation 4: In collaboration with faculty representatives, the VP for Research and Innovation together with the DU CIO should launch an evaluation of current policy, security and compliance requirements and processes for research toward a holistic risk-based institutional approach. The goal should be to look for minor adaptations to existing policies or seek new solutions (if warranted) in cases where the adaptations better balance research practices / needs with risk, but without compromising regulatory requirements or the protection of the institution against sensitive data loss. The effort should also reflect the need for nuanced policies or implementations for social science and campus researchers working with sensitive data that are not predicated or based on similar DUHS, School of Medicine, or School of Nursing clinical research policies or implementations, and to take care not to conflate clinical studies with other research endeavors. Ongoing research-related governance bodies should include sufficient faculty participation to incorporate representative feedback across the diversity of Duke's research enterprise and work towards increased transparency about the processes and decision making.*

Other Process and Organizational findings that had limited rather than universal support included:

- Communication and coordination are needed across the many oversight bodies as the process for the researcher has become complex, time-consuming, and a hindrance to research. (Social Sciences)
- Greater representation is needed from social scientists at the decision-making table, particularly regarding policies that impact Social Sciences research. For example, the updated Data Policy guidelines. (Social Sciences)
- Create support structures for team teaching in the humanities that may become a means of establishing initial 'connective tissue' among faculty and with IT (Humanities/Arts)

Technology Findings and Recommendations

While faculty across all working groups were aware of and praised the Duke Compute Cluster service and associated personnel, there was no shortage of concern expressed that the current offerings are too limited in terms of flexibility and scale of the compute offerings, the flexibility and cost of storage solutions over the lifecycle of research projects, and the technical guidance available to help discern and implement the best fit solution among local (on-premises) options and cloud approaches. These concerns were most pronounced—to a person—throughout Engineering, Natural Sciences and Basic Sciences, but were echoed by individuals on the Social Sciences and Humanities/Arts working groups. Sample feedback regarding Technology themed findings appear below; for a complete list see Appendix F, page 33.

- Many researchers are having to go outside Duke to peer institutions and national labs to use certain clusters; it would be a boon to the University and researchers alike to have the means to run small to medium scale jobs with a home cluster. (Natural Sciences Faculty Working Group)
- Several years ago, the Provost agreed to commit to covering 80% of all the computing needs of the faculty, recognizing the 20% of the most computationally intensive users would need to utilize grants or other approaches to meet those needs. Recognizing those needs may have changed over time, we should reassess whether the level of "entitlement" computing being provided by OIT still meets that 80% threshold. (ITAC discussion with Natural Sciences)
- There's an unmet need for mid-scale computing resources that would provide more than the Duke Compute Cluster but less than provided by a national lab. This includes fast interconnection and a diversity of hardware. (Engineering Faculty Working Group)
- NIH funded data storage retention requirements need to be recognized by both DHS and DU as a critical component of infrastructure. (Basic Sciences Faculty Working Group)

Finding D: The Duke Compute Cluster (DCC) and RAPID VM services, while quite valuable and cost effective, are not as expansive as many faculty require. For some, it is difficult or less attractive to use for certain teaching and specialized research tasks. On a continuing basis, and as our needs and technology

evolve, OIT should work with ITAC and academic leadership to evaluate and propose changes or additions to research computing services, and to enhance the guidance OIT offers to its users. See Appendix F (pages 33-36) as conveyed to varying degrees by all faculty working groups.

> *Recommendation 5: Evaluate the cost and funding approach to extend the capability of Research Computing by expanded HPC/GPU services so that they are better sized to meet demand and are easier to use; review with ITAC and academic leadership to determine viable approaches that can improve support and are financially feasible. Establish a more routine review cycle with ITAC faculty (or other faculty advisory bodies) to better evolve services with need.*

> *Recommendation 6: Establish one or more Faculty Working Groups to help guide OIT in tuning Research Computing offerings to better support data collection, statistical analysis, and machine learning, as well as common programming and graphics-intensive needs. New approaches need to be considered for the extremely data intensive needs of instrumentation, along with the financial models that can support them (funded via grants, indirect cost recover overhead, subvention, or otherwise).*

Finding E: Insufficient awareness exists about computational options and there is uncertainty of when to use which as many options are often presented and left to the faculty to choose. This lack of awareness, coupled with complexity in navigating the options, stymies researcher ability to easily access the resources they require. See Appendix F (pages 36-37), as conveyed to varying degrees by all faculty working groups.

> *Recommendation 7: Introduce guidance and simplify navigation of technical solutions for faculty and students by consolidating technical offerings to more broadly support research and academic use cases or needs (including labs with substantial data generation or acquisition from specialized instruments). Solutions should: encompass the full breadth of services available, including local and cloud options; be established as an evolving service; and be developed in consultation with faculty to 'field test' the efficacy of the guidance and navigation across the range of Duke's research domains.*

Finding F: The needs for research data storage have increased exponentially; current storage services address the need for active, near-term research storage but are not easy enough to use and share and do not incorporate solutions that are affordable for long-term and archival storage. Moreover, current options do not reasonably accommodate NIH and NSF 'unfunded mandates' for long-term data retention in ways that are either scalable or cost effective. See Appendix F (pages 37-39), as conveyed to varying degrees by all faculty working group (Recommendations 8 & 9) or as may represent an opportunity broadly for Duke but introduced most explicitly by Social Sciences (Recommendation 10).

> *Recommendation 8: Implement new long-term storage options for a variety of use-cases and ensure solutions are available to not only campus but also School of Medicine and Nursing researchers; establish funding models with leadership and faculty to ensure sustainability.*

> *Recommendation 9: Ensure the new long-term storage options along with existing and future storage solutions are approached as an eco-system, recognize the differences in storage needs across the research lifecycle and automating to the greatest degree the movement of the data across solutions to reduce the burden on faculty and risks to Duke. Make these options available to both campus and SoM users with attention to differences in datatypes and data storage needs in the science versus humanities versus social sciences.*

> *Recommendation 10: Evaluate the need and develop a proposal with estimated cost to establish a new service for researchers that would manage/license/administer shared access to expensive data resources/sets; carry out this work in collaboration with the prior Presidential Scholar who recently studied this need in greater detail on of the Provost and President and in conjunction with the Strategy Team 2030 implementation team.*

Other Technology findings that were less universal included:

- Provide additional physical computer labs (like Link Classroom 6) so students are not faced with using small laptop screens to access VMs for graphics-intense software like GIS. (Humanities/Arts)
- Provide faster interconnects for the DCC (needed to support certain kinds of research) (Engineering)
- A need for "bare metal" hardware environments, beyond what the DCC provides (Engineering)
- Better digital publishing services, especially extra-mural options w/ visualization + archive options; related, digital scholarship may require specialized tools and platforms for archiving. (Humanities/Arts)
- Establish clear guidance/direction for existing services for data collection, storage, and compute services that are available to researchers at no or low cost; clearly classify their appropriateness for use based on the sensitivity of the data or research being performed (Social Sciences, Basic Sciences)

Conclusion and Next Steps

To improve research IT support in an environment of ever-increasing requirements, complexity, and interdependence, will require a multilayered approach that focuses on incremental improvements to offer near-term relief while also moving forward on systemic issues. Broad input from Duke's faculty and researchers indicate the need to:

- **Simplify services** and reduce the amounts of time, effort, and money that researchers need to spend managing Information Technology and navigating among different offices or policies.
- Provide more **layered services** that match evolving needs of researchers based on acceleration of regulatory and technology requirements. (i.e., data management services vs. storage)
- Offer **domain specific services, protocols, and people** to better match evolving needs in different research domains. Universal across all sessions was a clear cry for more domain expertise from computational support staff.

Faculty urged repeatedly and vocally throughout the process that at least four areas of overall commitment are needed related to governance, financials, or service span in order to protect Duke from repeating the history that led to the current situation (a significant gap between research needs and support levels):

1. A faculty-centric oversight or governance body (or bodies) to help monitor and advise when and how services need to evolve, policies need to be established or revisited, etc.
2. A sustained budget commitment reflecting research support as a continuing and evolving need, requiring not just the one-time funds but also ongoing financial commitments and flexible financial models to address the changing needs and environmental factors (and research constraints) over time.
3. Recognition that future and ongoing services in support of research should avoid bright dividing lines or "firewalls" between Schools of Medicine/Nursing researchers and the rest of campus.
4. Clarity on overall process ownership of the organization(s) with responsibility, authority and accountability to sustain and evolve research support needs such as those characterized herein; when support or processes span multiple organizations, a commitment to a One Duke approach to solving.

An essential next step will be to share these findings with the greater Duke community to ensure feedback is consistent, validate that the recommendations (process oriented) are embraced by university and school leadership (Provost, EVP, Deans, and others), and identify and align partners who will help to implement recommendations, notably expected to include: VPRI, University Librarian, CIO, others, and the faculty bodies who will help to guide the implementation of recommendations (ITAC and others).

Following the validation process and report issuance, specific further steps are not proposed as part of this report. Rather, specific service and support proposals responsive to the report's findings and recommendations are expected to be developed over the coming months collaboratively among the CIO, VPRI and University Librarian.

The Information Technology Advisory Council, the Office of Information Technology, and the Office of the Vice President for Research and Innovation sincerely thank the many individuals who participated in the assessment project and contributed to this report.

