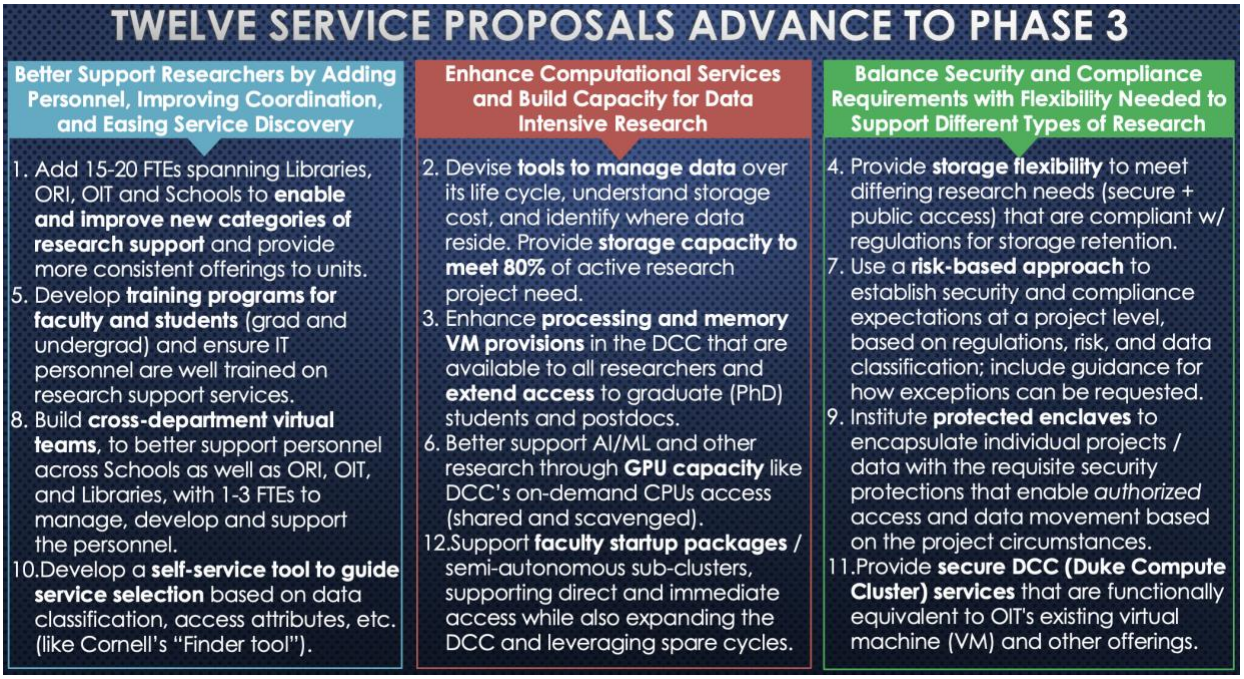
Executive Summary

In 2022, Duke conducted a comprehensive study of Research IT Needs, across all its non-clinical scholarly domains. Following the release of the December 2022 summary report, Duke's Vice President for Information Technology, Vice President for Research and Innovation and Vice Provost for Library Affairs teamed up with others to sponsor Phase 2 of the effort—developing service proposals to meet the Phase 1 needs expressed by faculty, which encompass much more than IT.

From January-May 2023, six cross-functional staff teams—each with faculty representation—drew up 39 proposed services to address Phase 1 expressed needs. After further service refinement and faculty and sponsor feedback, twelve services are recommended for implementation and they comprise three overarching *service clusters*, illustrated below, with the relative service priority of each enumerated:



Together these three overarching service clusters and their twelve services reflect a **coordinated and integrated research support program** across the Office for Research and Innovation (ORI), Office of Information Technology (OIT), Duke University Libraries (DUL) and others, in conjunction with Schools. Five of the twelve proposed service are already being actively pursued by service partners.

These services advance to Phase 3 which will focus on the funding approaches to implement the services (July-September 2023). Phase 3 will be led by financial experts and is likely to involve a multimillion-dollar funding increment on top of the indirect cost recoveries already in place today that support Duke's research endeavor (in excess of $300M). The funding approaches identified in the Phase 3 process are expected to vary for different services, from university (allocation) funded, to direct-to-grant funded, to overhead (indirect cost recovery) funded. Philanthropy may also be relevant and in some cases a service might be funded in part or in full by internal budget reallocation at the service provider level (but this is assumed to be the exception rather than the norm).

Beyond these twelve services, other proposed services were compelling and may represent targets for future funding or local (School-specific) pilots. As one survey respondent put it, *"the solutions are so refined that they ALL sound nearly equally compelling."* Of note, two pilots are already being pursued with Engineering for services that were not advanced to Phase 3 because their need was more localized.

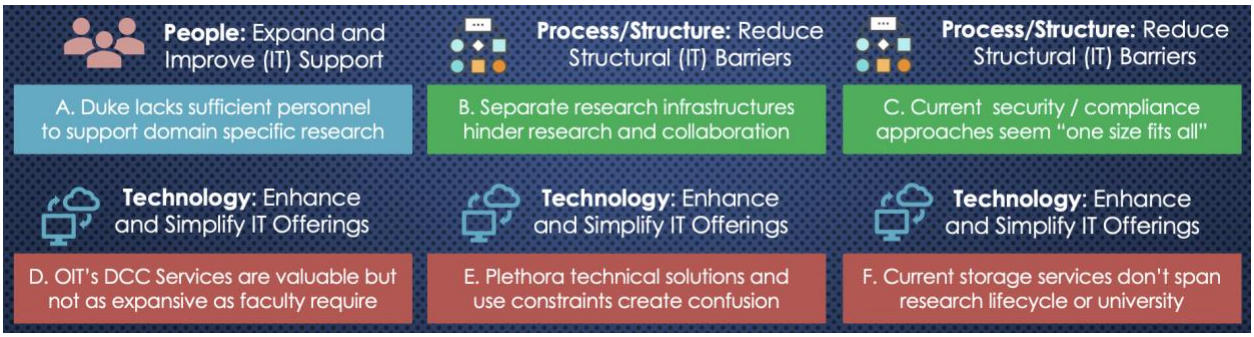
Background and Context for Phase 2

From February-November 2022, Duke's Information Technology Advisory Council (ITAC) undertook an assessment of Research IT Needs at Duke. The assessment invited participation from 37 faculty and 2 research/teaching staff[1] who were identified by deans and drawn from non-clinical scholarly domains. Their input was synthesized into six major findings and ten recommendations reflecting areas of common, perceived need. It is important to note that various findings and recommendations extended beyond the IT domain to encompass research support more generally.

The result of the assessment was presented to Academic Council on December 1, and a summary report[2] was simultaneously released to document the process and outcomes. Soon after, the Vice President for Information Technology, Vice President for Research and Innovation and Vice Provost for Library Affairs joined together as the three primary sponsors in establishing Phase 2 of the Research IT Needs assessment. The Phase 2 work was designed to be carried out by six cross-functional teams, whose purpose was to identify, develop and prioritize service proposals or projects that would be responsive to the recommendations identified by researchers in the Phase 1 effort.

Phase 2 Working Group Formation and Service Proposal Development and Consolidation

In January 2023, the primary sponsors identified charges and membership for six working groups (teams), one for each of the Phase 1 finding in areas of People, Process/Structure and Technology. The primary sponsors solicited eight other leaders (deans and other administrative executives), each of whom joined in sponsoring one of the Phase 2 working groups. The teams were identified as Groups A-F, corresponding to the finding from Phase 1 on which they were tasked to work.



Each team was largely comprised of staff drawn from ORI, DUL, OIT, and DHTS (Duke Health Technology Solutions), with others. The concentration of membership from these units was acknowledgement that those groups were likely providers of future services and solutions arising from Phase 2. Each group also included at least two faculty champions and other faculty consultees who were selected to monitor the working group's emerging service and project proposals to ensure the process remained researcher-centric and solutions were responsive to the faculty needs. A total of 55 individuals—26 staff and 29 faculty—were invited to participate in the six Phase 2 working groups, serving in distinct roles as Leads, Members, Faculty Champions, Faculty Consultees, Consulting Experts, and Staff Facilitators. Six of the faculty invited to engage in Phase 2 were active participants from Phase 1 and their inclusion provided a feedback loop between the phases. Other Phase 2 faculty were drawn from ITAC (9) to retain linkage to the body that initiated and oversaw Phase 1, and the balance (14) were drawn from faculty at large in response to Phase 1 feedback from deans, the provost, Academic Council, and others. (See Appendix A for membership and charges.)

[1] For simplicity, this report will refer to both the 39 faculty and 2 research/teaching staff participants in Phase 1 as faculty.
[2] The full report from Phase 1 is available at https://duke.is/72sjn.

Working groups met with sponsors in February to receive their charge, then each group met weekly over the next ten weeks to review Phase 1 findings, evaluate existing resources and associated gaps, and identify solutions they believed would be responsive to the needs expressed in Phase 1. Meeting frequency for Phase 2 participants was based on role, with each team's 3-4 leads called on more extensively (weekly) than members (bi-weekly) or faculty champions (3 times in the 10 weeks). Faculty had the option to engage as extensively as they wished, some choosing to participate deeply and others electing limited participation. Consultees /experts were called upon only as needed.

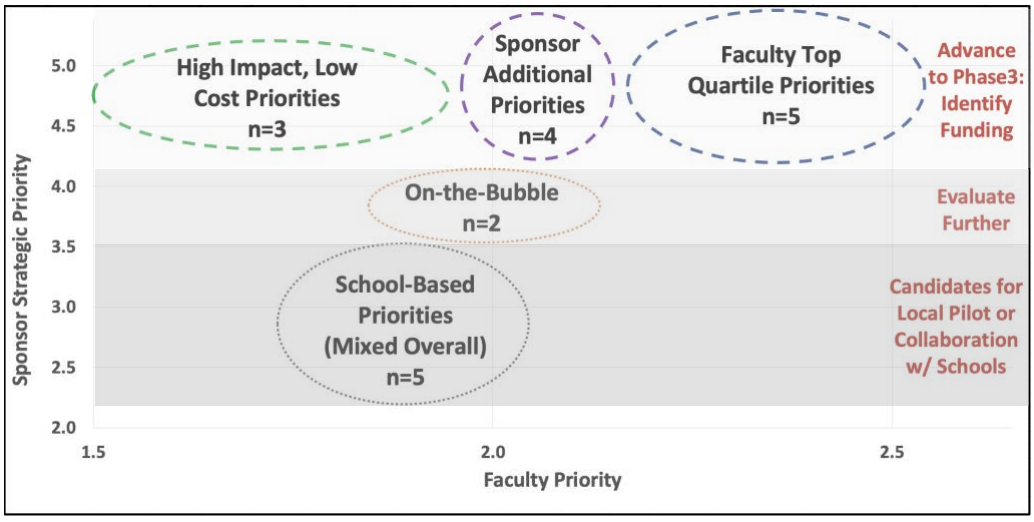
By early April, the six working groups had identified 39 potential services they believed could help meet the researcher needs expressed in Phase 1. These services are described in Appendix B. On April 6 a poster session was organized for all six groups, with sponsors and participants from both phases of the project invited. More than 50 individuals attended, about half representing Phase 2 staff leads, members, consulting experts and facilitators, and the other half consisting of faculty and sponsors. The poster session stimulated conversation across the various working groups and faculty provided feedback regarding the 39 proposed services (see Appendix C). As a result, ten service proposals were consolidated into other, similar proposals, leaving 29 distinct services / proposals for further consideration. (Appendix D details the consolidation process.)

During April and early May, a readout for each working group was held with its sponsors to discuss and refine proposed services. The status of the Phase 2 effort also was reviewed with several groups at the end of the academic year: with Deans Cabinet on April 24, with the Faculty Subcommittee of EROC (Executive Research Oversight Committee) on May 2 and with ITAC on May 11.

In May, poster session feedback, sponsor readouts, and guidance from a faculty expert in survey design were used to further pare the 29 distinct services/projects down the 21 service proposals that were most responsive to the overall needs of researchers as expressed in Phase 1. (See Appendix D.)

Prioritization Process and Resulting Categories of Recommended Outcomes

Proposed services were rated by faculty and sponsors, then graphed. As the following conceptual graph illustrates, they fell into five broad categories with three associated outcomes: twelve services are recommended to advance to Phase 3 for funding strategy development; two services require further evaluation; and at least five services may represent partnering opportunities with Schools.



Appendix E provides an actual graph showing how each distinct Phase 2 service aligns with the different categories (or falls outside them), based on faculty and sponsor ratings, as well as cost. Categories and related services are described next, with *estimated* costs noted in shorthand ($-$$$).

Faculty Top Quartile Priorities: Faculty from Phases 1&2 (n=58) were surveyed regarding proposed services. Response rate was an impressive 67% overall and based on mean faculty ratings[3], five top quartile services emerged as highest priority (each with an average score >2.2 score on 3-point scale). The faculty survey responses and write-in comments appear in Appendix F.

1. **Add 15-20 FTEs** spanning Libraries, ORI, OIT and Schools **to enable and improve new categories of research support** and provide more consistent offerings to units. ($$$)
2. **Devise tools to manage data over its life cycle**, understand storage cost, and clarify where data reside. Provide **storage capacity to meet 80%** of active research project need. ($$)
3. **Enhance VM provisions (processing / memory) in the Duke Compute Cluster (DCC)** that are provided to researchers; **extend access** to graduate (PhD) students and postdocs. ($)
4. Provide **storage flexibility** to meet differing research needs (secure + public access) that are **compliant w/ regulations for storage retention**. ($$$)
5. **Develop training programs for faculty and students** (grad and undergrad) and ensure IT personnel are well trained on research support services. ($$)

Sponsor Additional Priorities: Sponsors next rated strategic impact of the proposed services[4] and four more service priorities resulted. Each garnering an average sponsor rating >4.5 and was also highly rated by faculty (scoring above the median).

6. Better support AI/ML and other research through **GPU capacity** in the DCC, similar to the DCC's on-demand CPUs access (shared and scavenged). ($)
7. Use a **risk-based approach** to establish security and compliance expectations at a project level, based on regulations, risk, and data classification; include guidance for how exceptions can be requested. ($$)
8. Build **cross-department virtual teams**, to better support personnel across Schools and in ORI, OIT, and Libraries, using 1-3 FTEs to manage, develop and support the personnel. ($$)
9. Institute **protected enclaves** to encapsulate individual project data with requisite security protections; enable *authorized* access/data movement based on the project circumstances. ($$)

High Impact/Low-Cost Priorities: Finally, *very rough cost estimate ranges* were developed by working group staff, as low ($, <$150K), medium ($$, $150K-$750K), or high ($$$, >$750K). These *estimates* define the bubble size in the graph in Appendix E and are noted for each service above and below. Three additional services emerge as a result, each with high strategic impact and lower estimated cost:

10. Develop a **self-service tool to guide service selection** based on data classification, access attributes, etc. (like Cornell's "Finder tool"). ($)
11. Provide **secure DCC services** that are functionally equivalent to OIT's existing virtual machine (VM) and other offerings. ($)
12. Support **faculty startup packages**/semi-autonomous sub-clusters, delivering priority / immediate access to 'owners' while expanding the DCC and leveraging spare cycles. ($)

These twelve enumerated services reflect both needs of researchers and institutional priorities as conveyed by sponsors. They are recommended to advance to Phase 3, for the purpose of identifying appropriate ongoing (and as needed, bridge) funding to implement each service, ideally by FY25, either as pilots or production services.