Appendix A: Working Group Faculty Participants (Identified through Deans)

Natural Sciences (Identified by Mohammed Noor)

David Beratan, Chemistry
Eric Laber, Statistical Science, Biostatistics & Bioinformatics, and Global Health
Jianfeng Lu, Mathematics, Chemistry, and Physics
Dan Scolnic, Physics
Jenny Tung, Evolutionary Anthropology, Biology and DUPRI
Greg Wray, Biology, Evolutionary Anthropology, and Biostatistics & Bioinformatics

Social Sciences (Identified by Rachel Kranton and Don Taylor)

Kate Bundorf, Sanford School and Margolis Center for Health Policy
Sunshine Hillygus, Political Science and DISM (Duke Institute for Survey Methods)
Jim Moody, Sociology and DNAC (Duke Network Analysis Center)
Daniel Xu, Economics and Triangle Census Data Research Center
Chris Bail, Sociology and Duke Polarization Lab

Basic Sciences (Identified by Colin Duckett)

Lisa Cameron, Biology and Director, Light Microscopy Core
Alexandra Badea, Radiology, Neurology, Biomedical Engineering, BIAC
Greg Crawford, Pediatrics / Molecular Genetics and Microbiology
Lindsey Glickfeld, Neurosciences
David Herzfeld, Neurobiology
Danny Lew - Pharmacology and Cancer Biology

Engineering (Identified by Jerry Lynch and Claudia Gunsch)

Volker Blum, Mechanical Engineering and Materials Science
Cate Brinson, Mechanical Engineering and Materials Science
Jessilyn Dunn, Biomedical Engineering
Helen Li, Electrical and Computer Engineering
Miroslav Pajic, Electrical and Computer Engineering
Henry Pfister, Electrical and Computer Engineering
Amanda Randles, Biomedical Engineering, Computer Science, Duke Cancer Institute

Humanities/Arts (Identified by William Johnson, Paul Jaskot, John Brown and Victoria Szabo)

Digital Pedagogy:

Astrid Giugni, English
Thavolia Glymph, History
Hannah Jacobs, Art, Art History and Visual Studies (research/teaching staff)
Laura Lieber, Religious Studies
Clare Woods, Classical Studies

Digital Representation

Mark Goodacre, Religious Studies
Andrew Janiak, Philosophy
Pedro Lasch, Art, Art History and Visual Studies
Mark Anthony Neal, African and African American Studies
Amanda Starling Gould, Educational Programs & Digital Humanities, FHI (research/teaching staff)

Digital Environments

Annette Joseph-Gabrielle, Romance Studies
Josh Sosin, Classical Studies
Phil Stern, History
Edward Triplett, Art, Art History and Visual Studies
Augustus Wendell, Art, Art History and Visual Studies

Appendix B: Listening Session Summaries

From February to July, 60-90 minute sessions were held with each faculty Working Group listed in Appendix A and the following Summaries were circulated for signoff to Working Group Members after the Listening Sessions as a means of accurately capturing and ensuring participant concurrence with the Major Takeaways associated with each group

Natural Sciences/OIT (IT Support Discussion) Major Takeaways

- There is ongoing concern surrounding the uncertainty of the breadth of resources that may or may not be provided by the University, as well as what the costs from the University/Med School will be and what funding will be required. Would be helpful for OIT to assist in the engagement with SoM Technical Services regarding elements like HARDAC, and act as an expert voice of reason and support to the research groups.
- Seams are beginning to show in the overall research IT service model. There is a growing lack of understanding between OIT and SoM IT providers as to what services are provided by who and how to acquire them.
- Clearer support for data management, particularly in regard to new guidelines resulting from the forthcoming Data Policy updates, and how that all intertwines.
- Completing longer and more intricate jobs on a shared cluster has proven to be a consistent obstacle, and additional professional support has been needed in assisting jobs to completion. Currently, groups are being forced to share individual resources which can't maintain the needs of the many, and additional personnel would be a great step in rectifying this. This would standardize the approach taken by many of our peer institutions.
- Current support personnel, though limited, is an expert resource with great work to their name. However, it's currently not efficient or ideal on individual group budgets.
- Many researchers are having to go outside Duke to peer institutions and national labs to use certain clusters, would be a boon to the University and researchers alike to have the means to run small to medium scale jobs with a home cluster.
- Research groups are currently being forced into positions where it is necessary to utilize resources that are not funded by their labs. There are currently issues surrounding memory, I/O, and guaranteed jobs on the cluster. The current model doesn't supply the necessary amount of storage with fast I/O paired with low ram jobs.
- Many colleagues have begun using GPUs in their research, and it would be great to have access to additional resources that open that door of opportunity.
- Lowering the barrier to entry on computational elements would encourage greater growth and well-roundedness in the community. A desire for additional offerings of workshops and/or online reference materials to alleviate that obstacle.
- Make clear the appropriate route to access and utilize desktop support.
- A clear central hub for additional resources and relevant trainings.

Social Sciences/OIT (IT Support Discussion) Major Takeaways

- There is an ongoing and immediate need for scientific support. Specifically, the resources to educate and guide students and faculty to make use of new and existing technologies such as machine learning and survey in their research.
    - The introduction of classes/courses for graduate students in machine learning and related fields.
    - With SSRI no longer a viable source, and CCT appearing unlikely to fill this particular gap in the short-term, how do we meet these demands?
    - High-level overviews of new technologies/techniques are not useful. There is a need for a deep understanding of how to apply these techniques in our space.
- Greater representation from social scientists at the decision-making table, particularly regarding policies that impact Social Sciences research. For example, the updated Data Policy guidelines appear to have direct implications on how Social Science research is performed and yet social scientists were not involved in the creation of the policy.
- There is a challenge in meeting specialized needs for some research efforts. While there are good IT solutions available for general needs, and it scales well to meet general demand, there is a challenge getting help with specialized needs. This could be compute, storage, usage, or scientific support. A resource center that understands the ins and outs of options for researchers is key to solving the needs actively present and to removing unnecessary obstacles. (Example: The need for Full Stack Developers)
- Assistance in building and maintaining large data resources, that in turn could be actively shared and utilized throughout the Social Science community (internal and external communities). Additionally, specialized assistance in navigating the necessary regulatory needs and demands.
- Duke policies and actions dictated by security or compliance measures can actively derail research timelines, projects, and theses. Overregulation is commonly stymying what should be considered standard procedure.
- Regarding the use of sensitive data, there appears to be differing approaches when it comes to Institutional level research versus the Social Sciences research.
- Data Collection, Statistical Analysis, and Machine Learning are the three cores that require the most additional support and resources. Perhaps there is need for additional access to different types of compute and storage depending on analysis needs and/or sensitivity of data?

Basic Sciences/OIT (IT Support Discussion) Major Takeaways

- There is a growing need for access to long-term storage for imagery/data that is curated/managed by researchers, who as of now hold the responsibility of maintaining that information for long periods of time. There are concerns over whether the University will pull the current level of support, as well as questions surrounding where responsibility for maintaining these assets should fall and how to pay for this storage when associated grant funding runs out.

- In tandem with long-term storage needs, there is growing concern that researchers may be unaware of the storage options currently available to them, and have resorted to more rudimentary means (e.g. external hard drives).

- There is a need for training and education opportunities for students on applying techniques/approaches to their research areas (e.g. ML). There is a need for deeper subject matter expertise (how does this apply to my work?)

- The NIH mandates that information be readily available for public sharing, and currently external services are being utilized for such. It would be a boon to have a Duke service that allows for easy public sharing. Current usable services are limited and disorganized. All publicly available tools demand some level of tweaking and specific curation.

- Local servers and computation clusters would provide great benefit.

- There is a need for more specialized/dedicated local support personnel who can quickly and efficiently address problems as they arise, without the need to bounce a ticket back and forth with others. Current ticketing process and search for expert help is arduous and inefficient.

- Some security requirements impede current initiatives, a review of current standards would be helpful to ensure it's not too stringent. Specifically, local/shared account needs vs. NetID – Service Accounts, new computer setup (DHTS mandated requirements vs. researcher customization needs), and a general need to find the balance between security and creativity.

- Endpoint management needs to be reviewed to align with research practices. There are some systems that cannot be updated and still need access to network resources for a long time. Efficient and effective solutions are needed to bridge this gap between need of access and security needs.

- A better understanding of cost structure and availability of resources for cloud computing would be helpful. Researchers have been not able to include expected costs of cloud compute in any grants, which can lead to underestimations.

Engineering/OIT (IT Support Discussion) Major Takeaways

- There is uncertainty surrounding when researchers should pay for external services (e.g., GPUs), and why?

- For GPUs, current internal solutions take too long to reconfigure or reallocate, when the time needed to use the resources is minimal. Sometimes need better access to "bare metal". Getting access to cloud resources is proving difficult and time consuming, and it has become easier to just pay for outside solutions.

- When preparing grants, there is a general challenge of properly understanding and identifying what services and equipment are readily available, and what their associated costs are.

- Regarding specialized high-performance computing, there is a need for stable, on demand access to clusters with good support for MPI and InfiniBand connected storage, which currently isn't happening internally. This is forcing researchers to move towards national labs, which is an obstacle that our peers at other institutions aren't having to face.

- While there is no prevailing preference for where cloud resources live, there has been a great deal of effort put into Azure and establishing infrastructure there. It would be welcome to have Duke better facilitate access to other cloud services.

- Provided back-up services have been immensely beneficial to the research groups (noted for Pratt and referencing CrashPlan).

- Staff support from OIT has been worthy of great praise. Available expert personnel have been able to appropriately and efficiently navigate issues that have arisen. Additional dedicated personnel resources would be welcome.

- Regarding storage, there is a growing need for a sustainable mechanism that will maintain data for a long period of time. Often, grants will often require reserve of relevant data for 5+ years.

Digital Environments/OIT (IT Support Discussion) Major Takeaways

- There are active roadblocks surrounding your ability to properly manage computers and successfully integrate third-party software that is often be considered critical to research endeavors.

- There is an ongoing struggle with producing a sustainable studio environment for students. There is a continual need for a designated space where students can collaborate and access tools/resources that will aid them in producing more accurate and higher quality work (personal machines/devices aren't making the cut). The MPS Lab is a great example (adjacent to classroom 6), but still needs to evolve (version 2.0).

- There are places where a virtual OS can play a positive role, however there remains a number of obstacles on that front as well. Splashtop has been a success, but it has added to new issues (scheduling, connection to internal applications...).

- VCM Environments/Containers have worked well for CPU, but not GPU (which is critical). Researchers/students need to be sure they're logging into THEIR VM at start up. Additional training/resources surrounding VCM resources would be a boon, while IT further fleshes out GPU resources for the long-term.

- Students/Researchers are in need of a platform that is more robust than Sites@Duke for instances where there is a desire to host more than a static website. Using third-party solutions has been particularly effective on this front (e.g. Reclaim Hosting), wherein systems have been streamlined for efficiency and are well maintained over time.

- Solutions to managing large datasets would continue to be welcome, Box works to a varying degree, the expectation has become to not work with files at the same time on the same share.