[3] NB: Eight lower-rated services were excluded from the Faculty Survey but were placed along the X-axis of the Appendix E bubble graph based on Poster Session ratings of those services, relative to other service ratings.
[4] Sponsors rated all 29 consolidated services/projects, incorporating the eight lower/deferred priority items not presented in the faculty survey. This was in recognition that there could be certain services of high strategic value to the institution, but which would not necessarily be highly rated by individual faculty.

These twelve services recommended for Phase 3 form three broad *service clusters*, with the number beside each service corresponding to its overall priority as determined by the process detailed above.

On-the-Bubble Service Priorities: Two other services are worthy of further evaluation: both had reasonably high ratings by sponsors (4.0-4.25), but each received slightly lower faculty ratings and has an *estimated* annual cost that requires further financial validation:

- Create **Data Continuity Services that ensure data integrity and availability**, including providing the storage associated with maintaining data continuity. ($$-$$$)
- Create a **single, central protected research network** rather than the separate ones provided by OIT and DHTS. ($-$$)

Of note, a project to pilot moving a Basic Sciences unit to OIT's network is being explored.

These two bulleted services, while not recommended as initial targets for Phase 3, deserve further study as to their approach, feasibility, and cost. Teams from ORI, OIT, Libraries and DHTS should more fully evaluate each, with a goal of developing detailed service proposals for consideration and / or pilot implementations in FY25.

School Based Priorities: Variation in priorities across scholarly domains motivates further discussion. Appendix G gives more detail on services with high domain-specific ratings but not promoted in the global process. One service highly rated by Basic Sciences/Nursing (create a single, central protected research network) already appears in the On-the-Bubble category above, with its strategy identified.

In addition, five other services rated highly within one or two domains, but not overall. These reflect collaborative opportunities with specific service partners (designated to the right of each service):

a)	Improve and clarify **storage and computational options approved for regulated research**, highly ranked by both Engineering and Social Sciences/Policy. ($)	with ORI, OIT and, Libraries:
b)	Add 4th data classification; where feasible, **ease requirements on non-regulated sensitive data**, a priority with Basic Sciences/Nursing. ($)	
c)	Create an **education cluster** with CPU and GPU virtual machines to support course needs, among the highest priorities for Engineering. ($)	with OIT:
d)	Improve **web-based access to DCC resources**, highly rated by Humanities/Arts. ($)	
e)	Facilitate **Cloud AND On-Prem options**, including SOM researcher access OIT's services for non-clinical research where relevant, sought by Social Sciences/Policy. ($)	w/ ORI, OIT &, DHTS:

These five lettered services are not recommended to Phase 3 but reflect partnering opportunities among sponsors and the deans. Ideally, such collaboration can lead to school-specific / bounded deployments in FY24 or FY25 that could be funded at the unit level, and with the potential of perhaps later rising to institutional-level services.

Of course, other services not called out above as collaborative opportunities can certainly be pursued, especially where the service has low estimated implementation costs or where creative approaches to implementation might be pursued. Consider, for example, two other services that were of particular interest to faculty in Natural Sciences/Environment (Create/optimize a special-purpose **VM environment for graphical intensive work**) and Engineering (Formalize/extend special purpose **FastMPI cluster**). In both cases the sponsors perceive opportunities to develop proposals with domain faculty to fund these services via foundation or federal agency grants. In the case of the Engineering FastMPI cluster, the graph in Appendix E shows two points, differenting its two cost approaches: one if funded institutionally and the other reflecting funding through grants.

This type of creative partnering could lead to implementation of considerably more than the twelve services initially prioritized and advanced to Phase 3. Of significance, this Research IT Needs process has already spurred creative implementation approaches for two pilots with Engineering. The FastMPI cluster referenced above is being funded primarily by Engineering, then opportunistically, an education cluster (School priority (c) above) will be created through the "trickle down" of some legacy Engineering MPI equipment, already operated by OIT. Each cluster's usage model will follow the DCC model, where a primary queue provides priority access to the designated Pratt function(s), and lower priority queue(s) provide other researchers throughout Duke with 'scavenger' access to unused computational nodes.

Conclusion and Next Steps

The twelve services enumerated on page 3 now advance to Phase 3, due to their high likelihood of enhancing Research IT (and related) support across Duke. These services will form a coordinated and integrated research support program across the Office of Information Technology (OIT), Office of Research and Innovation (ORI), Duke University Libraries (DUL) and others, in conjunction with Schools. They will be pursued via the three Service Clusters articulated on page 4. Service partners have already launched initial work for five of the twelve priority services with low-cost estimates ($), even ahead of the essential and anticipated funding commitments in Phase 3.

Although implementing these services requires an incremental multimillion-dollar investment, the process through Phases 1 & 2 reinforces that even despite Duke's extensive existing investment in research, new technological, regulatory, and competitive challenges demand appreciable, further investment for Duke to remain preeminent among research universities.

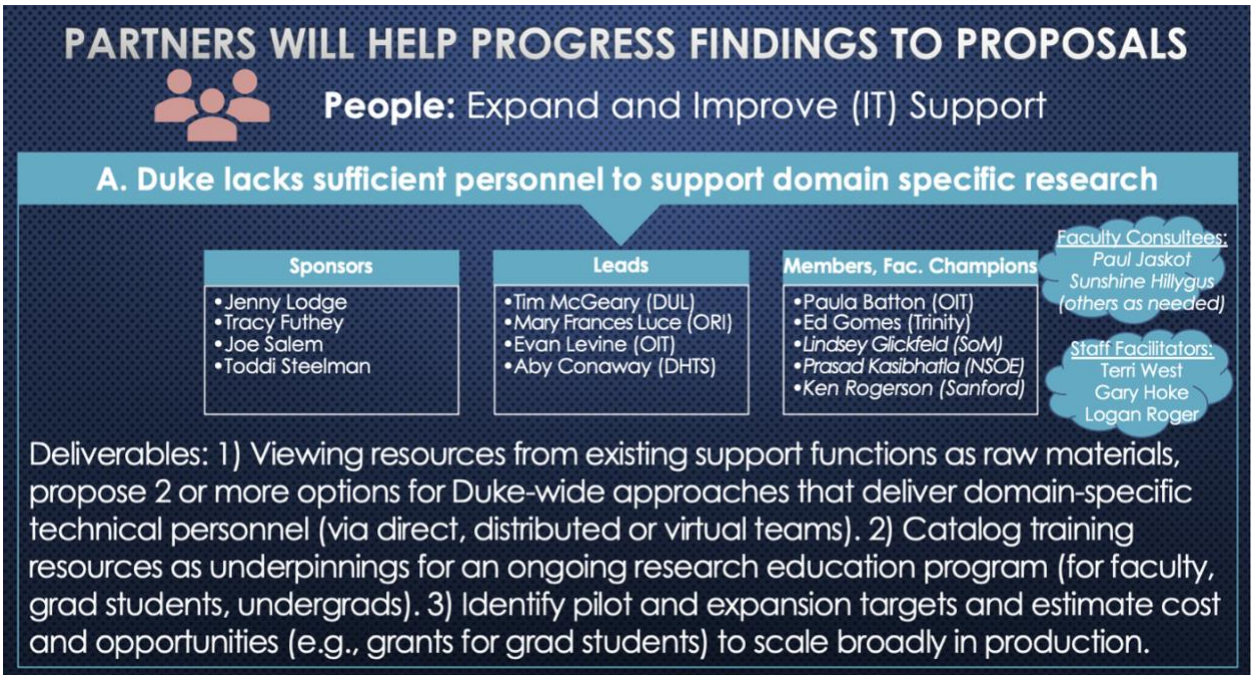
Phase 3 will aim to develop funding strategies for these services, in aggregate and individually. This will likely involve a combination of allocated funding, F&A/Indirect rate changes, services billed direct to grants, philanthropy or other approaches. Financial experts, along with sponsors and other leaders, will guide the Phase 3 work and develop financial proposals over the first quarter of FY24.

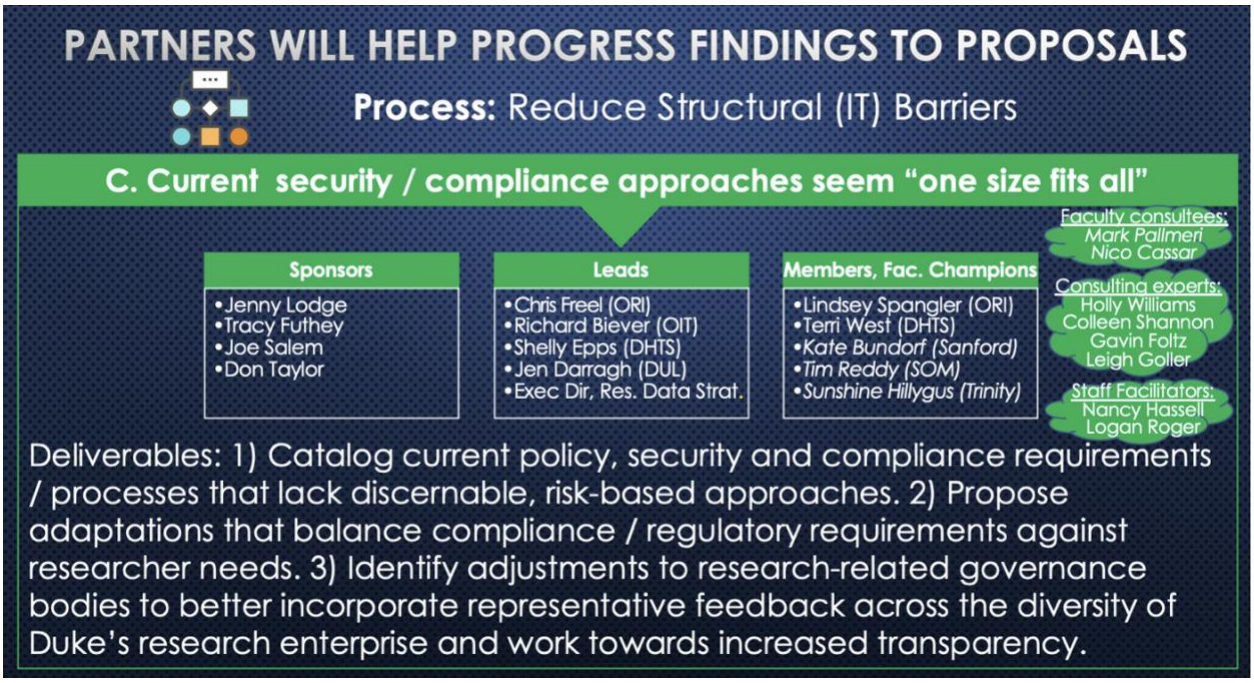
In parallel, two services identified on page 4 that do not advance to Phase 3 will be more fully studied by service providers / sponsors as potential targets for pilots in FY25. Further evaluation will refine their service definitions, cost estimates, and / or identify alternative approaches.

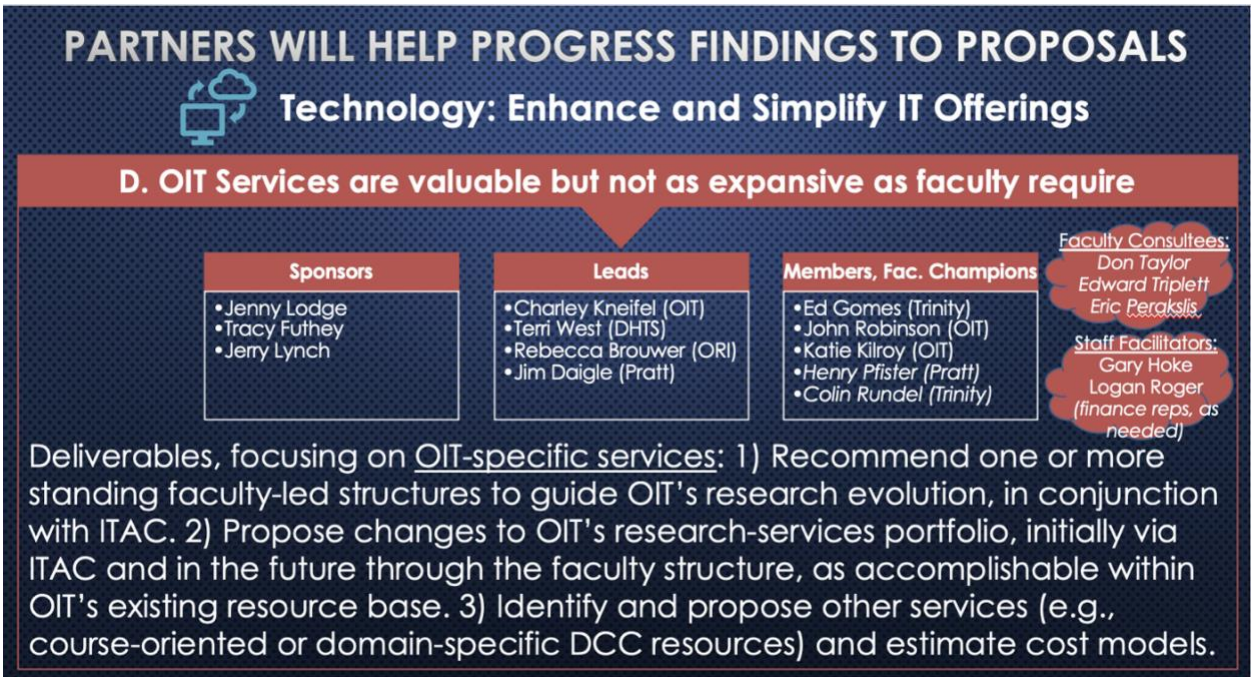
Finally, five further services designated as (a)-(e) on page 4 (and possibly others), reflect potential partnering opportunities among sponsors and individual Schools, at a pace and scale determined by the parties and based on localized priorities and resources at the service provider and School level. These services are estimated to be modest in cost (estimated <$150K), but because their value—at least initially—is believed to be more localized, they become opportunities to be funded by Schools.

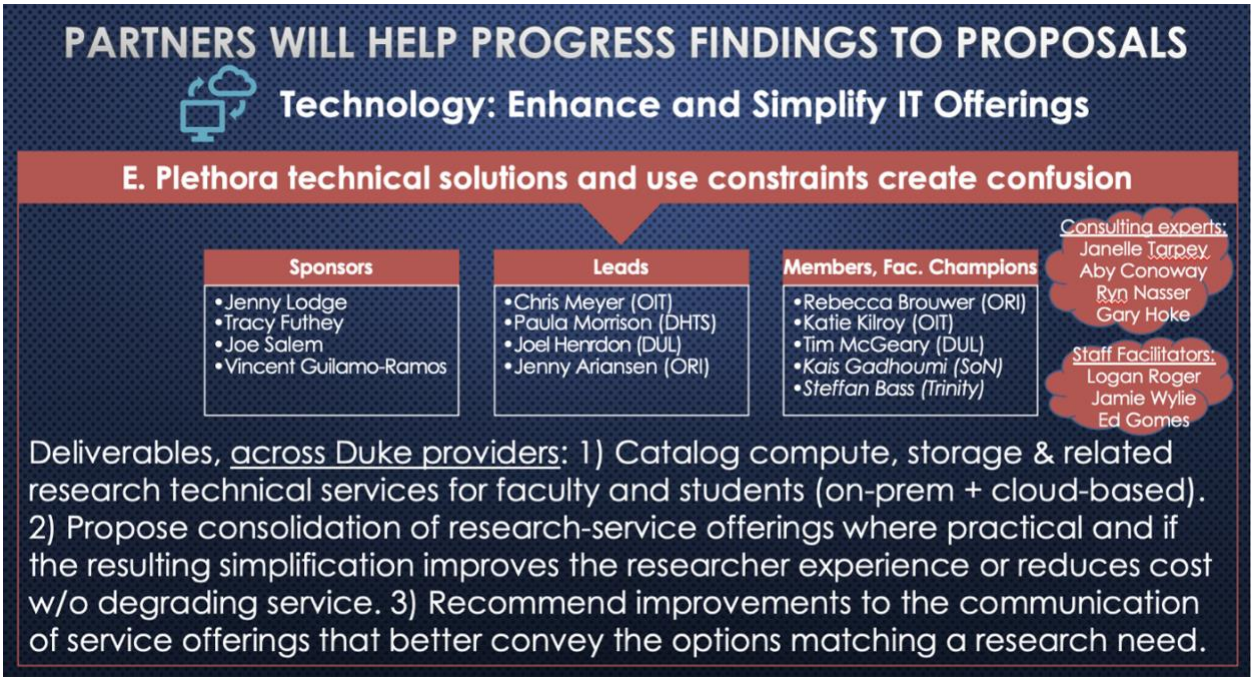
The sponsors acknowledge and thank the many faculty and staff who have contributed to the process to date and are optimistic that after Phase 3 and the twelve initial service implementations, that some non-prioritized services above may become candidates to pursue in a future stage of this project.

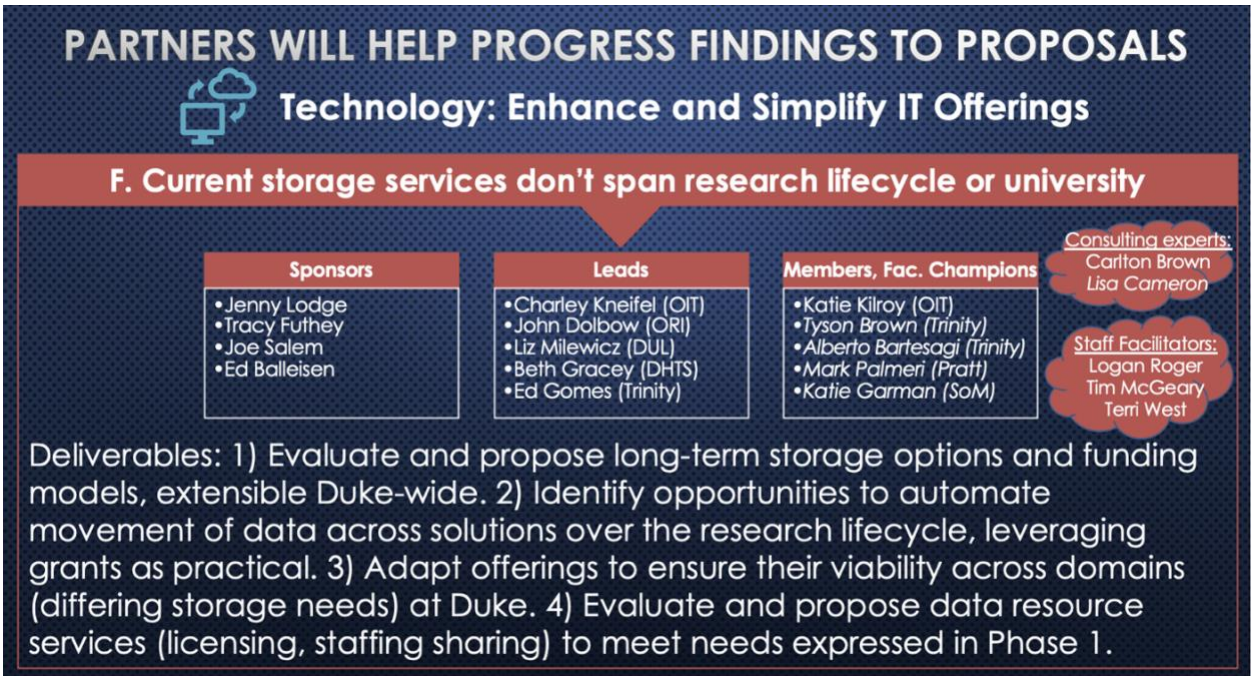
Appendix A: Membership and Charges for Phase 2 Working Groups




PARTNERS WILL HELP PROGRESS FINDINGS TO PROPOSALS
Process: Reduce Structural (IT) Barriers
C. Current security / compliance approaches seem "one size fits all"
Sponsors
•Jenny Lodge
•Tracy Futhey
•Joe Salem
•Don Taylor
Leads
•Chris Freel (ORI)
•Richard Biever (OIT)
•Shelly Epps (DHTS)
•Jen Darragh (DUL)
•Exec Dir, Res. Data Strat.
Members, Fac. Champions
•Lindsey Spangler (ORI)
•Terri West (DHTS)
•Kate Bundorf (Sanford)
•Tim Reddy (SOM)
•Sunshine Hillygus (Trinity)
Faculty consultees:
Mark Palmeri
Nico Cassar
Consulting experts:
Holly Williams
Colleen Shannon
Gavin Foltz
Leigh Goller
Staff Facilitators:
Nancy Hassell
Logan Roger
Deliverables: 1) Catalog current policy, security and compliance requirements / processes that lack discernable, risk-based approaches. 2) Propose adaptations that balance compliance / regulatory requirements against researcher needs. 3) Identify adjustments to research-related governance bodies to better incorporate representative feedback across the diversity of Duke's research enterprise and work towards increased transparency.

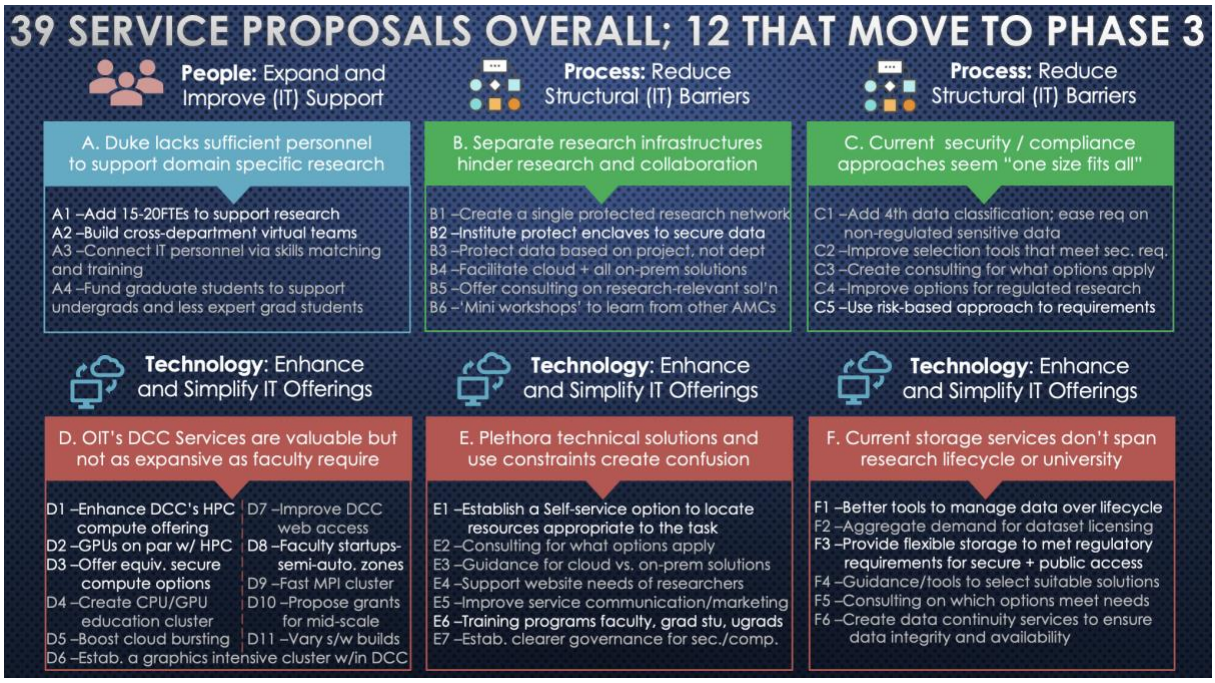

PARTNERS WILL HELP PROGRESS FINDINGS TO PROPOSALS
Technology: Enhance and Simplify IT Offerings
D. OIT Services are valuable but not as expansive as faculty require
Sponsors
•Jenny Lodge
•Tracy Futhey
•Jerry Lynch
Leads
•Charley Kneifel (OIT)
•Terri West (DHTS)
•Rebecca Brouwer (ORI)
•Jim Daigle (Pratt)
Members, Fac. Champions
•Ed Gomes (Trinity)
•John Robinson (OIT)
•Katie Kilroy (OIT)
•Henry Pfister (Pratt)
•Colin Rundel (Trinity)
Faculty Consultees:
Don Taylor
Edward Triplett
Eric Perakslis
Staff Facilitators:
Gary Hoke
Logan Roger
(finance reps, as needed)
Deliverables, focusing on OIT-specific services: 1) Recommend one or more standing faculty-led structures to guide OIT's research evolution, in conjunction with ITAC. 2) Propose changes to OIT's research-services portfolio, initially via ITAC and in the future through the faculty structure, as accomplishable within OIT's existing resource base. 3) Identify and propose other services (e.g., course-oriented or domain-specific DCC resources) and estimate cost models.


PARTNERS WILL HELP PROGRESS FINDINGS TO PROPOSALS
Technology: Enhance and Simplify IT Offerings
E. Plethora technical solutions and use constraints create confusion
Sponsors
•Jenny Lodge
•Tracy Futhey
•Joe Salem
•Vincent Guilamo-Ramos
Leads
•Chris Meyer (OIT)
•Paula Morrison (DHTS)
•Joel Henrdon (DUL)
•Jenny Ariansen (ORI)
Members, Fac. Champions
•Rebecca Brouwer (ORI)
•Katie Kilroy (OIT)
•Tim McGeary (DUL)
•Kais Gadhoumi (SoN)
•Steffan Bass (Trinity)
Consulting experts:
Janelle Tarpey
Aby Conoway
Ryn Nasser
Gary Hoke
Staff Facilitators:
Logan Roger
Jamie Wylie
Ed Gomes
Deliverables, across Duke providers: 1) Catalog compute, storage & related research technical services for faculty and students (on-prem + cloud-based). 2) Propose consolidation of research-service offerings where practical and if the resulting simplification improves the researcher experience or reduces cost w/o degrading service. 3) Recommend improvements to the communication of service offerings that better convey the options matching a research need.


PARTNERS WILL HELP PROGRESS FINDINGS TO PROPOSALS
Technology: Enhance and Simplify IT Offerings
F. Current storage services don't span research lifecycle or university
Sponsors
•Jenny Lodge
•Tracy Futhey
•Joe Salem
•Ed Balleisen
Leads
•Charley Kneifel (OIT)
•John Dolbow (ORI)
•Liz Milewicz (DUL)
•Beth Gracey (DHTS)
•Ed Gomes (Trinity)
Members, Fac. Champions
•Katie Kilroy (OIT)
•Tyson Brown (Trinity)
•Alberto Bartesagi (Trinity)
•Mark Palmeri (Pratt)
•Katie Garman (SoM)
Consulting experts:
Carlton Brown
Lisa Cameron
Staff Facilitators:
Logan Roger
Tim McGeary
Terri West
Deliverables: 1) Evaluate and propose long-term storage options and funding models, extensible Duke-wide. 2) Identify opportunities to automate movement of data across solutions over the research lifecycle, leveraging grants as practical. 3) Adapt offerings to ensure their viability across domains (differing storage needs) at Duke. 4) Evaluate and propose data resource services (licensing, staffing sharing) to meet needs expressed in Phase 1.

Appendix B: 39 Potential Services Proposed by Working Groups

The following graphic lists all 39 services in shorthand, and the remainder of this appendix provides more descriptive information about each. The 12 prioritized services that move to Phase 3 appear in white in the graphic below and are noted with red priority references in the detailed list that follows.



<u>Group A:</u> Duke lacks sufficient personnel to support domain specific research.

A1. Personnel: Add 15-20 FTEs spanning Libraries, OR&I, OIT and Schools to enable and improve new categories of research support and provide more consistent offerings, including to units with lower funding levels. Details still TBD, but early concept is resources would be heavily weighted to various data-related needs (~45% focused on data analysis, visualization, consultation, curation, securing), plus technically-focused (~20% experts on technologies like python, creating pipelines, computation, ML, selecting on-prem vs cloud, etc.), domain-based (~15% domain based w/ expertise in natural/basic sciences, social sciences, digital humanities, etc.), process-based (~10% 'process sherpas' to help navigate services/policies/procedures) and dissemination (10% aimed at website + metadata creation to support sharing/distribution of research results).

Priority #1

A2. Personnel: Build cross-department virtual teams, to better support personnel across Schools as well as ORI, OIT, and Libraries to manage, develop and support the personnel. Add 1-3 additional FTEs for research program / project oversight – coordination of virtual teams, skill and resource needs analysis, support governance/meetings/etc., such as building cross-department virtual teams and specific job expectations for research support professionals and linking the expectations to job descriptions, performance management, and feedback.

Priority #8

A3. Personnel: Better connect and utilize new and existing personnel by: creating an internal database that catalogs support personnel skills (technical and substantive) for matching skills with problems/projects; creating a centralized request list and process wherein employees and units can express specific needs to allow coordinated training.