Digital Representation/OIT (IT Support Discussion) Major Takeaways

- In general, there seems to be a sense that Humanities faculty don't seek out additional IT Support for their research. It would be a boon to determine a path to be more proactive, rather than reactive. To date, most individuals will wait until a crisis point is hit before seeking support. Many researchers are unaware of what technological resources are currently available to them, leading to more laborious efforts and less efficient work.

- There are many worthwhile resources readily available, but due to a lack of connective tissue between the researchers and technological support, faculty aren't utilizing them. Enhancing that level of enticement is going to be needed to help maneuver faculty out of their individual bubbles

- There are a number of incredible third-party interfaces/applications/services etc, and perhaps through pilot programs we can utilize these tools to better collaborate with peers outside our own institution. It would be beneficial to have some sort of working group that focuses on implementing inter-institutional platforms

- Within the Humanities, normal communications channels are crowded, and a lot of the key communications get lost in the noise. Often IT related subjects are less prioritized and pushed to a lower attention level.

- Scholars@Duke works great for a lot of disciplines, but not too well for the Humanities. Duke has a unique opportunity to be a leader in pioneering something akin to that, that does lift up the Arts.

- Faculty would like to see more Humanities students working in studio jobs early in their career at the University. Recruiting and marketing should have a second look taken, as well as be improved upon. There should be a better understanding that these opportunities also offer extended access for those who regularly work in those spaces.

Digital Pedagogy/OIT (IT Support Discussion) Major Takeaways

- There is a continued need for specialized support personnel who are well versed in the Humanities and the specific pitfalls and obstacles that come with the territory.
- Humanities students are in need of a more robust web space where they can perform greater experimentation (e.g. ReclaimHosting).
- Humanities students are also in need of regular remote access to GPUs and advanced processing means.
- On the topic of project archive, there is great need particularly in the visualization space and archiving models, etc. (E.g. A class building a digital archive and then needing to retain that for longevity).
- There is continued need of virtualized GPU resources for remote/visual display and intensive applications, which differs from other research groups' need of virtualized GPU resources for compute power.
- Needs are different in the summer versus the school year. Retaining students digital work online, archiving needs.
- A problem on the horizon, with a new course, is what to do when students need a specific digital tool (Matlab) and the obstacles they may run into when using it on their own. Who will provide the troubleshooting and technological support? It's far too difficult to be educator and tech support at the same time.
- Other Universities maintain a number of groups/centers with successful models of Humanities specialized technological support that has proven to be a great boon to their researchers (e.g. Stanford, MSU, Emory)

Appendix C: ITAC Minutes from Individual Working Group Discussions

From March to August, 60+ minute sessions of ITAC were devoted to hearing from each faculty Working Group listed in Appendix A to more fully explore the Major Takeaways Summarized in Appendix B

Natural Sciences, March 24, 2022:
https://itac.duke.edu/meetings-minutes/2022/march-24-2022-minutes/

Social Sciences, May 5, 2022:
https://itac.duke.edu/meetings-minutes/2022/may-5-2022-minutes/

Basic Sciences, May 19, 2022:
https://itac.duke.edu/meetings-minutes/2022/may-19-2022-minutes/

Engineering, July 14, 2022:
https://itac.duke.edu/meetings-minutes/2022/july-14-2022-minutes/

Humanities/Arts (Digital Environments), July 28, 2022:
https://itac.duke.edu/meetings-minutes/2022/july-28-2022-minutes/

Humanities/Arts (Digital Representation & Digital Pedagogy), August 11, 2022:
https://itac.duke.edu/meetings-minutes/2022/may-19-2022-minutes/

Appendix D: Feedback Session Summary Themes from Working Groups

In June and August, special 90 minute sessions of ITAC faculty were organized with Working Group faculty to ensure an accurate mapping of individual Working Group feedback from Appendix C into Summary Themes spanning all Working Groups

Consistent Theme – Data Challenges

- Assistance in building and maintaining large data resources, that in turn could be actively shared and utilized throughout the Social Science community (internal and external communities). Additionally, specialized assistance in navigating the necessary regulatory needs and demands. (Social Sciences)
- Data Collection, Statistical Analysis, and Machine Learning are the three cores that require the most additional support and resources. Perhaps there is need for additional access to different types of compute and storage depending on analysis needs and/or sensitivity of data? (Social Sciences)
- Local servers and computation clusters would provide great benefit. (Basic Sciences)
- Completing longer and more intricate jobs on a shared cluster has proven to be a consistent obstacle, and additional professional support has been needed in assisting jobs to completion. Currently, groups are being forced to share individual resources which can't maintain the needs of the many, and additional personnel would be a great step in rectifying this. This would standardize the approach taken by many of our peer institutions. (Natural Sciences)
- Many researchers are having to go outside Duke to peer institutions and national labs to use certain clusters, would be a boon to the University and researchers alike to have the means to run small to medium scale jobs with a home cluster. (Natural Sciences)
- Regarding specialized high-performance computing, there is a need for stable, on demand access to clusters with good support for MPI and InfiniBand connected storage, which currently isn't happening internally. This is forcing researchers to move towards national labs, which is an obstacle that our peers at other institutions aren't having to face. (Engineering)
- Solutions to managing large datasets would continue to be welcome, Box works to a varying degree, the expectation has become to not work with files at the same time on the same share. (Humanities/Arts)

Consistent Theme – Technology Challenges

- There is a growing need for access to long-term storage for imagery/data that is curated/managed by researchers, who as of now hold the responsibility of maintaining that information for long periods of time. There are concerns over whether the University will pull the current level of support, as well as questions surrounding where responsibility for maintaining these assets should fall and how to pay for this storage when associated grant funding runs out. (Basic Sciences) – STORAGE
- In tandem with long-term storage needs, there is growing concern that researchers may be unaware of the storage options currently available to them, and have resorted to more rudimentary means (e.g. external hard drives). (Basic Sciences) – STORAGE
- Regarding storage, there is a growing need for a sustainable mechanism that will maintain data for a long period of time. Often, grants will often require reserve of relevant data for 5+ years. (Engineering) – STORAGE
- On the topic of project archive, there is great need particularly in the visualization space and archiving models, etc. (E.g. A class building a digital archive that needs to be retained long-term, or individual student projects that need archiving). (Humanities/Arts) – STORAGE
- NIH funded data storage needs to be recognized by both DHS and DU as a critical component of infrastructure. (Basic Sciences) – STORAGE
- There is also a need for dynamic solutions for: A.) navigating data from "hot" to "cold" storage as necessitated by project stage and B.) sharing data both within Duke and across institutions to manage data availability requirements from the NIH. (Basic Sciences) – STORAGE
- A better understanding of cost structure and availability of resources for cloud computing would be helpful. Researchers have been not able to include expected costs of cloud compute in any grants, which can lead to underestimations. (Basic Sciences) – COMPUTE
- For GPUs, current internal solutions take too long to reconfigure or reallocate, when the time needed to use the resources is minimal. Sometimes need better access to "bare metal". Getting access to cloud resources is proving difficult and time consuming, and it has become easier to just pay for outside solutions. (Engineering) GPUs can be more complicated to use and require both access and additional training (e.g., researchers/students need to be sure they're logging into THEIR VM at start up) (Humanities/Arts) – COMPUTE
- While there is no prevailing preference for where cloud resources live, there has been a great deal of effort put into Azure and establishing infrastructure there. It would be welcome to have Duke better facilitate access to other cloud services. (Engineering) – COMPUTE

Consistent Theme –Personnel and Support Challenges

- There is a need for more specialized/dedicated local support personnel who can quickly and efficiently address problems as they arise, without the need to bounce a ticket back and forth with others. Current ticketing process and search for expert help is arduous and inefficient. (Basic Sciences)
- Current support personnel, though limited, is an expert resource with great work to their name. However, it's currently not efficient or ideal on individual group budgets. (Natural Sciences)
- Staff support from OIT has been worthy of great praise. Available expert personnel have been able to appropriately and efficiently navigate issues that have arisen. Additional dedicated personnel resources would be welcome. (Engineering)
- There is a continued need for specialized support personnel who are well versed in the Humanities and the specific pitfalls and obstacles that come with the territory. Tech support for students (who often work with faculty to advance their research) adds complexity to this issue. (Humanities/Arts)
- An ongoing and immediate need for scientific support. Specifically, the resources to educate and guide students and faculty to make use of new and existing technologies such as machine learning and survey in their research. (Social Sciences)
- There is a challenge in meeting specialized needs for some research efforts. While there are good IT solutions available for general needs, and it scales well to meet general demand, there is a challenge getting help with specialized needs. This could be compute, storage, usage, or scientific support. A resource center that understands the ins and outs of options for researchers is key to solving the needs actively present and to removing unnecessary obstacles. (Social Sciences)
- A variety of basic support issues often hinder research endeavors in the humanities and arts, where extramural funding may be less available, including basic computer as well as virtual OS management, integration of 3rd party software, studio environments for students, software purchases, and more robust website space/support. (Humanities/Arts)
- While OIT's DCC is effective in what it supports, other Universities have groups/centers with more successful models for Humanities specialized technological support for their researchers (e.g. Stanford, MSU, Emory) (Humanities/Arts)
- Despite many worthwhile resources readily available, a lack of connective tissue between the researchers and technological support, leads to underuse or lack of awareness. (Humanities/Arts, Basic Sciences)
- There is a need for more specialized/dedicated local support personnel who can quickly and efficiently address problems as they arise, without the need to bounce a ticket back and forth with others. Current ticketing process and search for expert help is arduous and inefficient. (Basic Sciences)
- Current support personnel, though limited, is an expert resource with great work to their name. However, it's currently not efficient or ideal on individual group budgets. (Natural Sciences)
- Staff support from OIT has been worthy of great praise. Available expert personnel have been able to appropriately and efficiently navigate issues that have arisen. Additional dedicated personnel resources would be welcome. (Engineering)
- There is a continued need for specialized support personnel who are well versed in the Humanities and the specific pitfalls and obstacles that come with the territory. Tech support for students (who often work with faculty to advance their research) adds complexity to this issue. (Humanities/Arts)
- An ongoing and immediate need for scientific support. Specifically, the resources to educate and guide students and faculty to make use of new and existing technologies such as machine learning and survey in their research. (Social Sciences)