A4. Personnel: Create a program to support undergraduate and less-expert graduate students, such as tapping the expertise of more experience graduate students. Could relate/tie to current programs focusing on 12-month funding for PhD students; also give Masters students more support.

Group B: Separate research infrastructures hinder research and collaboration.

B1. Infrastructures: Move to a single, central protected network (rather than the two separate networks that exist now: the Protected Network, aka PRDN, offered by OIT and the PACE environment provided by DHTS) and establish/maintain support resources that will meet the needs of different types of research within the protected network environment.

B2. Infrastructures: Institute protected enclaves to encapsulate individual projects / data with the requisite security protections that enable authorized access / data movement based on the project circumstances. The protected enclave approach should be within the protected network and should enable different levels of security protections as may be called for (sensitive vs. regulated data) based on standard 'templates' (common enclave protections for different types of research compliance/security needs. Encapsulate the data protections at a project/data level rather than at the border of the protected network. This will offer separate security perimeters on a project-by-project basis, but also can enable an easy, but secure, avenue for moving data in and out, or across projects (if/when authorized).

Priority #9

B3. Infrastructures: Reposition the relevant/necessary security boundaries for research data to surround the particular research group environment, rather than basing it on the unit/department a researcher resides.

B4. Infrastructures: Facilitate Cloud AND On-Prem options, including SOM researcher access OIT's services for non-clinical research, where relevant. Offer / support similar or equivalent types of solutions to enable research needs both on-prem and in the cloud, acknowledging research needs of different projects require different solutions.

B5. Infrastructures: Ensure ease-of-use and consultation to choose research-relevant solutions and that current available resources and future additions are easy to access and easy to utilize. Provide consultation around the available solutions (e.g., in decisions between computation on-prem or in the cloud, understanding the impact of latency and transmission, as well as configuration control over hardware and software). There should be clear indications to base determinations of what a researcher should proceed with using, based upon the characteristics of the project at hand. (Requires consultative expertise characterized in Group A solutions.)

B6. Infrastructures: Host 'mini-workshops' to learn from other universities with medical centers. Consult with other schools with academic medical centers via (4) 1/2 day 'mini workshops' to learn about other potential approaches for: (1) Leveraging a common research network serving the entire institution (inclusive of AMC) and supporting sensitive data transfers where access has been authorized via not only policy but also technology (e.g., virtualization, discrete protected enclaves, etc); (2) establishing common research computation and/or storage infrastructures that can be plumbed/piped into discrete environments to serve need of researchers requiring different network configuration and policies for security or performance reasons; (3) taking an institutional approach (whether centrally or at the school level or otherwise) to de-identifying data originating from the clinical function associated w/ other university AMC's or health system's (i.e., turning PHI into RHI, research health information, that can be shared without—or with many fewer—restrictions); and (4) developing an

institution-wide approach or framework that maps security protections and compliance requirements to objective criteria that a researcher can mostly figure out for themselves (i.e., one which effectively balance the level of protection required to an objective discernment of risk on a project-by-project basis, and which researchers consider practical).

Group C: Current security / compliance approaches seem 'one size fits all'.

C1. Security/Compliance: Update Data Classification to (a) increase the number of categories from 3 to 4, adding a "highly sensitive" category for regulated data; (b) cross-walk with the following scenarios as a means of better guiding faculty through the process: Data coming into Duke (e.g., from a Data Provider); Data generated at Duke (non-human research); Data from interactions with human participants (e.g., through surveys, mobile apps, etc.); Data leaving Duke (e.g., through collaboration); Data leaving Duke (e.g., through formal data sharing/publication).

C2. Security/Compliance: Improve tools (MRH, SecureIT, etc.) to match research use-cases, scenarios, and data/risk classification with appropriate technology solutions and support resources, including a matrix of services that are aligned with different data classifications and risks.

C3. Security/Compliance: Enhance consultative services ('process sherpas') to aid faculty in navigate challenges and requirements around data security and compliance requirements and options (similar to research navigators in SOM or research project managers being introduced in certain other parts of campus).

C4. Security/Compliance: Improve recommendations, offerings and solutions for regulated research (e.g., via data providers' secure enclaves, through automated provisioning of OIT&DHTS secure enclaves, or through a process to engage a (to be rolled out) campus resource to assist in vetting approaches non-standard situations).

C5. Security/Compliance: Use a risk-based approach to establish security / compliance expectations at a project level, based on regulations, risk, and data classification; include guidance for requesting exceptions. Use the risk-based approach to reviewing research and setting requirements. Streamline and merge research study reviews (e.g. pre- or post- submission) based on risk level, regulatory requirements and classification of data, including through an early look at the grant application process to offer an early decision on risk levels/reviews (e.g., not needed b/c low risk; not needed b/c high-risk and will use approved solution; will need review b/c not seen before; will need review b/c desire to use unsanctioned service).

Priority #7

Group D: OIT's DCC Services are valuable but not as expansive as faculty require.

D1. OIT Services: Enhance processing and memory VM provisions in the Duke Compute Cluster that are available to all researchers and extend access to graduate (PhD) students and postdocs. Increase capability of DCC and Research Toolkits configurations that are available to all faculty by increasing the standard faculty allocation for VMs to 8 Cores and 72GB; provide a sliding scale between RAM and cores; make available minimal (1) GPU capability to researchers via Research Toolkits and DCC's university-funded VM environment; extend access (allocations) directly to graduate (PhD) students and postdocs.

Priority #3

D2. OIT Services: Better support AI/ML and other research through GPU capacity like DCC's on-demand CPUs access (shared and scavenged). Expand DCC offerings to provide GPU capacity to researchers at Duke with on-demand access (which may involve a short delay of a few minutes to allocate and build the GPU server) to improve support for AI/ML and other research.

Priority #6

D3. OIT Services: Provide secure DCC services functionally equivalent to OIT's existing virtual machine (VM) and other offerings. These secure computing environments should provide the same types of services as RAPID VMs and other DCC services and these environments need to be pre-blessed by the IT Security Office for sensitive and/or restricted research / data. Priority #11

D4. OIT Services: Provide a separately funded cluster for education, with a reasonable mix of CPUs and GPUs to meet explicit demands of courses; during periods of under-utilization for class or education efforts enable researcher access through scavenger nodes.

D5. OIT Services: Create/facilitate more seamless interoperability w/ cloud environments: simple method for researchers to test/try cloud environments with no investment (cloud credits?); services to allow researchers to scale into cloud environments and to enable experimental environments; safety rails to prevent overspend and meet security needs (logs/monitoring); provide tools to move data between cloud(s) and on-prem.

D6. OIT Services: Establish a special purpose VM environment for Graphical interface intense-edge work (virtual worlds and AR space, game development, animations, GIS); leverage GPUs if/as separately funded via other service proposals described herein.

D7. OIT Services: Improve web-based access to DCC resources. Support and facilitate web-based access to DCC resources (CPUs, GPUs, Storage), including portals for data analysis, tools for display data, etc.

D8. OIT Services: Support faculty startup packages / semi-autonomous sub-clusters, supporting direct and immediate access while also expanding the DCC and leveraging spare cycles. Make it easier to create the semi-autonomous sub-clusters for faculty startup packages or other uses, with resources seeded by deans/chairs/DST but which can be leverage other DCC resources if spare cycles are available (scavenger nodes). Priority #12

D9. OIT Services: Formalize/extend a special purpose FastMPI cluster (10 nodes, each w/ 2 GPUs) to form a "materials science simulation cluster" including 100G InfiniBand network switches and nodes purchased in a model similar to DCC. May have further uses beyond materials sciences, and may help with DST recruiting.

D10. OIT Services: Pursue federal agency grants (NSF, DOD) to create a mid-scale computational environment as a collaboration between OIT and multiple faculty PIs (e.g., NSF's mid-scale solicitation [https://duke.is/6/d27n] - "Examples of projects that may be supported by Mid-scale RI-1 include, but are not limited to, infrastructure that supports high-priority research experiments or campaigns, major cyberinfrastructure that addresses community and national-scale computational and data-intensive science and engineering research, major shared community infrastructure and resources as may be required to enable community-scale research and upgrades and/or major new infrastructure for existing facilities. Proposals for infrastructure that advances research on climate science and the impacts of climate change are encouraged.")

D11. OIT Services: Establish/better support heterogenous software environments including support for a variety of operating systems (Ubuntu, RHEL/Centos/ALMA/Windows?)

Group E: Plethora technical solutions and use constraints create confusion.

E1. Service Navigation: Develop a Self-service tool to guide service selection based on data classification, access attributes, etc. The tool should enable researchers to 'interrogate' the options available to them – like Cornell University's Research Navigator "FinderTool" and should be easy to use while making it transparent how different services relate to / comply with security or regulatory compliance requirements. Priority #10

E2. Service Navigation: Add two more Research Navigators; inventory existing concierge services e.g., Duke Libraries, OIT, business analysis, people interfaces, departmental niche, etc. and identify gaps. (Requires consultative expertise characterized in Group A solutions.)

E3. Service Navigation: Guide researchers on the range of cloud and on-premises computational and data service offerings and solutions, e.g. Duke-internal, external and different tiers that differentially meet needs, whether for educational use, collaborating within Duke or across multiple institutions. Provide guidance (human and/or automated based on existing license agreements and experience at Duke) to researchers obtaining and paying for cloud services.

E4. Service Navigation: Assess and understand the variety of public-facing web solution needs of researchers across disciplines; guide researchers on meeting their needs through well-defined Duke-internal offerings or externally.

E5. Service Navigation: Evaluate effectiveness of existing communication and marketing channels of IT service and solution offerings (including training) available and adjust as necessary to ensure accurate information is available in a timely, comprehensive manner to researchers.

E6. Service Navigation: Develop training programs for faculty and students (grad and undergrad) and ensure IT personnel are well trained on research support services. Provide and promote generalized and specific training opportunities and materials for researchers on IT skills through existing channels (Roots, Pathways, Coursera, etc.); general and specific laboratory management, specific service domain (research computing and storage), and ensure research navigators cross train across disciplines as well as DUL support staff (reference librarians)

Priority #5

E7. Service Navigation: Establish clearer governance over security / compliance choices; assist researchers in understanding the governance infrastructure and established standards for conducting research that is consistent and aligned with institutional policies (e.g., exportation of controlled data) across OIT and DHTS.

<u>Group F:</u> Current storage services don't span research lifecycle or university.

F1. Storage Services: Devise tools to manage data over its lifecycle, understand storage cost, and identify where data reside; provide storage capacity to meet 80% of active research project need. Tools to manage data over lifecycle should span from acquisition through analysis, publishing, archiving, and/or long-term storage, while making more transparent the costs for storage options (financial, especially upfront when writing grants; time, in terms of delay in retrieving; and environmental). Tools should also provide visibility into where data are stored and in what form. This would be a shared service/task with the Duke Library and OIT/DHTS and incorporating lessons from OIT's NSF award to develop tools to help manage data across the lifecycle and with deep researcher engagement to ensure solutions meet research needs.

Priority #2

F2. Storage Services: Support the purchase, hosting, and (internal and/or external) sharing of data sets using methods akin to how software is licensed (sometimes for particular groups, sometimes broadly for the institution); assess data-set demand that may extend beyond an individual research group; provide support for aggregated purchases or university-purchases as a way to facilitate broader access to data sets and to share costs.

F3. Storage Services: Provide storage flexibility to meet differing research needs (secure and public access) that are compliant w/ regulations for storage retention. Ensure storage services are discerning in ways that meet the various security and/or public access requirements for their data storage, including providing the storage needed to meet regulatory retention requirements. [Item is separate from, but relates to items under Security/Compliance topics]

Priority #4

F4. Storage Services: Establish guidelines that document what storage Duke provides to them - via general services, college, departmental, or other services; and what is the base level of service available to any researcher (without further payment) and for what time period

F5. Storage Services: Create consulting services to: clarify what options are available and at what are the ongoing costs; interpret what is required for federal grant submissions (NIH, NSF, ...) as well during and after a grant is awarded; and convey what is required by publishers (Elsevier, PLOS, …). (Requires consultative expertise characterized in Group A solutions.)

F6. Storage Services: Create Data Continuity Services that ensure data integrity and availability, including providing the storage associated with maintaining data continuity and availability via backup and/or replication of data, and which ensure that only needed data is being backed up (intermediate results may not need to be backed up).

Appendix C: Faculty Feedback from April 6 Poster Session

On April 6th, 2023, facilitators of Phase 2 of the Research IT Needs project held an in-person gathering, inviting all affiliated sponsors, faculty champions and consultees, team leads and project participants, to showcase the proposed solutions crafted by the six teams over the course of the 10-week period comprising Phase 2. This forum gave attendees a chance to review the work and proposals of each group and provide feedback and express support for specific initiatives. Details follow on the major takeaways of that forum, for each Working Group.

Major Takeaways/Feedback on Proposed Solutions from each Working Group:

Group A: "Duke lacks sufficient personnel to support domain specific research"

The group's primary solution of "Add 15-20 FTEs spanning Libraries, OR&I, OIT [and Schools] to enable and improve new categories of research support and provide offerings to units with lower funding levels. Mix of university-funded (allocations) and direct-to-grant available resources." received overwhelming levels of support and was **the most discussed proposition of all groups during the forum**. Based upon conversations between Team Leads and forum participants, this initiative was deemed highest priority.

In tandem with Group A's primary solution of the addition of specific FTEs, the proposed solution to "Build cross-department virtual teams and specific job expectations for research support professionals. Link those expectations to job descriptions, performance management and feedback." was also met with high levels of support from project sponsors and participants, and a wide number of faculty who attended and engaged.

Additionally, the following proposed solutions from Group A gained moderate levels of support from both participating faculty and project sponsors, from highest priority to lowest, with the first enumeratied item below being recognized as the means through which the "cross-department virtual team..." secondary priority described above would be achieved.

1. Add 1-3 additional FTEs for research program and project oversight - coordination of virtual teams, skill and resource needs analysis, support governance/meetings/etc…

2. Create a program whereby graduate students can receive funding for supporting undergraduate students and less-expert graduate students…

3. Create an internal database that catalogs support personnel skills (technical and substantive) for matching skills with problems/projects…

4. Create a centralized request list and process wherein employees and units can express specific needs to allow coordinated training…

Group A also posed three questions to forum attendees which received the following responses:

1. Where are researchers currently going to seek out specialized support?

>"Each other. Buying commercial tools. Hiring consultants. Giving up."

>"Need options for core facilities. What do we tell researchers when they are coming from many departments?"

>"Non-OIT staff with technical expertise. Students. Colleagues at other institutions."

>"Understanding hardware + options available + effect on research pipeline."

2. What's currently working well regarding seeking out and acquiring specialized support?

> No Responses Provided

3. How do you assess gaps in expertise and bandwidth from the resources currently available?

>"Ask faculty what they need. Compare with existing expertise and projects. Hire, train, repurpose, and share key staff.

>" Define parameters for transitioning to grant-funded FTEs with projects that achieve scale."