- There is a challenge in meeting specialized needs for some research efforts. While there are good IT solutions available for general needs, and it scales well to meet general demand, there is a challenge getting help with specialized needs. This could be compute, storage, usage, or scientific support. A resource center that understands the ins and outs of options for researchers is key to solving the needs actively present and to removing unnecessary obstacles. (Social Sciences)
- A variety of basic support issues often hinder research endeavors in the humanities and arts, where extramural funding may be less available, including basic computer as well as virtual OS management, integration of 3rd party software, studio environments for students, software purchases, and more robust website space/support. (Humanities/Arts)
- While OIT's DCC is effective in what it supports, other Universities have groups/centers with more successful models for Humanities specialized technological support for their researchers (e.g. Stanford, MSU, Emory) (Humanities/Arts)
- Despite many worthwhile resources readily available, a lack of connective tissue between the researchers and technological support, leads to underuse or lack of awareness. (Humanities/Arts, Basic Sciences)

Consistent Theme –Organization Challenges

- Greater representation from social scientists at the decision-making table, particularly regarding policies that impact Social Sciences research. For example, the updated Data Policy guidelines. (Social Sciences)
- Duke policies and actions dictated by security or compliance measures can actively derail research timelines, projects, and theses. Overregulation is commonly stymying what should be considered standard procedure. (Social Sciences, Natural Sciences)
- Regarding the use of sensitive data, there appears to be differing approaches when it comes to Institutional level research versus the Social Sciences research. (Social Sciences)
- Seams are beginning to show in the overall research IT service model. There is a growing lack of understanding between OIT and SoM IT providers as to what services are provided by who and how to acquire them. (Natural Sciences)
- Lowering the barrier to entry on computational elements would encourage greater growth and well-roundedness in the community. A desire for additional offerings of workshops and/or online reference materials to alleviate that obstacle. (Natural Sciences)
- Some security requirements impede current initiatives, a review of current standards would be helpful to ensure it's not too stringent. Specifically, local/shared account needs vs. NetID – Service Accounts, new computer setup (DHTS mandated requirements vs. researcher customization needs), and a general need to find the balance between security and creativity. (Basic Sciences)
- When preparing grants, there is a general challenge of properly understanding and identifying what services and equipment are readily available, and what their associated costs are. (Engineering)
- Better mechanisms are needed for communicating IT users' needs/requirements to the policy makers and financial managers. ITAC is an excellent mechanism for this on the University side, but there is need for a comparable avenue in SoM. (Basic Sciences)
- There are inefficiencies and a lack of support for research oversight which is leading to weakened compliance. (Basic Sciences)
- There is ongoing concern surrounding the uncertainty of the breadth of resources that may or may not be provided by the University, as well as what the costs from the University/Med School will be and what funding will be required. Would be helpful for OIT to assist in the engagement with SoM Technical Services regarding elements like HARDAC, and act as an expert voice of reason and support to the research groups. (Natural Sciences)

Appendix E: Feedback Session to Prioritize Findings

In August, a series of draft "Action Points" (later becoming the report's Findings / Recommendations) were developed by consolidating the common Themes in Appendix D and based on both Listening Session Takeaways (Appendix B) and ITAC Discussion points (Appendix C). The August special 90 minute sessions of ITAC invited all available Working Group faculty review and "vote" (via sticky dots) on these "Action Points" (Findings / Recommendations); faculty participants also augmented the draft Action Points and added (and voted for) write in points where they perceived missing priorities.

Top priorities (based on number of sticky dot votes)

Technology:

- Create new service offerings and guidance for specialized compute needs (e.g. HPC/GPU clusters that are easier to use and sized to meet demand)
- Introduce guidance (e.g., concierge service) for which cloud and on-premises options can support particular research and academic needs, including what resources are available to faculty and students across what cloud services or locally
- Implement new long-term storage options for a variety of use-cases, considering both archival as well as sharing/external access needs; ensure solutions recognize the differences in storage needs across the research lifecycle

Process/Structures

- Objectively assess costs/benefits of dual research infrastructures (OIT/DHTS)

Personnel

- Establish domain-specific technical support to assist with specialized needs, including compute, storage, data analysis, and new applications of technologies (e.g., application of machine learning, creating digital environments, building and maintaining large, shared data resources)
    - Leverage these resources to offer education programs faculty / students

Write in:

- Security requirements hampering ability to do work/research

Moderate priorities (based on number of sticky dot votes)

Process/Structures:

- Provide clear processes/service/steps to support researcher navigation of security, compliance or DUA (Data Use Agreement) requirements, so faculty don't have to know how to navigate separate offices (OIT/ITSO/DORI)
    - Include consultative support for grant preparation related to these requirements
- Establish clear guidance/direction for existing services for data collection, storage, and compute services that are available to researchers at no or low cost; clearly classify their appropriateness for use based on the sensitivity of the data or research being performed (see related Technology bullet 3)
- Create domain-specific training for faculty, grad students and post docs on applying new technology options to research areas; create support structures for undergraduates in support of Humanities and Arts researchers

Write ins:

- Shared access to expensive data resources) doesn't seem (to me) to be reflected
- Differential approaches for faculty, grad students, undergrads (e.g., honors theses)

- Increased needs for digital publishing services, especially extra-mural options w/ visualization + archive options (Reclaim Hosting)
- Centralized (biomedical) assistance for machine learning and deep learning use applications in image analysis

Domain-specific or lower priorities (based on number of sticky dot votes)

Technology:

- Improve local options for shared compute needs
- Establish specialized labs with workstations, licensed software and other tools for specific needs (e.g. digital work)

Personnel:

- Improve faculty representation on institutional/administrative decision making or review groups (DORI)
- Re-evaluate basic technology support and research support levels and allocated resources in the humanities and arts and benchmark against programs at other leading universities.
- Create support structures for team teaching in the humanities as a means of establishing initial 'connective tissue' among faculty and with IT
- Evaluate expansion of +Programs to embrace to digital humanities projects

Write ins:

- Need to benchmark against other universities outside humanities as well-e.g., all research support
- Communication of policy changes, with time to respond, before they happen
- Many issues are human – tech will not address all situations

Appendix F: Detailed Findings Based on Highest Priority from Working Groups

Based on the highest priorities Findings / Recommendations in Appendix E, the Takeaways from Listening Session (Appendix B) together with details from ITAC Discussion Points (Appendix C) were comprehensively mapped to each from across all Working Groups. These details are recognized to be an essential part of the shortened Report, but are shared via this appendix so as not to overwhelm or exhaust the reader.

People Findings and Recommendations

Finding A: The focus of OIT's "Research Computing" function as the prevailing embodiment of IT support for research led to a robust and responsive infrastructure comprising networking, computation, and storage supporting general needs. However, it did not build a base level of personnel support for research endeavors that rely on data collection, analysis and storage or the application of technology to specific scientific domains. By leaving the human component to the lab, department, and school level, the result has been inconsistency in specialized research support as conveyed by all faculty working groups, even those with substantial extramural funding.

The lack of sufficient domain-specific technical personnel pervades all disciplines:

- Despite many worthwhile resources readily available, a lack of connective tissue between the researchers and technological support, leads to underuse or lack of awareness. (Humanities/Arts and Basic Sciences Faculty Working Groups)
- There is a challenge in meeting specialized needs for some research efforts. While there are good IT solutions available for general needs, and it scales well to meet general demand, there is a challenge getting help with specialized needs. This could be compute, storage, usage, or scientific support. A resource center that understands the ins / outs of options for researchers is key to solving the needs actively present and to removing unnecessary obstacles. (Social Sciences Faculty Working Group)
- "The biggest need is for human infrastructure." (Hannah Jacobs in ITAC discussion with Humanities/Arts)
- There is a need for more specialized/dedicated local support personnel who can quickly and efficiently address problems as they arise, without the need to bounce a ticket back and forth with others. Current ticketing process and search for expert help is arduous and inefficient. (Basic Sciences Faculty Working Group)
- People are important and some degree of domain knowledge is also really important—what physical chemists need, or what particle physicists need, or what genomicists need—they're not really the same thing. Compared to some of our peers we are under-resourced in that kind of domain computational expertise. The funding models of our peers are all over the map, from fully allocated to heavily charged out, but just in pure body count we have fewer of those people than many of our peers do right now. (ITAC discussion with Natural Sciences)
- We really need help from high level IT specialists in building these computer environments and testing and troubleshooting them. Ideally, we would have a sandbox where we can kind of play around and as we add new software or new hardware we'd be able to make sure that the environment and analysis has appropriate precision of timing [between instrumentation, storage and computation]. (ITAC discussion with Basic Sciences)
- Completing longer and more intricate jobs on a shared cluster has proven to be a consistent obstacle, and additional professional support has been needed in assisting jobs to completion. Currently, groups are being forced to share individual resources which can't

maintain the needs of the many, and additional personnel would be a great step in rectifying this. This would standardize the approach taken by many of our peer institutions. (Natural Sciences Faculty Working Group)