>"Ask of research staff available to collaborate. Ask of specialized support for domain specific software and desktop support."

Group B: "Separate research infrastructures hinder research and collaboration."

Group B's proposed solution of "Base the necessary security boundaries surrounding data around a specific project, rather than based upon the unit/department a researcher lives in" received very high levels of support across the board from project sponsors and affected faculty/researchers. This proposition, alongside the following, was deemed the highest priority of change for Group B. However, there was a write-in comment worthy of further consideration relative to this proposed solution "May cause confusion in single research group if there are different projects with different risk also confusion as projects change over time"

Alongside basing security boundaries around specific projects, Group B's proposed solution of "Institute a protected enclave that offers an easy, but secure, avenue for moving data in and out." was also met with high levels of support from the research community present and was deemed equal in priority.

Group B's remaining proposed solutions were met with moderate additional support, from highest priority to lowest…

1. Move to a singular, central protected network, rather than two separate instances, but maintain support resources that are specific to particular types of research within the same environment."

2. Support and offer solutions on campus as well as in the cloud, specifically when the cloud isn't the right solution. Understand that latency and transmission can be critical components, as well as control over hardware and software.

3. Consult with other schools with academic medical centers, bringing them to campus if/as needed to see what we might learn from the approaches of those who have already solve; unified common research network, approach to protected data enclaves, easier (authorized) data migration across enclaves, and resource sharing across different protect enclaves.

4. Ensure that current available resources and future additions are both easy to access and easy to utilize. There should be clear indications to base determinations of what a researcher should proceed with using, based upon the characteristics of the project at hand. (Requires consultative expertise characterized in Group A solutions.)

Group B also posed three questions to forum attendees which received the following responses:

1. Do you believe the proposed solutions will assist in reducing the current obstacles faced by the research community?

> "Separate enclaves is a great idea."

> "Degree of protection. Current OIT vs. Draconian DHTS."

2. What, if any, additional issues are not addressed by these proposed solutions and would remain a significant blocker to your research?

> "How do we evaluate security needs on a project by project basis? Eg. Basic Science Labs working on bacterial genetics shouldn't need to be treated like PHI."

3. Specific other schools with good models to probe?

> No Responses Provided

Group C: "Current security / compliance approaches seem "one size fits all""

Group C's proposed solutions were met with lukewarm reception, and little additional visible faculty support when compared to the other groups present, however the group's primary proposition of "Update the Data Classification to add a 4th"highly sensitive" category for regulated data; detail how each category relates to: Data coming into Duke (e.g., from a Data Provider), Data generated at, Duke (non-human research), Data from interactions with human participants (e.g., through surveys, mobile apps, etc.), Data leaving Duke (e.g., through collaboration), Data leaving Duke (e.g., through formal data sharing/publication)", was supported as the highest priority for implementation by forum attendees and project sponsors.

Group C's remaining proposed solutions garnered some additional support from the research community, from highest priority to lowest…

1. Improve tools (MRH, SecureIT, others) to match research use-cases, scenarios, and data/risk classification with appropriate technology solutions and support resources

2. Improve recommendations as well as offerings and/or solutions for regulated research

First preference - Utilize data providers' secure enclaves when available (e.g. UK BioBank, the Health and Aging Data enclave, MiCDA). Second preference - Automate provisioning of resources in OIT and DHTS-managed secure enclaves (on-prem and Azure) for easy provisioning by researcher or research support groups. Third preference - Process to engage CR-PSSC for assistance in vetting non-standard solution

3. Streamline and merge research study reviews (e.g. pre- or post- submission) based on risk level, regulatory requirements and classification of data. This could involve looking at the grant application process and ensuring that there's an early decision on risk levels/reviews (e.g., not needed b/c low risk, not needed b/c high-risk and will use approved solution, will need review b/c not seen before, or desire to use unsanctioned service). Formalize new role of the Executive Director, Research Data Strategy & Governance in determining, implementing and supporting strategies associated with research data, governance, security, and compliance; expand role of the CR-PSSC to develop and/or support reviews so that researchers are not bounced and referred to multiple offices.

4. Implementing changes and/or new solutions are dependent on new people resources to help as consultants/project sherpa's (navigator/RPM role) and technical support for research groups. [These are encompassed in Group A solutions.]

Group C also posed three questions to forum attendees which received the following responses:

1. Who is your first contact regarding data storage and security questions that concern your research projects?

> "Image data, repository options - required for publication. NIH and Publications require image data to be posted in a repository. What are Duke options?"

> "Seems to change every time. Many catch 22's. Shelley Epps? Corey Ennis?"

2. Describe the main roadblocks you run into when needing approval for technology solutions or access to data?

> "Solutions for maintaining legacy software/hardware."

3. Provide use-cases/workflows for how you work with research data or produce data.

> No Responses Provided

4. Do you plan on or see a need to pursue contracts/grants from the DoD/DARPA? NIH? CMS? State of North Carolina?

> No Responses Provided

5. What gaps are there in educational resources needed to complement technical solutions (e.g., to build awareness in and develop confidence in making correct decisions).

> "We lack consistency in risk classification and solutions!"

Group D: "OIT Services are valuable but not as expansive as faculty require"

Group D proposed services received high levels of input (interest at the poster session) from the research community (both faculty and project sponsors) throughout the course of the event, with three of the group's proposed solution receiving very high levels of support across the board and sharing top billing for priority of implementation. Those solutions being…

1. Establish special purpose VM environment for graphical interface intense-edge work

2. Raise bar on university-funded services of DCC and Research Toolkits [Ed: write in comments here clarified that what is really needed is to "lower" the bar, in order to make more powerful computational services available and easier to use]

3. Create/facilitate more seamless interoperability w/ cloud environments

Group D's remaining propositions also garnered moderately consistent additional support from participating attendees, from highest priority to lowest…

1. Support and facilitate web-based access to DCC resources (CPUs, GPUs, storage)

2. Provide a GPU cluster for all researchers at Duke, with access through several mechanisms and separately "seeded" from DCC

3. Provide secure computing environments that provide the same types of services as RAPID and the DCC

4. Plan to pursue NSF or similar funding to create mid-scale computational environment

5. Establish/provide greater support for heterogenous software environments

6. Formalize/extend special purpose FastMPI cluster for "materials science simulation cluster"

Group D also posed three questions to forum attendees which received the following responses:

1. Would the above proposed solutions meet your needs?

> "More consultants to help researchers configure and use dcc."

2. What else might your research need in terms of computation?

> "Sandbox for spinning up GPU instance for mc."

3. Are you willing to participate in grant development to develop new, special purpose clusters?

> No Responses Provided

Group E: "Plethora technical solutions and use constraints create confusion"

Group E's primary proposition of "Establish a Self-service Option – like the Research Navigator Cornell University "Finder Tool" was met with high levels of additional support from attending research community members, alongside the project's sponsors. This was a clear frontrunner for priority out of Group E.

Coinciding with their foremost proposition, Group E's proposition to "Make Personnel Available – Research IT Staffing (assess and understand intricacies of needs and timing) [esp. DUL]" was equally lauded by participants and project sponsors and deemed of equal priority for implementation.

Additionally, the following proposed solutions from Group E gained moderate levels of support from both participating faculty and project sponsors, from highest priority to lowest…

1. Understand and meet the Public-facing web solution needs (Duke v. External)

2. Ensure Communication and Marketing of IT offerings

3. Data and Compute – Cloud and On-Premises solutions

4. Establish and follow a governance infrastructure, consistent w/ institutional policies (e.g. export controlled data

5. Training Assessment

Group E also posed three questions to forum attendees which received the following responses:

1. What training would you like to see offered?

> "Cross training to include Libraries staff (Ref Desk, Research Consult, Instruction). Specific research navigator knowledge."

2. What are the biggest pain points in Duke's data storage environments?

> "Infrastructure for maintaining data through life cycle. Hot > Cold migration, standardization, metadata, public access."

> "Public facing Hot storage + project publication venues."

> "How can we share openly with collaborators?"

3. How do IT staff learn what is available centrally versus in their own department?

> "Faculty and trainees."

4. How likely are you to use a self-service tool and/or a person and under what circumstances?

> "When speaking with libraries staff (ref desk, research, consult, instruction) and would like them to assist/direct.

Group F: "Current storage services don't span research lifecycle or university"

Alongside Group A's proposition for the introduction of additional specialized FTEs, Group F's proposal to "Offer and support easily accessible tools for a) managing data over its life cycle, especially for supporting the purchase, hosting, and sharing of data sets (internally and externally); b) more easily understanding the costs of varying storage options (both on premises and cloud-based), including performance delays and environmental impacts, to help inform researchers' storage decisions; and c) identifying where data lives at any given time," was the standout solution provided by the groups at the showcase. Group F's primary provided solution was met with significant additional from nearly half of all forum participants and the project's sponsors. Again, like Group A's FTE proposal, this proposition was lauded and placed with the highest priority.

Group F's proposition of "Offer services that help researchers [and students] understand what storage options are available [at what costs], the requirements surrounding a federal grant submission, and what is required by publishers," was also very well received by participating members of the research community and project sponsors. Following the previous proposition, this was met with second highest priority for Group F.

Additionally, the following proposed solutions from Group F gained significant levels of support from both participating faculty and sponsors, from highest priority to lowest…

1. Data Support Policies: Update policies that document what Duke provides to researchers, the base level of free/entitlement services and their duration, expectations for data retention and publishing, and roles and responsibilities for ownership and stewardship of data.

2. Storage Service: Provide storage services that meet the needs outlined above to researchers at Duke – both on premises and cloud-based storage.

3. Preservation & Access Services: Offer data repository services that meet requirements for FAIR data and code and for public data sets, provide data storage in support of group/team/project websites for data sharing, provide data storage for sensitive data sets, and provide public access metadata discovery services to meet fed. funding agency requirements.

4. Preservation & Access Services: Offer data repository services that meet requirements for FAIR data and code and for public data sets, provide data storage in support of group/team/project websites for data sharing, provide data storage for sensitive data sets, and provide public access metadata discovery services to meet fed. funding agency requirements.

5. Data Continuity Services: Offer services that ensure data integrity and availability via backup and/or replication of data, and that ensure only needed data is being backed up (intermediate results may not need to be backed up).

Group F also posed three questions to forum attendees which received the following responses:

1. Do you believe the proposed solutions will assist in reducing the current obstacles faced by the research community?

> "Will help them better understand the better and more efficient ways to store."

2. What, if any, additional issues are not addressed by these proposed solutions and would remain a significant blocker to your research?

> "Be careful…there is no "one size fits all" data storage solution."

> "Standardize risk classifications for data across Duke + consistent w/ peer institutions."

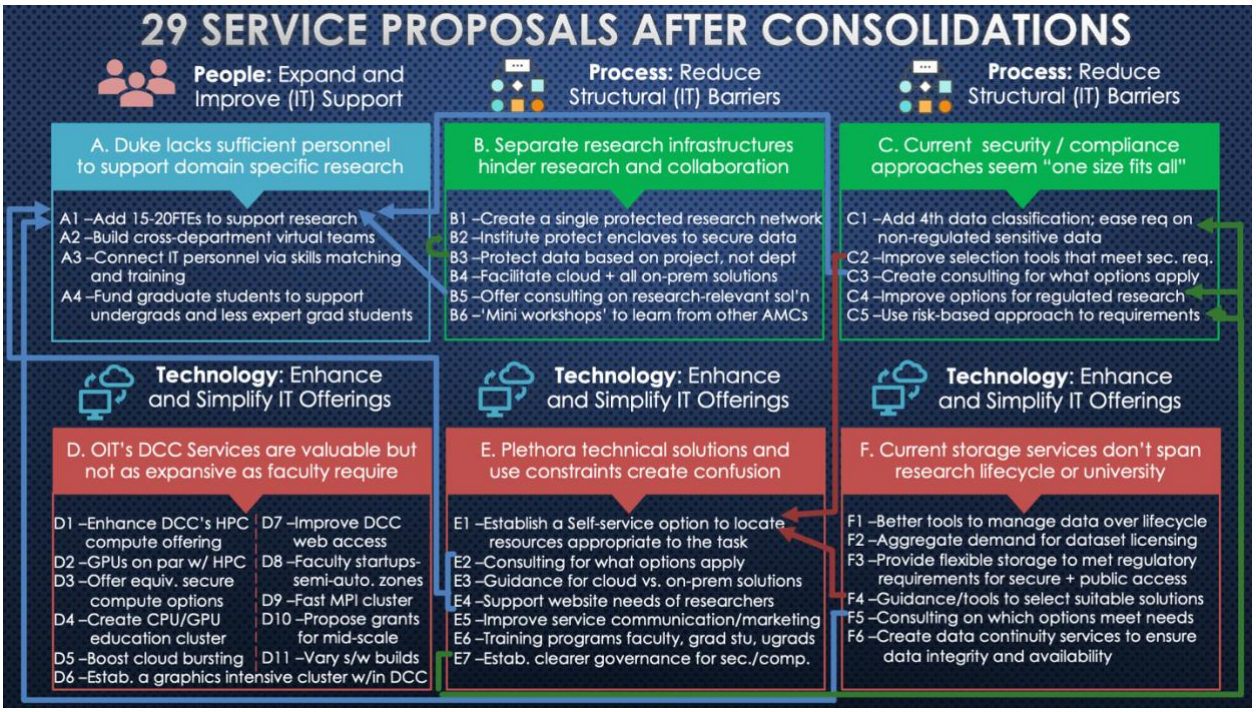
> "Help solve the problem of "time." How can I store data and metadata fast and easy without involving lots of extra time."

Appendix D: Service Proposal Consolidation and Triage

From the outset, project sponsors and team leads recognized that the nature of the Phase 1 findings would lead to overlap among the six Phase 2 working groups. For example, perceived complexity in navigating two research infrastructures, each with different services and constraints on its use, can manifest as a concern that too many technical solutions exist and cause confusion, or as a concern that there aren't enough personnel to help navigate which computational service applies to a particular need, or even a concern that faculty in one domain should have access to services in the other.

This cross-group inter-dependency was acknowledged with working group participants when Phase 2 was launched, and with the understanding that the individual groups should proceed with their recommendations and subsequent consolidations would be made, as the need arose.

Based on poster session feedback, input from sponsors, and discussion among working group leads, nine of the initial 39 initiatives proposed by workgroups were consolidated into services/projects proposed by other groups. Those consolidations are depicted in the following graphic:

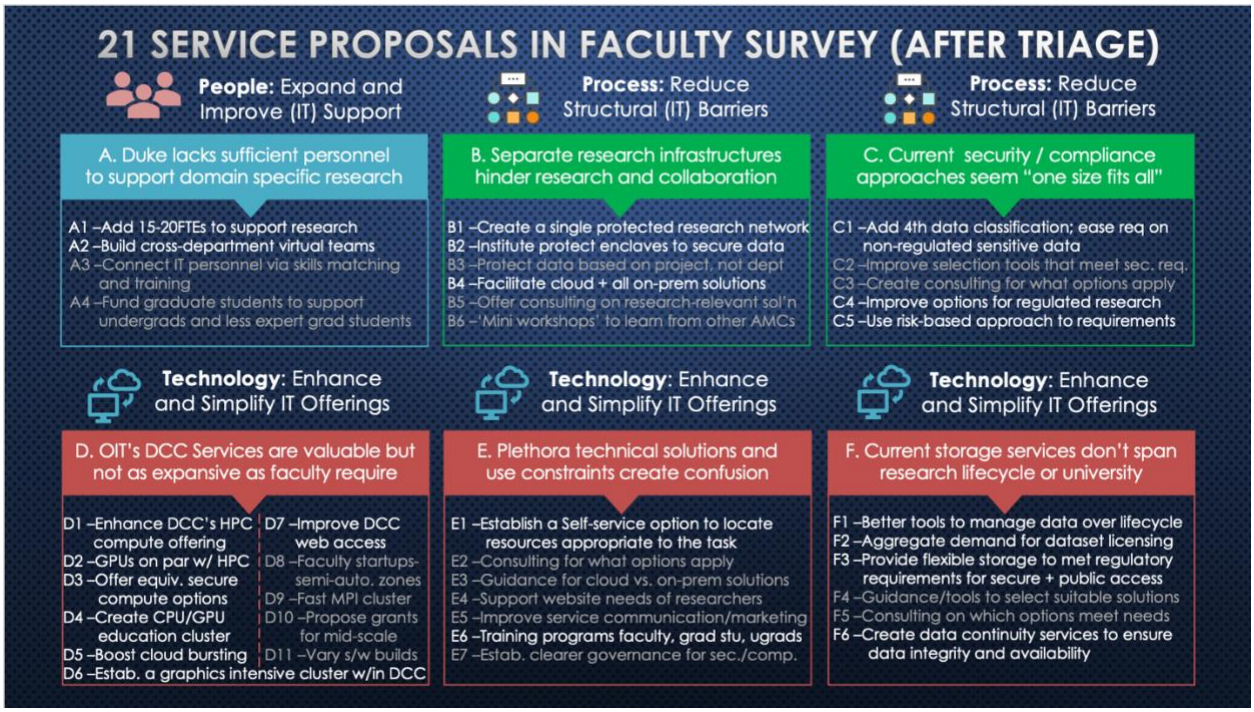


Most, notably, six service proposals were consolidated into A1 (B5, C3, E2, E3, E4, F5). Two other service proposals (C2 and F4) were incorporated into E1, and one service proposal (B3) was incorporated into B2. Finally, one service proposal (E7) was consolidated among three other service proposals (C1, C4, C5).

Following the consolidations described above, the 29 services/projects remaining were scored based on faculty voting at the April 6 poster session. Eight priorities had particularly low votes and so were subsequently **excluded** from the faculty prioritization survey distributed on May 15. Here is the list of excluded service, with a notion for each that specifies the number of votes it received at the poster session:

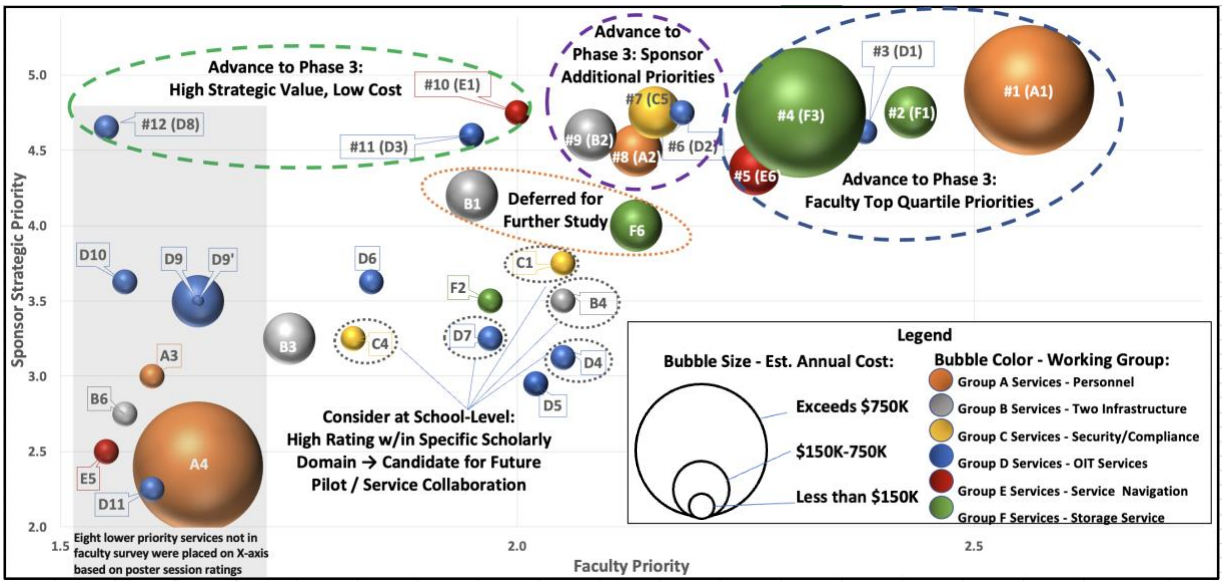
A3- Personnel: Connect IT personnel thru skill matching and training (6 votes)
A4 - Personnel: Fund graduate students to support undergraduates and less expert grad students (4 votes)
B6 - Infrastructures: Host 'mini-workshops' to learn from others (6 votes)
D8 - OIT Services: Support faculty startup packages or other semi-autonomous sub-clusters (0 votes)
D9 - OIT Services: Formalize/extend special purpose FastMPI cluster (2 votes)
D10-OIT Services: Propose federal grants to create a mid-scale computational environment (4 votes)
D11-OIT Services: Improve support for heterogeneous software environments (3 votes)
E5 - Service Navigation: Evaluate and Improve Communication and Marketing of IT services and solutions (5 votes)

These two rounds of consolidations and triage resulted in 21 services being advanced to faculty for prioritization, shown in white font below.



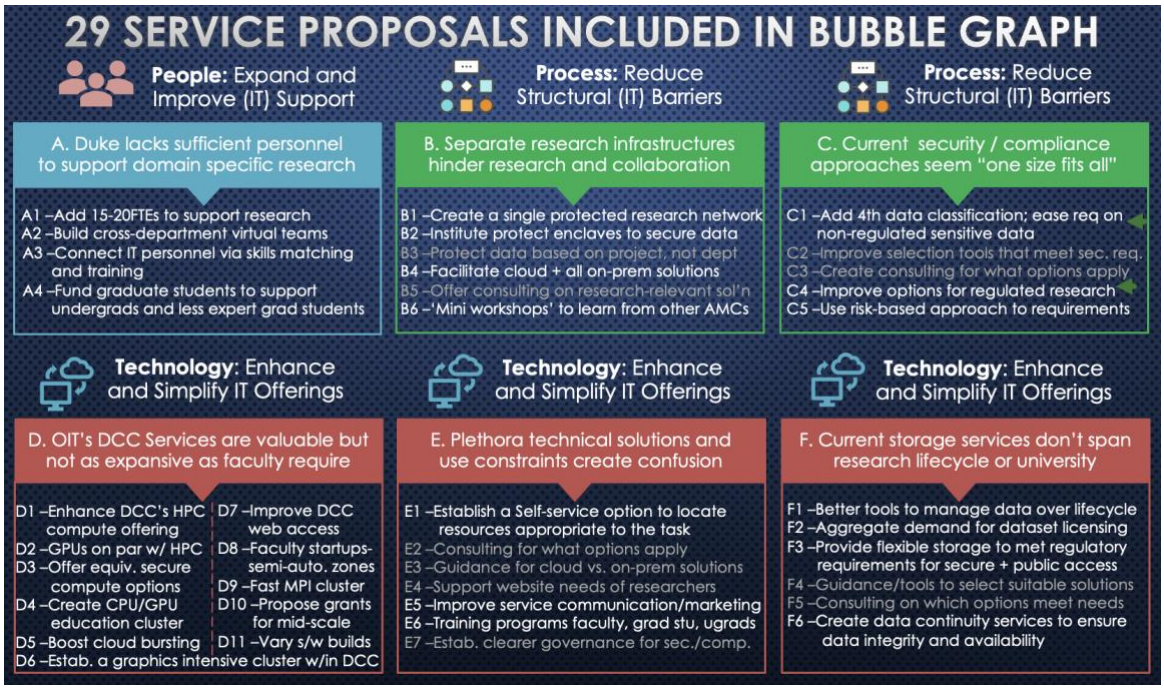
Appendix E: Proposed Services Based on Faculty / Sponsor Priorities and *Estimated* Cost

The 21 highly rated services from the poster session formed the basis of the survey sent on May 15 to all 58 faculty participants from Phases 1 and 2. After two[5] reminder emails to non-respondents, the survey was closed on June 21, with an overall 67% response rate and response rates of 56%-75% from each domain surveyed[6]. Faculty feedback[7], sponsor input and cost estimates were graphed:



Graph 1: Bubble labels #1-12 rank the highest priority services, described on page 3. Other bubble labels refer to lower-priority services proposed by specific work groups (A-F), but not prioritized through the overall process. Appendix B lists services in detail.

Services plotted above[7] are listed below with their short-description label in white for the services included in the survey, and grayed out if excluded (i.e., consolidated into others; see Appendix D).



[5] One domain area for which the initial response rate was below 50% received a third reminder.

[6] Natural Sciences/Environment, Engineering, Social Sciences/Policy, Basic Sciences/Nursing, Humanities/Arts

[7] Lower-rated services excluded from the Faculty Survey were placed along the X-axis based on relative ratings from Poster Session

Appendix F: Faculty Survey Results and Write-in Comments Regarding Proposals[8]

	Count	Mean	StdDev	Var	Priority Rating and Count Low	Medium	High	Sum
A1 - Personnel: Add 15-20 FTEs to improve research support (incorporates need for consultation conveyed by all groups)	39	2.51	0.64	0.40	7.69% (3)	33.33% (13)	58.97% (23)	98
A2 - Personnel: Build cross-department virtual teams to better connect (existing & new) research support personnel in ORI, OIT, Librarires, Schools	39	2.08	0.76	0.58	25.64% (10)	41.03% (16)	33.33% (13)	81
B1 - Infrastructures: Move to a single, central protected research network rather than the separate ones provided by OIT and DHTS	39	1.90	0.78	0.60	35.90% (14)	38.46% (15)	25.64% (10)	74
B2 - Infrastructures: Institute protected enclaves to encapsulate individual projects/data with the necessary security protections	39	2.03	0.77	0.59	28.21% (11)	41.03% (16)	30.77% (12)	79
B4 - Infrastructures: Facilitate Cloud AND On-Prem options, including SOM researcher access OIT's services for non-clinical research where relevant	39	2.00	0.78	0.62	30.77% (12)	38.46% (15)	30.77% (12)	78
C1 - Security/Compliance: Add 4th data classification; where feasible, ease requirements on non-regulated sensitive data	39	2.00	0.85	0.72	35.90% (14)	28.21% (11)	35.90% (14)	78
C4 - Security/Compliance: Improve and clarify storage and computational options approved for regulated research	39	1.77	0.70	0.49	38.46% (15)	46.15% (18)	15.38% (6)	69
C5 - Security/Compliance: Use risk-based approach to reviewing research and setting requirements	39	2.10	0.71	0.50	20.51% (8)	48.72% (19)	30.77% (12)	82
D1 - OIT Services: Enhance the Duke Compute Cluster (DCC) standard VM provisions with additional capacity (processing and memory)	39	2.33	0.73	0.53	15.38% (6)	35.90% (14)	48.72% (19)	91
D2 - OIT Services: Provide GPU capacity in similar ways to the DCC's CPU computation	39	2.13	0.72	0.52	20.51% (8)	46.15% (18)	33.33% (13)	83
D3 - OIT Services: Provide secure compute options across OIT services that are functionally equivalent to OIT's non-secure options	39	1.90	0.67	0.45	28.21% (11)	53.85% (21)	17.95% (7)	74
D4 - OIT Services: Create an education cluster with CPU and GPU virtual machines to support course needs	39	2.00	0.75	0.56	28.21% (11)	43.59% (17)	28.21% (11)	78
D5 - OIT Services: Improve services/interoperability with cloud to facilitate bursting for large computations , ML training, etc.	39	1.97	0.73	0.54	28.21% (11)	53.85% (21)	17.95% (7)	74
D6 - OIT Services: Create/optimize a special-purpose VM environ. for graphical intensive work	39	1.76	0.72	0.52	38.46% (15)	43.59% (17)	17.95% (7)	70
D7 - OIT Services: Improve web-based access to DCC resources	39	1.92	0.66	0.43	25.64% (10)	56.41% (22)	17.95% (7)	75
E1 - Service Navigation: Establish a Self-service Option like Cornell's "Finder tool" to guide service selection based on data classification, access characteristics, etc.	39	1.95	0.64	0.41	23.08% (9)	58.97% (23)	17.95% (7)	76
E6 - Service Navigation: Develop training programs for faculty as well as for students (graduate and also undergraduate); ensure IT personnel are also well trained on research support services.	39	2.21	0.72	0.52	17.95% (7)	43.59% (17)	38.46% (15)	86
F1 - Storage Services: Provide better data management tools to manage data and storage over the data lifecycle, and provide the associated storage for active projects	39	2.38	0.62	0.39	7.69% (3)	46.15% (18)	46.15% (18)	93
F2 - Storage Services: Support the purchase, hosting, and internal / external sharing of data sets via methods akin to how software is licensed (sometimes for particular groups, sometimes broadly for the institution)	39	1.92	0.69	0.48	28.21% (11)	51.28% (20)	20.51% (8)	75
F3 - Storage Services: Provide more flexiblity in storage to meet differing research needs for both secure access and public access, including any required storage retention to meet regulatory requirements	39	2.26	0.63	0.40	10.26% (4)	53.85% (21)	35.90% (14)	88
F6 - Storage Services: Create Data Continuity Services that ensure data integrity and availability, including providing the storage associated with maintaining data continuity	39	2.08	0.66	0.43	17.95% (7)	56.41% (22)	25.64% (10)	81

[8] Two additional surveys were submitted after the end of the survey period; their ratings did not materially affect the services advanced to Phase 3 and so are excluded from the above analysis, but their write-in comments are included below.