- Staff support from OIT has been worthy of great praise. Available expert personnel have been able to appropriately and efficiently navigate issues that have arisen. But additional dedicated personnel resources would be welcome. (Engineering Faculty Working Group)
- Current support personnel, though limited, is an expert resource with great work to their name. However, it's currently not efficient or ideal on individual group budgets. (Natural Sciences Faculty Working Group)
- While OIT's Duke Compute Cluster (DCC) is effective in what it supports, other Universities have groups/centers with more successful models for Humanities specialized technological support for their researchers (e.g. Stanford, MSU, Emory) (Humanities/Arts Faculty Working Group)
- Sometimes it is hard to get to [support personnel] through filing tickets. We need to have people on the ground close to us so that when something goes bad they are able to solve problems right away. Such problems (at times due to security software kicking us out of the network) may occur in the middle of an experiment, and we need to have assistance rather quickly. A little bit of handholding would make progress happen much faster. Everyone would benefit from consultation and education from people with more IT experiences than we have. (ITAC discussion with Basic Sciences)
- A variety of basic support issues often hinder research endeavors in the humanities and arts, where extramural funding may be less available, including basic computer as well as virtual OS management, integration of 3rd party software, studio environments for students, software purchases, and more robust website space/support. (Humanities/Arts Faculty Working Group)
- The biggest issue for the Humanities is when there is no one or only one person consulting with faculty in domains like romance study-specific technology, and where that same person is also helping with subject-specific classroom and co-curricular projects. (Humanities/Arts Faculty Working Group)
- There is a need for staff to help figure things out or help when things go wrong such as help from people who could be more proactive with vendors to get cloud environments working is needed. And help with hardware especially around computation and networking, as well as with storage issues, and other things are what faculty would say are the biggest needs. Research needs and technology changes are moving so quickly, and the challenge is keeping up with these. (Amanda Randles and Henry Pfister in ITAC discussion with Engineering)
- There is a continued need for specialized support personnel who are well versed in the Humanities and the specific pitfalls and obstacles that come with the territory. Tech support for students (who often work with faculty to advance their research) adds complexity to this issue. (Humanities/Arts Faculty Working Group)
- Machine learning needs of researchers came up in the working group meeting, while recognizing Duke has a very active machine learning community here on campus. These needs are expected to become a bigger and bigger part of what researchers do. (ITAC discussion with Natural Sciences)
- There are needs for personnel to consult on resources for classes. One senior design course for machine learning and signal processing involved projects for this class where students decided on a repository and applied machine learning. Matching all versions can be a nightmare on an OIT VM so the easiest thing is a Google GCC machine where root admin access is provided, and everything can be installed with required versions. But this can be expensive. Also, training with multiple GPUs has been problematic. In one project case a student used Jukebox, a neural network that generates pop music when provided with a

parameter such as: "I want a song by Drake with these lyrics." This was expensive to do on Google because it runs for a long time. So, they got it running on the DCC with the help of an OIT technical expert; this was a success story. Another group started with the Salesforce Research AI Economist package. This involved reinforcement learning that starts with simple agents that pretend to be an economy and you see how the economy works. This utilizes a reinforcement library which is on top of a parallelization library called Ray. Ray is very finicky, and the project has never gotten it to work well on anything other than bare metal. (ITAC discussion with Engineering)

- [Regarding the process of digital mapping, and the essential role of undergraduates supporting the research…] … it's hand labor in the tracing, though I had a student in my AI for Arts and Humanities class who was testing an automated training and neural network to identify footprints in Sanborn maps. I think that's one of the beauties of our courses being intersection with students coming from computational backgrounds. (ITAC discussion with Humanities/Arts)
- In the School of Medicine, many of the researchers say the School lacks computational expertise to assist, but we also lack Linux support and Linux experience, such as how to work with pipelines with Python and R. (ITAC discussion with Basic Sciences)
- Open Science Grid support is incredibly useful and facilitates faculty research in terms of making sure that we are matched up with what our needs require, especially when faculty do not know a priori what they need. The sort of technical personnel who can help bridge this gap are essential: computational power is a necessary but not sufficient condition to meet faculty need and the technical personnel deliver on that sufficiency condition. (ITAC discussion with Natural Sciences)

Recommendation 1: In consultation with faculty, Duke should devise a service that builds and supports teams of domain-specific technical personnel who are well versed in both on-site and cloud options. The personnel should be equipped to assist with specialized needs in the different domains, including compute, storage, data collection and analysis, and new applications of technologies (e.g., applying machine learning, creating digital environments, building and maintaining large, shared data resources). Evaluation of support structures should extend beyond department vs. central IT resources, also encompassing prospective engagement from areas like Libraries or other programs. The effort should start with pilots and expand with experience to take advantage of early lessons, include faculty feedback, and have central assistance in recruiting personnel.

Beyond domain-specific technical assistance, educational and training programs for faculty / students are needed:

- There is an ongoing and immediate need for scientific support. Specifically, the resources to educate and train students and faculty to make use of new and existing technologies such as machine learning and survey in their research. (Social Sciences Faculty Working Group)
- Whether student environments for teaching are provided via DCC or cloud vendors, it is important to recognize that that some students have programming experience and some do not, and so appropriate support (training) is required. (ITAC discussion with Engineering)
- Lowering the barrier to entry on computational elements would encourage greater growth and well-roundedness in the community. A desire for additional offerings of workshops and/or online reference materials to alleviate that obstacle. (Natural Sciences Faculty Working Group)
- The ability for faculty to bring this technology into the classroom is essential and is how we want to educate students. Using the DCC to educate students would be easier since it is painful to spin up a Google cloud instance and it would be nice for students to be able to

learn using the cluster without the Google-setup hurdle. Nonetheless, there is a sense that teaching real-world "how to set up and run" in the commercial environments is also important. Having both options (onsite with DCC and cloud-based) is desirable as students can sometimes get discouraged before they even start programming. (Jerry Lynch, Henry Pfister, Volker Blum and Amanda Randles in ITAC discussion with Engineering)

- An important thread is that Social Sciences (and others) have issues around involving novice graduate and undergraduate student researchers in projects. This is may not be an OIT issue per se, but it's an important layer of our digital ecosystem that can get lost. (Paul Jaskot in ITAC discussion with Social Sciences)
- GPUs can be more complicated to use and require both access and additional training (e.g., researchers/students need to be sure they're logging into *their* VM at start up) (Humanities/Arts Faculty Working Group)

Recommendation 2: Develop and catalog training resources to support an ongoing education program targeting faculty and graduate students but also serving undergraduates who participate in research endeavors. Such a program should include applying established approaches (e.g. data analysis using common applications such as R and emerging machine learning) and should leverage the domain-specific technology support resources, as well as existing co-curricular programs such as via Libraries (e.g., CVDS), Center for Computational Thinking (CCT), OIT (e.g., Co-Lab Roots), various curricular programs offered by academic departments, and others.

Process and Organizational Findings and Recommendations

Finding B: Separate IT infrastructures supporting research have operated for more than two decades (OIT-Campus Schools, DHTS-Schools of Medicine and Nursing), along with distinct lab and departmental IT personnel. Despite each organization striving to deliver great service, and while OIT and DHTS coordinate on many items, the impact for faculty collaborating across these complex infrastructures for research has become increasingly problematic. For example, decisions of SoM research policy and support are well understood to be the purview of the School, however their impact can reverberate across campus in terms of research administration and policies as well as directly impact faculty in campus schools such as Pratt and Trinity.

Structurally, the separate infrastructures introduce confusion, frustration, and inefficient solutions:

- The current model has pretty much prevailed over many years, which is that DHTS is responsible for supporting all School of Medicine associated researchers as well as the hospital system (which has very different needs and demands) and OIT is responsible for supporting all non-school of medicine, and I know that's not exactly true, because I know OIT personnel have spent a lot of time with School of Medicine so I know that that there isn't a formal support 'firewall' between the two organizations, but there does seem to be somewhat of a barrier that could be lowered and if that were the goal it should be a primary goal. (ITAC discussion with Basic Sciences)
- Regarding the perceived contrast in support for campus vs. SoM, "I'm really glad that there actually is a group of people who think about this stuff on the campus side. I don't take that for granted because of living a little bit in both worlds." (Greg Wray in ITAC discussion with Natural Sciences)
- As for existing tools to support researchers like myresearchhome, and myresearchpath, they can become overwhelming and some of the information is up to date, and some it's not up

to date. While it's very good to have centralized resources and information such as 'this is the cluster' and 'this is how you access it' and 'this is how fast you can get to it', we still need a balance with local consultation and support. (ITAC discussion with Basic Sciences)

- Seams are beginning to show in the overall research IT service model. There is a growing lack of understanding between OIT and SoM IT providers as to what services are provided by whom and how to acquire them. (Natural Sciences Faculty Working Group)
- Better mechanisms are needed for communicating IT users' needs/requirements to the policy makers and financial managers. ITAC is an excellent mechanism for this on the University side, but there is need for a comparable avenue in SoM. (Basic Sciences Faculty Working Group)
- Sharing resources between the medical side and campus side of Duke is tricky because of the incredible diversity on the university side. This is especially challenging when data sharing involves undergraduates and graduate students who have never worked on issues of data compliance and data transfer. Finally, cost-sharing while often needed, does not address the needs of graduate students and faculty who do not have funding and undergraduates who are being pushed toward research involvement. (Social Sciences Faculty Working Group) [Also applies to recommendation 4.]
- Large data sets need to be moved quickly between the imaging system and associated reconstruction computation, which today requires some sort of heavy-duty local compute and keeping the storage very close to the imaging scanners. This is because you need to offload the data from the scanner to a backup, which means network speed becomes a concern, and which often drives a need for local (departmental) compute solutions. Guidance and assistance navigating other options is needed, especially as SoM is increasingly emphasizing (and financially incentivizing) cloud solutions. (ITAC Discussion with Basic Sciences) [Also applies to recommendation 3.]
- Things have often seemed to be a moving target in DHTS like take object storage which we've talked about–it's great, it's fantastic, it's cheap but if we can't get our data off when it moved from on prem object storage, we were all caught. When researchers are suddenly saddled with needing to move it's frustrating and there doesn't seem to be a long term plan other than throwing the next point solution out. (ITAC Discussion with Basic Sciences) [Also applies to recommendation 9.]
- There is concern regarding costs of the SoM direction of moving computation completely or largely into the cloud, as well as with the seemingly abrupt timeline researchers are expected to make that shift, with limited or no material support for making the conversion. Moreover, for groups with data hosted on an OIT VM there often is no ingress or egress path into their secure storage blobs in the Azure (cloud) system that researchers can use, and the solutions offered were less than ideal. This may drive basic sciences researchers to adopt the DCC environment. (ITAC discussion with Natural Sciences) [Also applies to recommendation 7.]
- We often struggle to maintain a compute infrastructure that exists, or almost exists, today in our local labs, but every now and then needs to be reconfigured in the proper way. Lab personnel can use it to meet the CPU and GPU operational needs. We have moved more towards using deep learning and we have our own resources, but they are a challenge to maintain and operate. Is this something that the university is addressing at all? We need not only the proper infrastructure but the systems that can support the image acquisition and instrumentation requirements that we have. (Alexandra Badea and Lindsey Glickfeld in ITAC Discussion with Basic Sciences) [Also applies to recommendation 6.]
- We need a better connection of local support personnel (with domain knowledge) to the larger IT support system, be it DHTS or OIT. Today we have a system where we file online tickets and although everybody has the best intention, sometimes it's hard to find the

person who's qualified and has the security clearance to actually implement the solution and oftentimes it comes back to us. It's not super-efficient when you have to interact with 6 people, and they are doing their best with everybody trying to locate the right person. I would love a simple system where we know directly who knows enough to solve the problem and maybe could have the security clearance to fix my problem. (ITAC Discussion with Basic Sciences)