Write in Comments Regarding Service Priorities and Process

storage access is always tricky -- make it easy via web interface for the majority and/or provide APIs for more advanced users. This was a challenge for dhts as they provided web based wrappers for object storage, but didn't provide low level api access (or a very limited subset of api calls). If you have millions of files -- a web interface is just not sufficient for real work.
enhance personalized, and direct IT support to PIs students, beyond standardized options
Through this long process, the solutions are so refined that they ALL sound nearly equally compelling. I also feel torn between answering questions based on my own research needs and answering in a custodial capacity on behalf of others with legitimate and pressing needs that I don't share. Nice job, all concerned!
Thanks for collecting this feedback!
Thank you for such an in-depth review of the needs!
Storage- I have a hard time dissociating these options since managing data through the "life-cycle" (a) requires public access (c) and data continuity (d). Personnel- I feel strongly that adding 15-20 FTEs across campus (a) will not be enough unless there is synergy from cross-departmental teams (b) which can access a broader knowledge base. Infrastructure and Security/Compliance- I am in favor of any options that move us further from "one size fits all". OIT services- I don't have strong opinions on what is needed here other than improving access since I currently have not had success using these resources.
My biggest need at Duke would be a simple, large compute cluster with fast interconnect, without VM or any intervening infrastructure, facilitating compute intensive simulations locally.
More CPUs and GPUs in cluster :)
It can't be stated enough that we need research support staff with both IT and academic domain expertise, not only to support existing research (or research that could exist but does not currently have the needed support) but also to support and promote any/all of the needed services listed in the survey. I've shared the example of Stanford's Academic Technology Specialists, and I strongly encourage OIT to explore this model.
It appears that many of these options relate to more sophisticated (or perhaps complex) data uses than my graphical, humanist research requires. That being said, I indicated the priorities for my research.
I don't use the secure computing infrastructure which seems to be costly. I wonder if there are ways for researchers who use that to support those costs on their grants by adding service payments.
From the Social Sciences. I get that answers may differ depending on discipline. Thanks for all the work on this.
-licensing data like software is a very bad idea. do not do this. -please don't make us sit through yet another training program. Design the software such that it is intuitive and obvious to use. Training sessions are just markers for bad design (ever wonder why you don't need a training session to use an iPhone?) -we do not need secure enclave consolidation -- that will put campus side under SOM, which will dominate and is not necessary.
Assist core facilities to inform researchers/users about data storage and transfer options and temporary storage locations for analysis. I keep finding that students or lab members in SoM basic science labs or basic research labs in A&S know nothing about backing up data and plans for how to transfer and store their data. This is on their PI, but better informing and training needs to happen. Just letting you know that the information about data storage and transfer options are not getting to individual labs/PIs (and I need and update about options, too). Thanks
I still think we need to support on-campus Graphics PC labs, but I did not see this on the list. It seems we are still years away from providing a reliable docking station & VM approach to graphics computing, and we absolutely need to teach students the procedures, and physical postures of desktop graphics computing - which are very different from their modes of working with their laptops. To free up the desktops that we do have, GPU rendering in the cloud / render farms are also very important. Without them, students have been, and will continue to tie up our desktop machines for rendering, or photogrammetry processing etc.

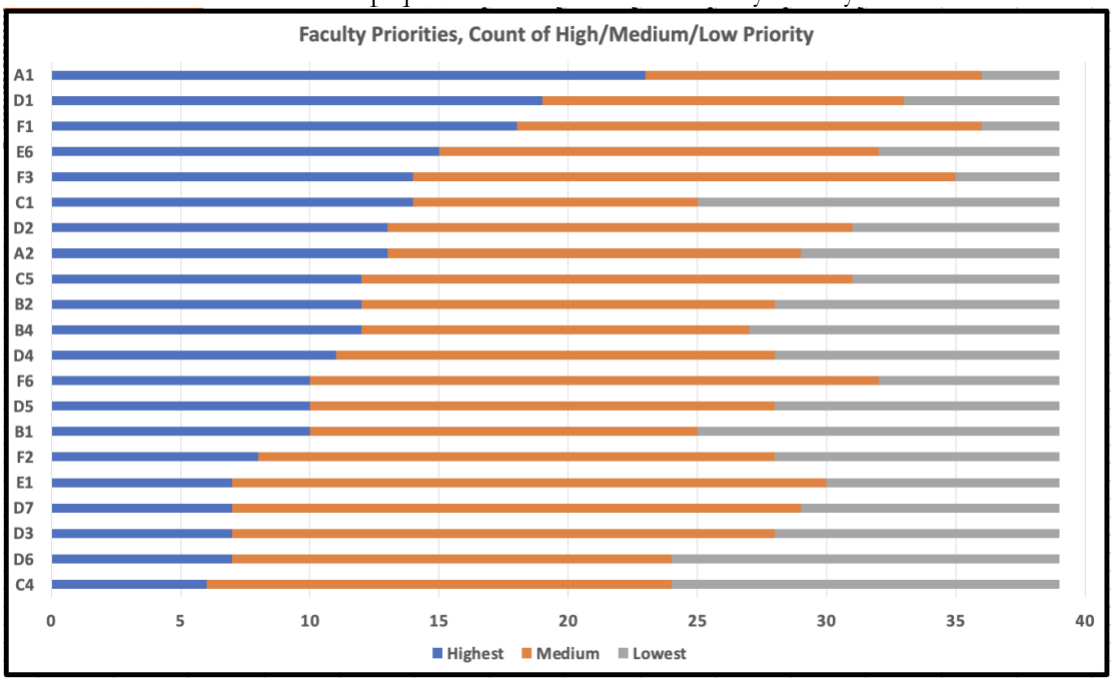
Appendix G: Discussion of Survey Responses and Variations by Individuals and Domains

The survey presented a three-point rating scale for each service, with a low priority rating for an individual service corresponding in the analysis to a score of 1, a medium priority as a 2, and highest priority corresponding to a 3. This meant that with 39 respondents, the maximum possible score for a given service was 117.

All faculty were urged to spread their scores and did so. For the 39 respondents, the average rating given was a 2.06, with the standard deviation of 0.25. Only seven respondents (18%) had average scoring beyond one standard deviation (max 2.43, min 1.19).

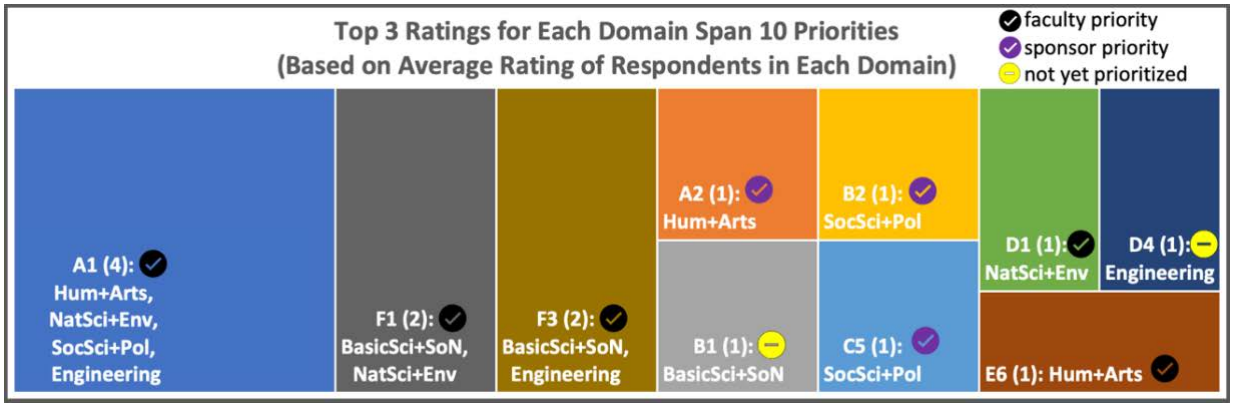
Appendix E shows all 21 services rated by all 39 faculty on the three-point scale, with the total (Sum) in the far-right column reflecting the total score for that service. The service receiving the highest sum of all rankings scored a 98 (service A1) out of the maximum possible 117, while the lowest total score (Sum) for a service was 69 (service C4). Interestingly, even for service C4, the service with the lowest number of total votes, 15% of faculty rated it as among their highest priorities, especially in Engineering and Social Sciences/Policy (see below regarding high service ratings at the School level).

As this next chart shows, each of the 21 services in the survey was ranked highest by some but lowest by others. The chart shows services in descending order of 'highest priority' ratings. Note that the first five services from this chart corresponded to the five highest mean scores, which was the metric used to determine the top quartile of services as rated by faculty.

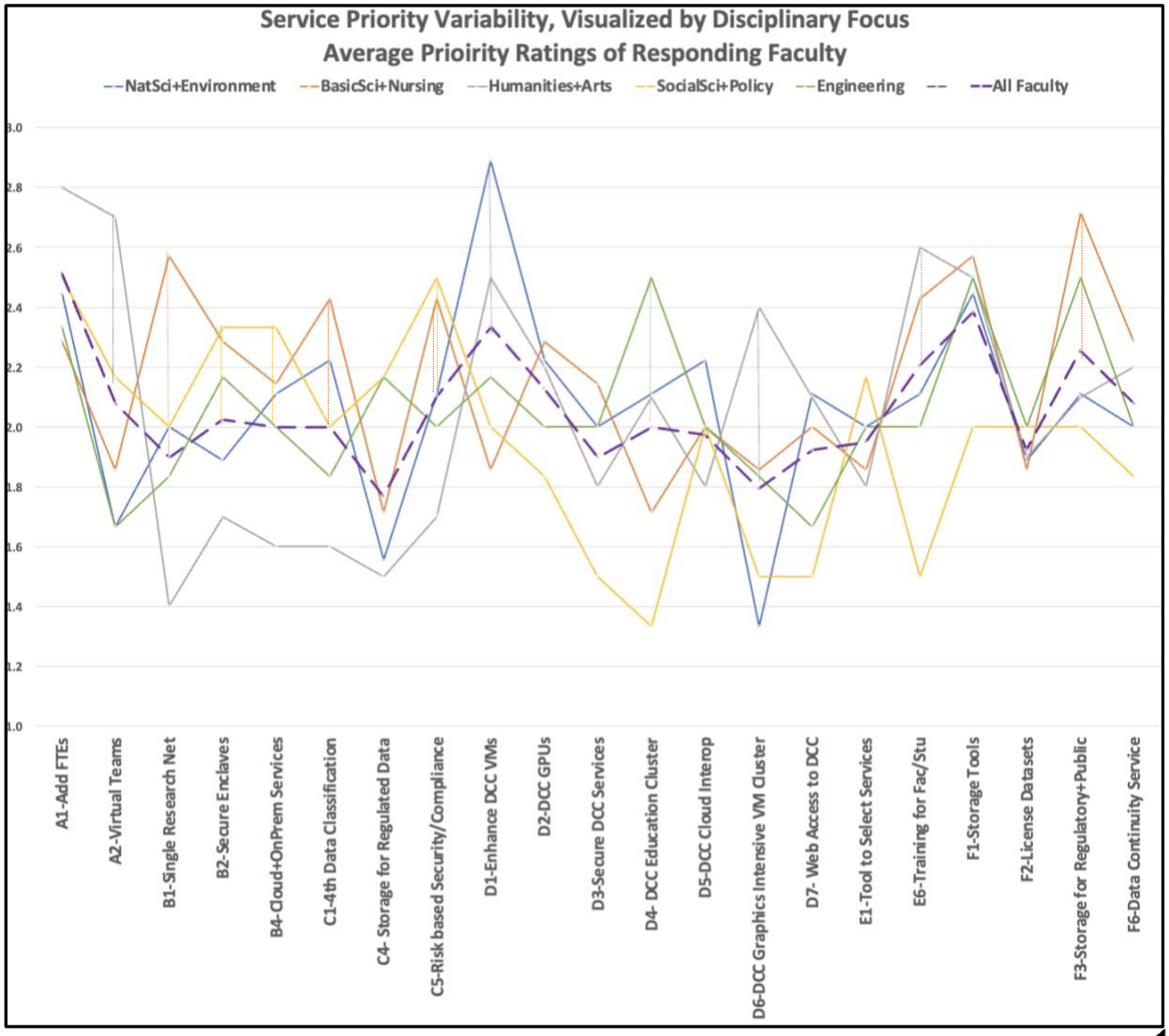


Beyond the considerable variation in faculty responses at an individual level, there was variability based on aggregations of respondents by academic domain. Specifically, the survey analysis identified the top three priorities from each scholarly domain cluster (Humanities/Arts, Natural Sciences, Basic Sciences/Nursing, Engineering, Social Sciences/Policy). Those top three priorities for the five scholarly clusters reflected ten service priorities. Eight of those, shown below with black or purple checks, were already prioritized in the twelve services that advance to Phase 3. The two services that were not prioritized by the global process (marked in the graphic below with a ⊖) are of

keen interest to Basic Sciences/Nursing (*B1 - Create a single, central protected research*) and Engineering (*D4 - Create an education cluster with CPU and GPU virtual machines to support course needs*).



For more than half of the 21 services in the faculty survey, particular domain faculty rated a service considerably (>0.5) or moderately (>0.35) higher than the global faculty rating, as called out below with dotted vertical lines extending above the purple dashed all faculty.



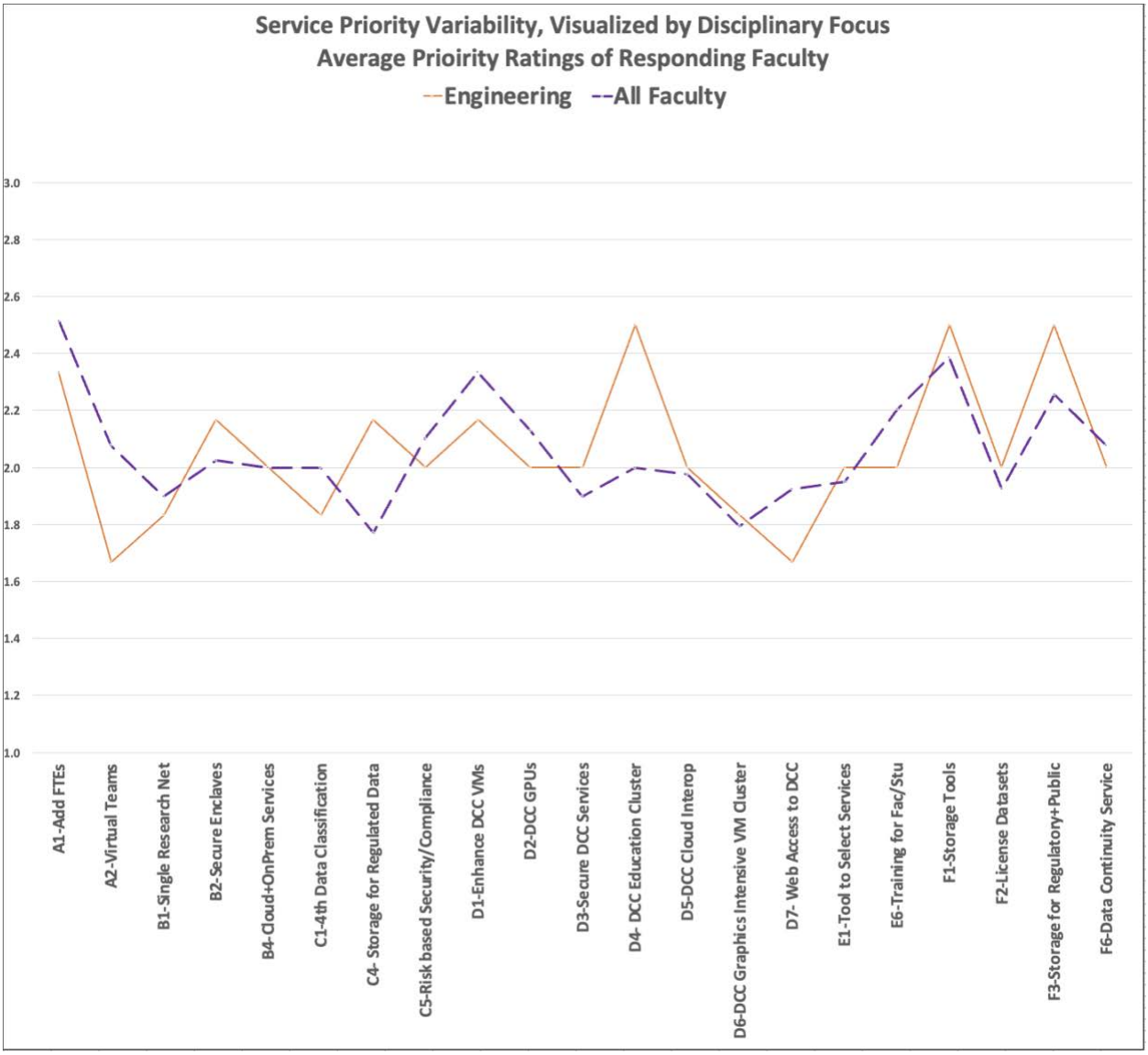
Many cases of higher domain-based rating underscored interest in a service proposal already prioritized by overall faculty ratings (A2, D1, E6, F3) or sponsors (B2, C5). In other cases, it reflected a localized perspective on a proposed service as a high priority, which that was not conveyed (prioritized) based on overall faculty ratings or sponsor ratings. Domain-by-domain analysis follows.

Natural Sciences / Nicholas

- Faculty from Natural Sciences and Nicholas rated an already-prioritized service considerably higher than the overall faculty averages *D1 - Enhance the Duke Compute Cluster (DCC) standard VP provisions)* (scoring 2.89 for Natural Sciences/Nicholas vs 2.33 for faculty overall).
- Three services rated slightly higher than for overall faculty: *C1 - Update Data Classification to add 4th regulated data classification* (2.22 vs 2.0), *D6 - Establish special purpose VMs for Graphical interface intense-edge work* (2.22 vs 1.97), and *D8 - Support faculty startup packages* (2.11 vs 1.92)
- Two other services rated lower for Natural Sciences/Nicholas faculty relative to all faculty: A2 (create/support virtual teams, 1.67 vs 2.08) and D7 (enable web access to the DCC, 1.33 vs 1.79)

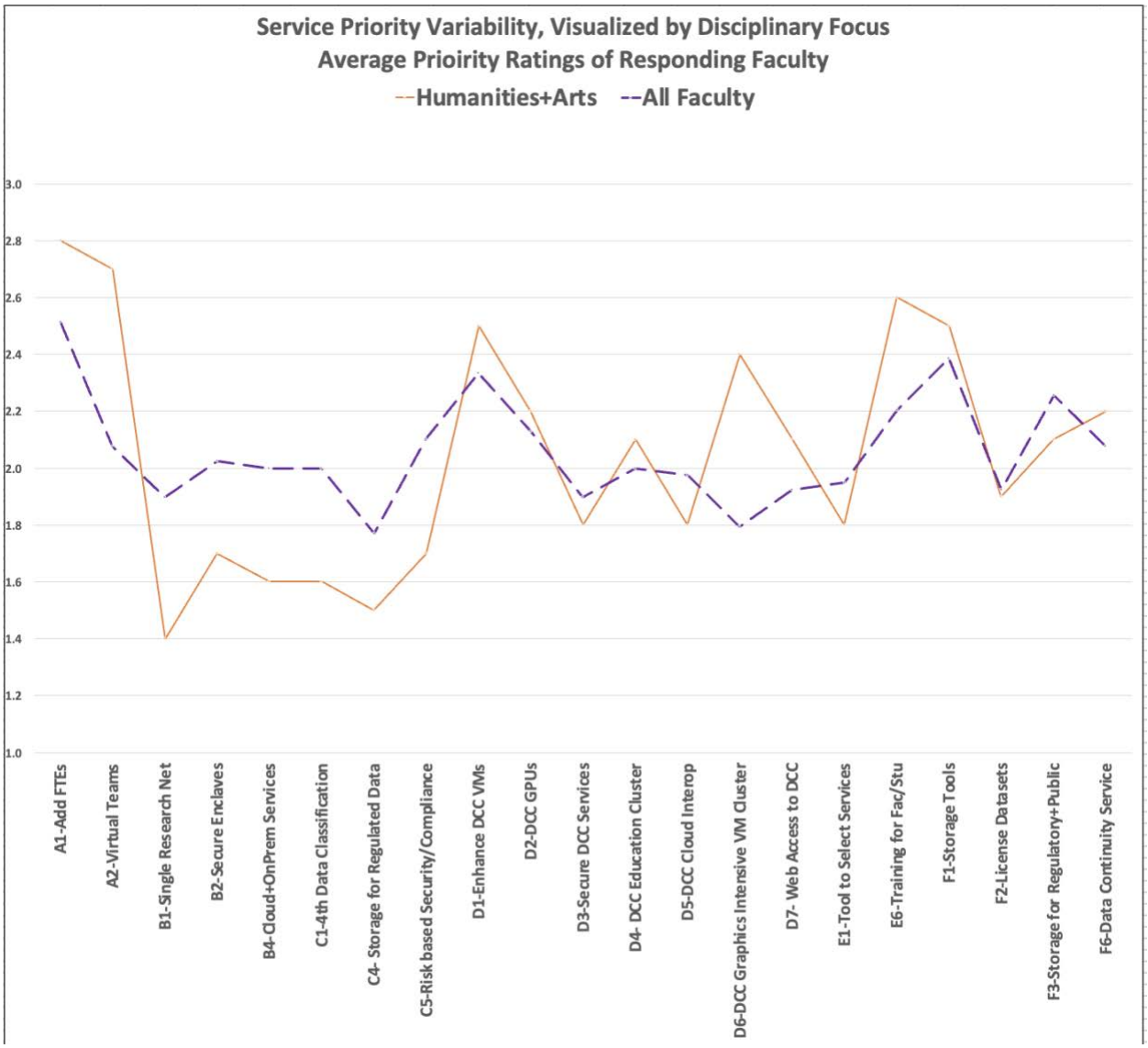
Engineering

- Engineering faculty conveyed a strong preference for *D4 - Create an education cluster with CPU and GPU virtual machines to support course needs* (2.17 for Engineering vs 1.77 for faculty overall). They also more highly rated *C4 - Improve and clarify storage and computational options approved for regulated research,* relative to the overall faculty average (2.17 vs 1.77).
- They expressed a preference relative to faculty overall for an already prioritize service, *F3 - Provide storage flexibility to meet differing research needs (secure and public access, compliant w/ regulations for storage retention* (2.5 vs. 2.26) and a slight preference for service B2 (protected enclaves, 2.17 vs 2.03), which fell into the 'on-the-bubble' category requiring further evaluation.
- Engineering faculty rated service A2 notably lower (creating virtual teams to support research, 1.67 vs. 2.08) and D8 somewhat lower (supporting faculty startup, 1.67 vs 1.92).



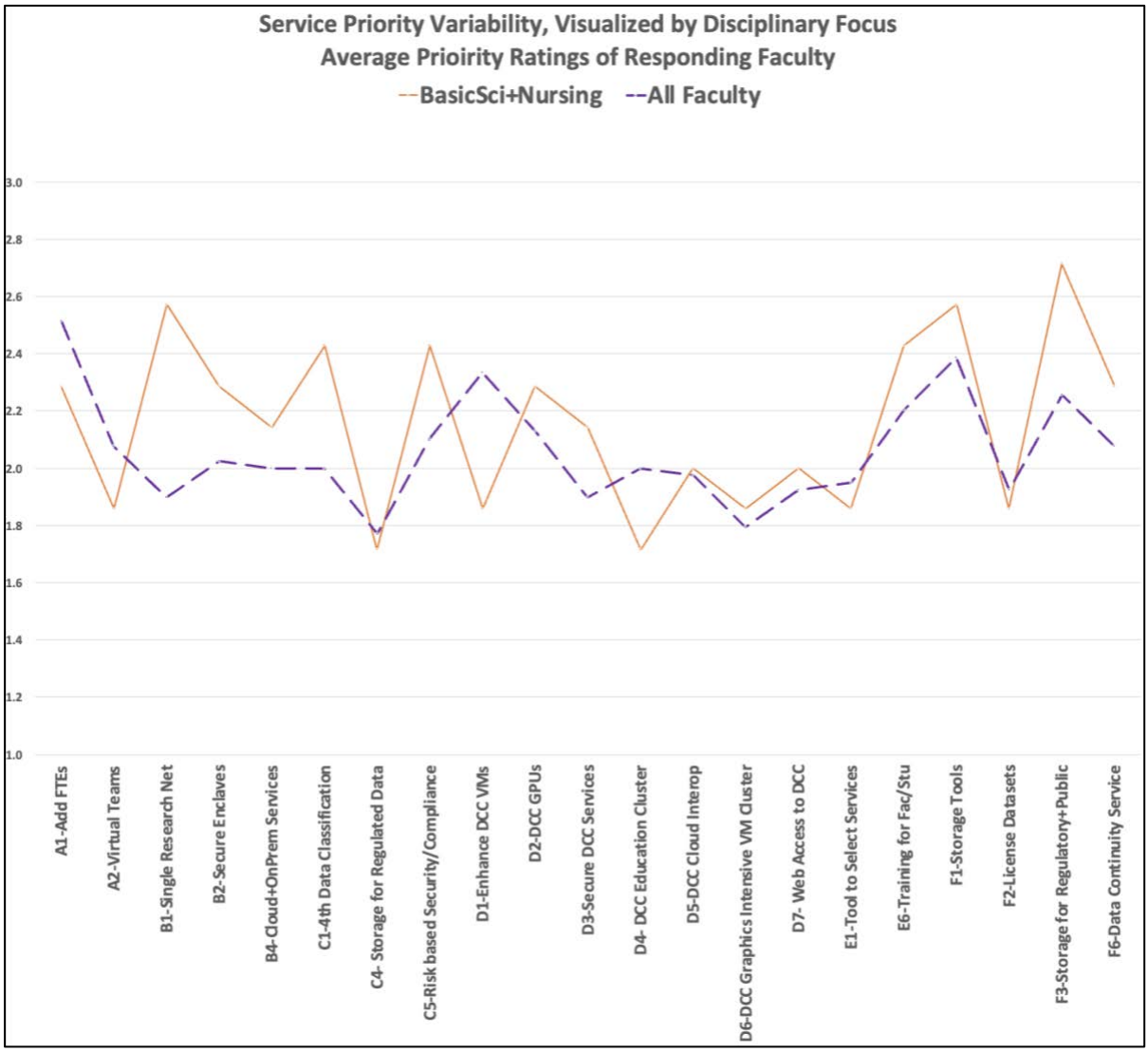
Humanities/Arts

- Humanities/Arts faculty expressed much stronger interest than faculty overall in service *D7: Improve web-based access to DCC resources* (2.4 vs 1.79). They also scored already-prioritized services *A2 - Build cross-department virtual teams to better connect (existing & new) research support personnel* and *E6 - Develop training programs for faculty as well as for students (graduate and undergraduate); ensure IT personnel are also well trained on research support services* higher than the norm (2.7 vs 2.08 and 2.6 vs 2.21, respectively).
- They consistently rated lower every one of the services that emerged from Working Groups tasked with Findings B (Separate research infrastructures hinder research and collaboration) and C (Current security and compliance approaches seem 'one size fits all'). This is unsurprising given relatively fewer instances of sensitive research, and / or collaboration with SOM/SON.



Basic Sciences/Nursing

- Basic Sciences/Nursing faculty ratings had considerably more variability relative to faculty overall and they were much more likely to prioritize *B1- Create a single, central protected research network rather than the separate ones provided by OIT and DHTS* (2.57 vs. 1.9). They also gave moderately higher scoring to *C1 - Add 4th data classification; where feasible, ease requirements on non-regulated sensitive data* (rated at 2.43 compared to 2.0 for faculty overall).
- They also highly rated two already prioritized services, *F3 - Provide flexibility in storage for secure access and public access*, (2.71 vs 2.26) and *C5 - Use risk-based approach to reviewing research and setting requirements* (2.43 compared to 2.1) and gave slightly higher ratings to three other already-prioritize service, D3 (provide secure DCC services, 2.14 vs 1.9), E6 (training for faculty / students, 2.43 vs 2.21) and F1 (tools to manage data over its lifecycle, 2.57 vs 2.38).
- Two other services were slightly higher rated than for faculty overall, including B2 (institute secure enclaves, 2.28 vs 2.03), and F6 (data continuity services, 2.29 vs. 2.08).
- Finally, two services ranked considerably or moderately lower than for faculty overall, D1 (enhance the DCC, 1.86 vs 2.33) and D4 (education cluster, 1.71 vs 2.0).



Social Sciences/Policy

- Social Sciences/Policy faculty scored two security and compliance related services and two infrastructure services moderately higher than faculty overall: *C4 - Improve and clarify storage and computational options for regulated research* (2.17 vs 1.77), *B4 - Facilitate Cloud AND On-Prem options, including SOM researcher access OIT's services for non-clinical research* (2.33 versus 2.0), and the already-prioritized *B2 - Institute protected enclaves to encapsulate individual projects/data* (2.33, 2.03), and *C5 - Use risk-based approach to reviewing research and setting requirements* (2.5 vs 2.1).
- In other cases, ratings were considerably lower for E6 (training for faculty and students 1.5 vs 2.21) and D3 (offer secure DCC services, 1.5 vs 1.9), moderately lower for F1 (storage tools for data lifecycle, 2.0 vs 2.38), D4 (education cluster, 2.0 vs 2.33), and D8 (support faculty startup packages, 1.5 vs 1.93), and slightly lower for D1 (enhance DCC, 2.0 vs 2.33), D2 (GPUs in DCC, 1,83 vs 2,13), D7 (improve web access to DCC, 1.5 vs. 1.79), F3 (provide storage to meet regulatory requirements, 2.0 vs 2.26), and F6 (data continuity services, 1.83 vs 2.08)