- There is concern of shortcomings in research IT support for campus researchers who collaborate with SoM or utilize resources housed there. The perception is that there is not a comparable counterpart to OIT's level of support within the School of Medicine. Related, there is a perceived concern over a lack of strategic vision for computing with the School, and a lack of clarity about who owns what parts of the environment and support, or who's the decision authority. (ITAC discussion with Natural Sciences)
- Data exchange across groups can be difficult, including with collaborating faculty on the campus side, in the statistics department. Together we study voxel based morphometry to understand how brain atrophy happens in mouse models of Alzheimer's disease. We have developed our own brain atlases with about 300 regions per brain, and we can count the number of tracks or connections between each pair of these regions. We construct connectivity matrices, which we analyze together with our collaborators. For example, we have worked with Dr. David Dunson using tensor network principal component analysis to relate brain connectivity to traits, and we have developed methods for characterizing track bundle properties. (ITAC discussion with Basic Sciences)
- One of the hard things about being in the basic sciences is that my trainees are undergraduates. They're in BME, they're in computer sciences, they're graduate students they span the whole university, and that's where a lot of the challenges come up. (ITAC discussion with Basic Sciences)
- In BME we run studies in the hospital and the need to balance between security in DHTS & OIT creates difficulty for fluidity of data flow—we have to set up VMs to log in and move data because you can't get two networks to talk to each other and you're creating virtual hot points for multiple users getting permissions in different groups. How much bandwidth and effort gets consumed just because we're copying files internally to mitigate between different security requirements? (Mark Palmeri in ITAC discussion with Natural Sciences) [Also substantiates recommendation 4.]
- It is important that we have the right balance between local compute versus cloud compute and the concerns that we hear from people is that the value systems for security seems so weighted to securing devices that it creates difficulty when trying to secure local computer systems by specialized IT Support on the ground. We need a sandbox environment where we can install, uninstall, test things, which ideally would enable us to install less standard software, test and run as well as the ability to develop and share. We also need improved data storage/backup and to be able to effectively share data with outside collaborators. These are all very difficult to do within the School of Medicine security framework and it limits the applications we can use and thereby inhibits the research. And then how do we deal with these issues from the framework to ensure a smooth SoM and campus interaction? (ITAC discussion with Basic Sciences) [Also applies to recommendation 4.]
- Today there seems to be kind of a firewall between the SoM and campus. This restricts the sharing of data with the rest of campus and the use of resources from OIT. Also, it would be nice to have the data be more easily shareable across universities, since many of us have collaborators at other institutions. It's also becoming an NIH requirement that raw data be accessible to the public and so, a seamless, hierarchical system of data storage that is also shareable, would really be advantageous. (ITAC discussion with Basic Sciences)

Further, that such difficulty exists despite the best intentions of many individuals suggests an endemic, structural issue rather than a transient service, personnel, or organizational shortcoming:

- DHTS personnel are really trying hard to understand the needs of the basic science researchers and working very hard to understand requirements and the options available given funding realities. So, there's a lot of things that that we're trying to work through, and I know Terri West has been helping us with all of the business requirements, and so I think there are some anomalies of things that that are harder for DHTS to support and provide services to. I think what they're trying to do with the OASIS team is to get that more customized support available to the researchers. But it's been very insightful to hear some of the issues that we have and have been raised here; I think we know some of them, but we don't know all of them. (DHTS feedback in ITAC discussion with Basic Sciences)
- Another pain point is people who have to move data between special installations, such as going from a DHTS environment to OIT. This affects even people who only work on the campus side because many of the core facilities are in the DHTS network and deliver their data to a file share in the DHTS network. And then the person in campus side has to move it over into wherever they're going to store it stage it work with it. And so that file transfer process can be a bit fraught because you need to work across that firewall. (ITAC discussion with Natural Sciences) [Also applies to recommendation 4.]
- There has been an interesting discussion among SoM basic scientists research needs and the way that DHTS is configured to support those needs and the way OIT it is configured to support those needs. It appears that the structure of that support and the way that scientists access the support is configured quite different in the two organizations. So the real question is: What are the administrative ways to make it simpler and more likely that research computing users in the basic sciences have access to and know about the support that OIT can provide so that we don't have to duplicate efforts? OIT has some great research computing expertise. DHTS has all the responsibility for PHI [Protected Health Information] data and all of the protocols that go with that. Isn't it more reasonable to sort of divide the expertise, and then make sure that the users know where to go, and that OASIS is plugged into OIT's team in ways that lower the barrier to get from the SoM side to the resources that OIT provides? So that's why I say this is a question I think for the highest level of administrators. The support needs have changed, of course, as science has changed. Should we be considering how to address those needs through a different model? To be clear, I'm impressed with all of the efforts that are being made on behalf of the SoM and in particular Terri West's group is doing fantastic support for us through OASIS and Aby Conaway's group. I think we have a lot of great people I just think that maybe the way they're organized to support specifically basic science is not optimal. And I'm just asking is there a way to rethink that structure. (Terry Oas in ITAC discussion with Basic Sciences)

Recommendation 3: Conduct an external review with peer expert leaders (from academic medical centers that are tightly coupled with their universities) to objectively assess costs/benefits of dual and decentralized research infrastructures (OIT/DHTS/Schools/Institutes); present for review to University leadership including the Provost, SoM Dean and other University and DUHS leaders. Regardless of the outcome of the review, seek to discontinue or drastically limit the practice of SoM policies and technologies being imposed on the campus without input from faculty.

Finding C: Disparate and sometimes conflicting compliance and security measures (Campus and SoM) and a seeming lack of gradations of risk for various data types and uses can inhibit intra- and inter-institutional research collaboration. While researchers acknowledge and accept the necessity to comply with requirements to protect sensitive and restricted data, they desire approaches that better define options and balance the risk of actual data loss with the perceived adverse impact to Duke's research endeavor itself.

The complexities associated with the institutional policies and approval processes for data use, coupled with differing approaches for campus and SoM, create difficulty for not only for research within SoM but also for researchers and collaborators on the rest of campus.

- It is important that we have the right balance between local compute versus cloud compute and the concerns that we hear from people is that the value systems for security seems so weighted to securing devices that it creates difficulty when trying to secure local computer systems by specialized IT Support on the ground. We need a sandbox environment where we can install, uninstall, test things, which ideally would enable us to install less standard software, test and run as well as the ability to develop and share. We also need improved data storage/backup and to be able to effectively share data with outside collaborators. These are all very difficult to do within the School of Medicine security framework and it limits the applications we can use and thereby inhibits the research. And then how do we deal with these issues from the framework to ensure a smooth SoM and campus interaction? (ITAC discussion with Basic Sciences) [Also applies to recommendation 3.]
- There is an opportunity to develop shared data resources maintained centrally to reduce duplication of effort, improve oversight and compliance, and increase access to data among Duke investigators. Additionally, specialized assistance is needed in navigating the necessary regulatory needs and demands. (ITAC Discussion with Social Sciences) [Also applies to recommendation 10.]
- In BME we run studies in the hospital and the need to balance between security in DHTS & OIT creates difficulty for fluidity of data flow—we have to set up VMs to log in and move data because you can't get two networks to talk to each other and you're creating virtual hot points for multiple users getting permissions in different groups. How much bandwidth and effort gets consumed just because we're copying files internally to mitigate between different security requirements? (Mark Palmeri in ITAC discussion with Natural Sciences) [Also applies to recommendation 3.]
- In terms of how security can limit our research, for collaborations across campus we've used OneDrive or Box, and it seems that this is not easy because of permission problems, and it becomes even more difficult if we try to do it with collaborators outside of Duke. If it's secure data, it needs to be packaged up and restricted highly. Then you can think about encrypting it, and then sharing it in some fashion, or you can think about using Box. The limitation with OneDrive is that it isn't allowed to be shared off campus. It would be nice if it were because OneDrive has some great features for people collaboratively writing together. For example, we have collaborators at Johns Hopkins and they're using OneDrive to share with us, but we can't share back with them due to security limitations here. (Lindsey Glickfeld, Charley Kneifel and Alexandra Badea in ITAC discussion with Basic Sciences)
- Regarding use of sensitive data, there appear to be differing approaches for Institutional level research versus Social Sciences research. (Social Sciences Faculty Working Group)
- Another pain point is people who have to move data between special installations, such as going from a DHTS environment to OIT. This affects even people who only work on the campus side because many of the core facilities are in the DHTS network and deliver their

> data to a file share in the DHTS network. And then the person in campus side has to move it over into wherever they're going to store it stage it work with it. And so that file transfer process can be a bit fraught because you need to work across that firewall. (ITAC discussion with Natural Sciences) [Also applies to recommendation 3.]

"One size fits all" security approaches and complexity of different components of the review / approval process further exacerbate efficient research conduct, especially when that "one size" is tailored to the most extreme risks. Related was the need for a more transparent process for IRB and other research data related approvals:

- Duke policy or actions dictated by security / compliance controls can actively derail research timelines, projects, and theses. Overregulation commonly stymies what should be viewed as standard procedures. (Social Sciences and Natural Sciences Faculty Working Groups)
- There is a lack of coordination among the many oversight departments including ITSO, RFDS, ORS, and IRB. It is unclear from the faculty perspective when to involve whom. IRB has some decision rules, but the process needs to be improved for faculty as the process is confusing and time-consuming and there are cases of researchers giving up. […] This is often because oversight approval gets hung up. While the mission of institutional protection is understood, it would be good to know what decision rules are being used, whether these rules can be appealed, and why these rules differ from the rules of other boards. It would be good to think these processes through from the researcher's perspective. Faculty input and decreasing delays are needed. (Sunshine Hillygus in ITAC discussion with Social Sciences)
- Regarding data management, policies and security, it feels like in some cases, we have a very binary decision on either all research is sensitive or it's all public and so there's not enough gray area recognized in between on how to deal with that. Moreover, it would be good if Duke could lower the barrier to entry for securely managing data and provide better training on use of the available resources across the disciplines. (ITAC discussion with Natural Sciences)
- Better communication and coordination are needed across the many oversight bodies involved in the Research Oversight processes and policy, as the process for the researcher has become complex, time-consuming, and a hindrance to research. (ITAC discussion with Social Sciences)
- Nirvana for the social sciences researchers at Duke would be a place where investigators could go when needing to access a data set. There would be coordination across the oversight organizations and bodies where all could go for help on IRB (Institutional Review Board) requirement fulfillment, and for needed training on maintaining confidentiality. There would be clear, organized, and maintained systems about constraints for accessing data. A safer and more efficient environment would be provided for researchers including undergraduate and graduate student researchers. If attractive data sets are provided and oversight requirements are clear and efficient, researchers will access them. (Kate Bundorf, in ITAC discussion with Social Sciences) [Also applies to recommendation 10.]
- Because our systems are so specialized, one thing that has been really disruptive to us over the past few years is the increasing oversight on the updates and the security of our computers. On multiple occasions this has caused us to be completely unable to do experiments. We come in the next day, and things have just totally broken. Since some of our hardware is network attached, we can no longer see those devices. When there is an update that's rolled out to us it goes to every single computer and it's automatically done and we have no real choice in the matter. What we really need is a little bit more of an

individualized plan. And since we don't deal with patient health information, having some compromises there would be really helpful to us. (ITAC discussion with Basic Sciences)

- Sharing resources between the medical side and campus side of Duke is tricky because of the incredible diversity on the university side. This is especially challenging when data sharing involves undergraduates and graduate students who have never worked on issues of data compliance and data transfer. Finally, cost-sharing while often needed, does not address the needs of graduate students and faculty who do not have funding and undergraduates who are being pushed toward research involvement. (Social Sciences Faculty Working Group) [Also applies to recommendation 3.]
- Some security requirements impede current initiatives, a review of current standards would be helpful to ensure it's not too stringent. Specifically, local/shared account needs vs. NetID – Service Accounts, new computer setup (DHTS mandated requirements vs. researcher customization needs), and a general need to find the balance between security and creativity. (Basic Sciences Faculty Working Group)

Recommendation 4: In collaboration with faculty representatives, the VP for Research and Innovation together with the DU CIO should launch an evaluation of current policy, security and compliance requirements and processes for research toward a holistic risk-based institutional approach. The goal should be to look for minor adaptations to existing policies or seek new solutions (if warranted) in cases where the adaptations better balance research practices / needs with risk, but without compromising regulatory requirements or the protection of the institution against the sensitive data loss. The effort should also reflect the need for nuanced policies or implementations for social science and campus researchers working with sensitive data that are not predicated or based on similar DUHS, School of Medicine or School of Nursing clinical research policies or implementations. Ongoing research-related governance bodies should include sufficient faculty participation to incorporate representative feedback across the diversity of Duke's research enterprise and work towards increased transparency about the process and decision making.

Technology Findings and Recommendations

Finding D: The Duke Compute Cluster (DCC) and RAPID VM services, while quite valuable and cost effective, are not as expansive as many faculty require. For some, it is difficult or less attractive to use for certain teaching and specialized research tasks. On a continuing basis, and as our needs and technology evolve, OIT should work with ITAC and academic leadership to evaluate and propose changes or additions to research computing services, and to enhance the guidance OIT offers to its users.

Regarding the need for extending the infrastructure for the Research Computing relative to Finding D:

- For GPUs, current internal solutions take too long to reconfigure or reallocate, when the time needed to use the resources is minimal. Sometimes research or teaching need better access to "bare metal". Getting access to cloud resources is proving difficult and time consuming, and has become easier to just pay for outside solutions. (Engineering Faculty Working Group)
- In Physics our biggest users are heavy consumers of computational resources and they have needs that far outstrip what OIT could conceivably offer, so we have to go off-site to major facilities, which involves writing grant applications to get our CPU hours. The competition for these is getting more and more commoditized, and it is more and more woven into a lot of things that we do in the research world. Generally speaking, it behooves an institution

like Duke to provide a base level of capabilities in terms of computing and storage that can satisfy these general needs, setting aside the millions of hours per year that really high-end labs need. (Steffen Bass in ITAC discussion with Natural Sciences; Similar observations made by Chemistry and Engineering faculty)

- Access to very large High-Performance Computing (HPC) resources is possible through external proposals but can mean time-consuming, yearly repeated proposals at multiple centers, and resources can fluctuate wildly. Moreover, HPC in teaching is not routinely possible at external HPC resources and key expertise cannot be adequately incorporated into classes, and undergraduate and Masters research is similarly hampered by reliance on external resources. Stable, accessible mid-scale HPC at Duke would address a significant gap and solve the competitive disadvantage for HPC-based research and teaching. (Volker Blum in ITAC discussion with Engineering)
- Local servers and computational clusters would provide great benefit. (Basic Sciences Faculty Working Group)
- Many researchers are having to go outside Duke to peer institutions and national labs to use certain clusters; it would be a boon to the University and researchers alike to have the means to run small to medium scale jobs with a home cluster. (Natural Sciences Faculty Working Group)
- Several years ago, the Provost agreed to commit to covering 80% of all the computing needs of the faculty, recognizing the 20% of the most computationally intensive users would need to utilize grants or other approaches to meet those needs. Recognizing those needs may have changed over time, we should reassess whether the level of "entitlement" computing being provided by OIT still meets that 80% threshold. (ITAC discussion with Natural Sciences)
- There's an unmet need for mid-scale computing resources that would provide more than the Duke Compute Cluster but less than provided by a national lab. This includes fast interconnection and a diversity of hardware. (Engineering Faculty Working Group)
- NIH funded data storage retention requirements need to be recognized by both DHS and DU as a critical component of infrastructure. (Basic Sciences Faculty Working Group)
- Students remote into desktops in Classroom 6 and also have their own their own [OIT provided] VM. We're working with the [OIT] VM group and they're very enthusiastic about testing scalability of VM for these more intensive applications so we've given them two case studies of courses that would typically not work well at a low level VM but we're going to be testing them out. (Augustus Wendell and Ed Triplett in ITAC Discussion with Humanities/Arts)
- VMs should be expanded to provide many GPUs. (ITAC Discussion with Engineering)
- Colleagues in physics and chemistry report insufficient raw compute power in the DCC to meet common needs that are fairly routine within their field; rather, these need to be pursued with collaborators at other institutions who have more raw compute power. (ITAC discussion with Natural Sciences)

Recommendation 5: Evaluate the cost and funding approach to extend the capability of Research Computing by expanded HPC/GPU services so that they are better sized to meet demand and are easier to use; review with ITAC and academic leadership to determine viable approaches that can improve support and are financially feasible. Establish a more routine review cycle with ITAC faculty (or other faculty advisory bodies) to better evolve services with need.

Further to Finding D and the need for tuning of Research Computing solutions to satisfy different uses:

- We have a challenge with students wanting to work on their laptops which, in the computational media realm, are often underpowered for the work that's needed. This is a very difficult dialogue to have with students and comes back to this issue of where do they work and what's the studio culture when we're moving very quickly through commercial software that's very intensive. We're working now to explore VM options. (ITAC Discussion with Humanities/Arts)
- Data collection, statistical analysis, and machine learning are the three cores that require the most additional support and resources. We need support for data collection and analysis (e.g., survey design) and perhaps there is need for additional access to different types of compute and storage depending on analysis needs and/or sensitivity of data. (Social Sciences Faculty Working Group)
- I have had experiences where hardware issues have caused students to drop my GIS course. I am a huge advocate for GIS and other spatial software, so it is disappointing when I see a graduate student get discouraged and move away from mapping or modeling for their thesis project because of access issues. I also think physical labs are so important because this is where students teach each other and learn by looking over each other's shoulders. Dedicated computing spaces are critical for building communities in our project-based courses and projects – especially Bass Connections, Code+, Data+, Story+ etc. (Ed Triplett in ITAC discussion with Humanities/Arts)
- We often struggle to maintain a compute infrastructure that exists, or almost exists, today in our local labs, but every now and then needs to be reconfigured in the proper way. Lab personnel can use it to meet the CPU and GPU operational needs. We have moved more towards using deep learning and we have our own resources, but they are a challenge to maintain and operate. Is this something that the university is addressing at all? We need not only the proper infrastructure but the systems that can support the image acquisition and instrumentation requirements that we have. (Alexandra Badea and Lindsey Glickfeld in ITAC Discussion with Basic Sciences) [Also applies to recommendation 3.]
- Hannah Jacobs does a huge amount of work with us and helping manage the digital art history and visual culture lab where she'll work with us to develop the image for the computers every year and help make sure that the faculty members and student teams that are creating work there has have input into that process so that we're well equipped for the upcoming year. (ITAC Discussion with Humanities/Arts)
- Using Duke standard VM for Python works pretty well but one issue is that versions change all the time. So we will get something up and working and then, the version will change and OIT will upgrade the version, leaving the code that previously worked not work anymore, and though it is not hard to fix, it is a continuous loop that requires time. There have also been compatibility issues with getting JAX to work with Python. JAX is a Google library that is useful for automatic differentiation. These compatibility issues led to the use of Google's CoLab environment which my students and post-docs requested and costs $10/month. (Henry Pfister in ITAC Discussion with Engineering)
- I am starting to brush up against machine learning resources needed in classes. This is where we need a GPU enabled lab for courses that require that sort of horsepower and capacity. (Augustus Wendell in ITAC Discussion with Humanities/Arts)
- Communication system simulations use Python code running on 100 to 200 cores in parallel. Doing this in the Duke Compute Cluster (DCC) works great and the scavenger ability applied at night can grab hundreds of CPUs. One challenge is that simulations have to be broken into pieces to run and then, be put back together. However, with good timing and some luck it's possible to use the scavenge queue to get a machine with sufficient cores (e.g.,

128) so that all of the parallelization can be done from one Jupyter notebook. So the DCC has both pros (can grab a large number of scavenger/common nodes) and cons (setup/teardown). (ITAC Discussion w/ Engineering)

- Regarding use of machine learning resources in classes, this is where we need a GPU enabled lab for courses that require that sort of horsepower and capacity. Sometimes the timeline and deployment into fully managed labs is challenging and it's totally understood that there is a long timeline for that. This gets harder when revisions to software is happening relatively quickly and we may not have 9-12 months to prepare for the impact of those changes in our VR environments. (ITAC Discussion with Humanities/Arts)
- There is a need for commodity hardware (including GPUs) and a lot of specialized software. Also, laser scanning photogrammetry is needed. (Humanities/Arts Faculty Working Group)

Recommendation 6: Establish one or more Faculty Working Groups to help guide OIT in tuning Research Computing offerings to better support data collection, statistical analysis, and machine learning, as well as common programming and graphics-intensive needs. New approaches need to be considered for the extremely data intensive needs of instrumentation, along with the financial models that can support them (funded via grants, indirect cost recover overhead, subvention, or otherwise).

Finding E: Insufficient awareness exists about computational options and there is uncertainty of when to use which as many options are often presented and left to the faculty to choose. This lack of awareness, coupled with complexity in navigating the options, stymies researcher ability to easily access the resources they require.

Specific to the need to improving the navigation of the varied options on-premises and in the cloud:

- There is concern regarding costs of the SoM direction of moving computation completely or largely into the cloud, as well as with the seemingly abrupt timeline researchers are expected to make that shift, with limited or no material support for making the conversion. Moreover, for groups with data hosted on an OIT VM there often is no ingress or egress path into their secure storage blobs in the Azure (cloud) system that researchers can use, and the solutions offered were less than ideal. This may drive basic sciences researchers to adopt the DCC environment. (ITAC discussion with Natural Sciences) [Also applies to recommendation 3.]
- While there is not a prevailing preference among cloud solutions, there has been a great deal of effort put into Azure and establishing infrastructure there. It would be welcome to have Duke better facilitate access to other cloud services. (Engineering Faculty Working Group)
- Large data sets need to be moved quickly between the imaging system and associated reconstruction computation, which today requires some sort of heavy-duty local compute and keeping the storage very close to the imaging scanners. This is because you need to offload the data from the scanner to a backup, which means network speed becomes a concern, and which often drives a need for local (departmental) compute solutions. Guidance and assistance navigating other options is needed, especially as SoM is increasingly emphasizing (and financially incentivizing) cloud solutions. (ITAC Discussion with Basic Sciences) [Also applies to recommendation 3.]
- Support for classes has been enabled in the past through connections with Azure. But every year, this has to be reestablished – every year students leave not wanting to work on the cloud because of this experience which is not good. (ITAC Discussion with Engineering)
- Recent thesis committees or committee meetings involved students making heavy use of GPUs. Both pushed the limits of what's available to them via the DCC, and the uses would

be even more difficult if the students tried to go off-site and spin up cases to use TensorFlow or resources in Google's cloud. (ITAC discussion with Natural Sciences)

- There are pros and cons to formalizing a Google CoLab service offering for Duke. Advantages: more flexible for reconfiguration than OIT VMs; many repos include scripts designed for Google CoLab; easy to share code with people outside of Duke; disadvantages: no common file storage; billing is complicated. The question is whether institutional access to Google CoLab could help to resolve this, and how within Duke could we enable the capability broadly for teaching and research? (ITAC Discussion w/ Engineering)

Recommendation 7: Introduce guidance and simplify navigation of technical solutions for faculty and students by consolidating technical offerings to more broadly support research and academic use cases or needs (including labs with substantial data generation or acquisition from specialized instruments). Solutions should: encompass the full breadth of services available, including local and cloud options; be established as an evolving service; and be developed in consultation with faculty to 'field test' the efficacy of the guidance and navigation across the range of Duke's research domains.

Finding F: The needs for research data storage have increased exponentially; current storage services address the need for active, near-term research storage but are not easy enough to use and share and do not incorporate solutions that are affordable for long-term and archival storage. Moreover, current options do not reasonably accommodate NIH and NSF 'unfunded mandates' for long-term data retention in ways that are either scalable or cost effective.

The urgent need for storage solutions to meet research needs and satisfy compliance requirements was consistently conveyed:

- There is a growing need for access to long-term storage for imagery/data that is curated and managed by researchers over long periods of time. There are concerns over whether the University will continue the current level of support, where responsibility for maintaining these assets should fall, and how to pay for this storage when associated grant funding runs out. (Basic Sciences Faculty Working Group)
- On the topic of project archives, there is great need, particularly in the visualization space, for archiving models. (E.g. A class building a digital archive needs to be retained long-term, and individual student projects need archiving). (Humanities/Arts Faculty Working Group)
- The Humanities have an ongoing need for support for public-facing "hot" storage of research data and also the development of bespoke research products such as websites, repositories, databases, digital projects, etc. Many of us are fulfilling these needs with external hosting solutions, prepackaged systems that minimize the need for local programming and support but can be very expensive. These create reliance on external providers and because they not under the Duke "umbrella" it creates sustainability challenges. (ITAC Discussion with Humanities/Arts)
- The biggest simulation in one lab, a full-body simulation at large scale, required 140.7 TB to keep the fluid in memory. Since maxxing out memory is an issue, the lab needs to run on Amazon Aurora. We need to be able to store our data into the future as required by NIH grants so long-term storage is needed. Blue gene/q total system memory requires 1.6 PBs. Then, if we are processing 1PB of data and creating 1 PB of data, and running millions of steps, we run into lots of data problems. (Engineering Faculty Working Group)
- Data management is particularly tricky for projects funded by NIH, which has a relatively short span of funding on a given grant, and which is shorter than the required data

retention period. We need a reasonable approach going forward to manage these costs, and deal with all these requirements. (ITAC discussion with Basic Sciences)

- More space is needed to experiment, for students working in the web or in other digital space who need to spin up applications. (Humanities/Arts Faculty Working Group)
- Storage (especially long-term) is critical and becoming more and more of an issue (currently maxed out at 700TB). Need storage for quick access for Machine Learning (ML) workloads; also, need archival storage. (ITAC Discussion with Engineering)
- Solutions to managing large datasets would continue to be welcome, Box works to a varying degree, the expectation has become to not work with files at the same time on the same share. (Humanities/Arts Faculty Working Group)

Recommendation 8: Implement new long-term storage options for a variety of use-cases and ensure solutions are available to not only campus but also School of Medicine and Nursing researchers; establish funding models with leadership and faculty to ensure sustainability.

Likewise, the difficulties for faculty and risks to Duke of the current model where faculty are solely responsible for navigating among storage options were compellingly described:

- In tandem with long-term storage needs, there is growing concern that researchers may be unaware of the storage options currently available to them, and have resorted to more rudimentary means (e.g., external hard drives). (Basic Sciences Faculty Working Group)
- Hot (active) storage is through an Isilon system and SoM has established a new S3 storage system for cold storage. The steps to get data from hot to cold storage is inefficient, manual and complex. "I'm personally responsible for dragging and dropping those data […] and so that is a lot of overhead. It makes me nervous given that I'm also the one to delete the data off of Isilon. Having more automated systems with checks and things would be really helpful." (Lindsey Glickfeld in ITAC Discussion with Basic Science)
- The ideal hybrid storage environment enables us to move into the cheapest storage possible, based on need. Much of the data requires slow and cheap (cold) storage but then there can be issues with later finding the data, which is being addressed by OIT. OIT is being deliberate about how to back up data. Opensource CEPH storage may provide a useful option and it is being looked at. Also, partnering with national labs would be helpful so we can learn from them. E.g., can we learn from CERN? (Engineering Faculty Working Group)
- Things have often seemed to be a moving target in DHTS like take object storage which we've talked about-- it's great, it's fantastic, it's cheap but if we can't get our data off when it moved from on prem object storage, we were all caught. When researchers are suddenly saddled with needing to move it's frustrating and there doesn't seem to be a long term plan other than throwing the next point solution out. (ITAC Discussion with Basic Sciences) [Also applies to recommendation 3.]
- Dynamic solutions are needed to: A.) navigate data from "hot" to "cold" storage based on project stage and B.) share data both within Duke and across institutions to manage data availability requirements from the NIH. (Basic Sciences Faculty Working Group)

Recommendation 9: Ensure the new long-term storage options along with existing and future storage solutions are approached as an eco-system, recognizing the differences in storage needs across the research lifecycle and automating to the greatest degree the movement of the data across solutions to reduce the burden on faculty and risks to Duke. Make these options available to both campus and SoM users with attention to differences in datatypes and data storage needs in the science versus humanities versus social sciences.

Finally, a potentially transformative storage and data-related need was convincingly championed largely by Social Sciences but with the prospect that the need could extend broadly:

- There is an opportunity to develop shared data resources maintained centrally to reduce duplication of effort, improve oversight and compliance, and increase access to data among Duke investigators. Additionally, specialized assistance is needed in navigating the necessary regulatory needs and demands. (ITAC Discussion with Social Sciences) [Also substantiates recommendation 4.]
- Assistance in building and maintaining large data resources, that in turn could be actively shared and utilized throughout the Social Science community (internal and external communities). (Social Sciences Faculty Working Group)
- Since my research is interdisciplinary, I can no longer look to only databases from my discipline. There is an opportunity to develop shared data resources maintained centrally to reduce duplication of effort, improve oversight and compliance, and increase access to data among Duke investigators. Coming here from Stanford they had a person who helped to do this. Since many of the data sets we work with are purchased, they have clearly defined restrictions on how they could be used and shared, which is often defined as don't share outside of the campus, so these data sets were mainly resources for faculty associated with Stanford. A similar commitment to sharing data at Duke would help to make data more available and would require having specific staff to help acquire the datasets, and address issues upfront around data protections, sharing limitations and other technical issues. (Kate Bundorf in ITAC discussion with Social Sciences)
- Nirvana for the social sciences researchers at Duke would be a place where investigators could go when needing to access a data set. There would be coordination across the oversight organizations and bodies where all could go for help on IRB (Institutional Review Board) requirement fulfillment, and for needed training on maintaining confidentiality. There would be clear, organized, and maintained systems about constraints for accessing data. A safer and more efficient environment would be provided for researchers including undergraduate and graduate student researchers. If attractive data sets are provided and oversight requirements are clear and efficient, researchers will access them. (Kate Bundorf, in ITAC discussion with Social Sciences) [Also applies to recommendation 4.]

Recommendation 10: Evaluate the need and develop a proposal with estimated cost to establish a new service for researchers that would manage/license/administer shared access to expensive data resources/sets; carry out this work in collaboration with the prior Presidential Scholar who recently studied this need in greater detail on behalf of the Provost and President and in conjunction with the Strategy Team 2030 implementation team.