Research Support Needs Initiative: Assessment, Recommendation, Action

John Board - ECE/CompSci faculty and Associate CIO
Tim McGeary - Assoc Univ Librarian for Digital Strategies & Technology
Rebecca Brouwer - Director, Research Initiatives
Overview

- Take you through a 4-phase process (underway) at Duke to improve research support for researchers across the University
  - Process
  - Findings
  - Next steps
The story begins

- CIO Tracy Futhey opted to revisit how OIT supported research computing from top to bottom
  - How do we support the non-HPC crowd?
  - Especially in the social sciences and humanities who have few resources?
  - Federal government is rolling out new data management requirements
  - Managing digital storage is hard
  - Challenge of having separate IT structures on campus and the health system

- Plan: Let’s engage the faculty and ask them!
First stop: IT Governance at Duke

- Tracy Futhey (CIO) convenes biweekly the IT Advisory Council (ITAC), with 15 faculty members, plus undergraduate and graduate students, and senior IT leadership from OIT and our schools.
- Desired more comprehensive solicitation of faculty opinion on research support. So another 50 faculty, in 7 groups, representing most non-clinical research disciplines were added.
Four-Phase Process

**Phase 1**
Assess Needs Duke-wide
Faculty Driven
February-November 2022

**Phase 2**
Propose Solutions
Service Partner Driven
January-July 2023

**Phase 3**
Determine Structures
Institutionally Determined
July 2023-present

**Phase 4**
Implementation
Service Partner Driven
PHASE 1

Assess

2022
Process

- Listening sessions with 7 groups of faculty (N=37) in:
  - Natural Sciences
  - Social Sciences
  - Basic Sciences
  - Engineering
  - Humanities/Arts (3 working groups)
- Faculty-prepared summaries shared with ITAC
- Two poster sessions distilled faculty feedback
Summary Findings and Recommendations
(REPORTS available at: https://duke.is/research-support-needs)

People: Expand and Improve (IT) Support

A. Duke lacks sufficient personnel to support domain specific research
1. Build and support new teams of domain-specific technical personnel
2. Develop & catalog training resources as an ongoing education program

B. Separate research infrastructures hinder research and collaboration
3. Objectively assess costs/benefits of dual and decentralized research IT infrastructures, which confuse and frustrate faculty

C. Current security / compliance approaches seem “one size fits all”
4. Evaluate current policy, security and compliance IT-related requirements and processes toward a holistic risk-based institutional approach

Process: Reduce Structural (IT) Barriers

D. OIT Services are valuable but not as expansive as faculty require
5. Evaluate approaches to extend OIT’s computational services (HPC, GPU)
6. With faculty input, tune services to better support faculty need (data, ML)

E. Plethora technical solutions and use constraints create confusion
7. Clarify / simplify technical solutions for particular research uses, stressing common services available to both campus and SoM (both cloud & local)

F. Current storage services don’t span research lifecycle or university
8. Implement long term storage options spanning campus and SoM
9. Automate data migration over lifecycle
10. License datasets as we do software

Technology: Enhance and Simplify IT Offerings

STRATEGIC ADVANCES

SERVICE ENHANCEMENTS
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.
So it’s not IT alone... Co-Sponsorship

- OIT
  - Research Computing
  - Information Security
- Duke University Libraries
  - Center for Data Visualization Sciences
  - ScholarWorks: Center for Open Scholarship
- Office for Research and Innovation
  - Protected Data Services
  - Scientific Integrity
  - Resource Navigation
PHASE 2

Recommendations
Spring 2023
Working Group Formation / Proposal Development

● Three co-sponsors identified leaders, service providers, faculty champions and stakeholders to take part in 6 working groups
  ● The 55 individuals met weekly for 10 weeks
  ● Identified 39 potential services that could meet the needs described in Phase 1
● Consolidated priorities to 29, then 21
● Surveyed faculty + estimated costs
● Landed on top 12 priorities to advance to Phase 3
39 Service Proposals From 6 Working Groups

**People: Expand and Improve IT Support**

A1 – Add 15-20 FTEs to support research
A2 – Build cross-department virtual teams
A3 – Connect IT personnel via skills matching and training
A4 – Fund graduate students to support undergrads and less expert grad students

**Process: Reduce Structural IT Barriers**

B1 – Create a single protected research network
B2 – Institute protect enclaves to secure data
B3 – Protect data based on project, not dept
B4 – Facilitate cloud + all on-prem solutions
B5 – Offer consulting on research-relevant sol’n
B6 – ‘Mini workshops’ to learn from other AMCs

C. Current security / compliance approaches seem “one size fits all”

C1 – Add 4th data classification; ease req on non-regulated sensitive data
C2 – Improve selection tools that meet sec. req.
C3 – Create consulting for what options apply
C4 – Improve options for regulated research
C5 – Use risk-based approach to requirements

**Technology: Enhance and Simplify IT Offerings**

D. OIT’s DCC Services are valuable but not as expansive as faculty require

D1 – Enhance DCC’s HPC compute offering
D2 – GPUs on par w/HPC
D3 – Offer equiv. secure compute options
D4 – Create CPU/GPU education cluster
D5 – Boost cloud bursting
D6 – Establish a graphics intensive cluster w/in DCC

E. Plethora technical solutions and use constraints create confusion

E1 – Establish a Self-service option to locate resources appropriate to the task
E2 – Consulting for what options apply
E3 – Guidance for cloud vs. on-prem solutions
E4 – Support website needs of researchers
E5 – Improve service communication/marketing
E6 – Training programs faculty, grad stu, ugrads
E7 – Establish clearer governance for sec./comp.

F. Current storage services don’t span research lifecycle or university

F1 – Better tools to manage data over lifecycle
F2 – Aggregate demand for dataset licensing
F3 – Provide flexible storage to meet regulatory requirements for secure + public access
F4 – Guidance/tools to select suitable solutions
F5 – Consulting on which options meet needs
F6 – Create data continuity services to ensure data integrity and availability
Faculty survey

- All participants (N=58) from phases 1 & 2 invited to complete
- 67% response rate
- 56%-75% from each domain surveyed
- Faculty heeded request to spread their scores across the 21 services
  - Even lowest rated service overall was rated highest by some
29 Distinct Services (Faculty/Sponsor Priorities) + Cost Estimates

- High Impact, Low Cost Priorities n=3
- Sponsor Additional Priorities n=4
- Faculty Top Quartile Priorities n=5
- On-the-Bubble n=2
- School-Based Priorities (Mixed Overall) n=5

- Advance to Phase3: Identify Funding
- Evaluate Further
- Candidates for Local Pilot or Collaboration w/ Schools
29 Distinct Services (Faculty/Sponsor Priorities) + Cost Estimates

- **Advance to Phase 3: High Strategic Value, Low Cost**
- **Advance to Phase 3: Sponsor Additional Priorities**
- **Advance to Phase 3: Faculty Top Quartile Priorities**

### Cost Estimates
- Exceeds $750K
- $150K - 750K
- Less than $150K

### Priorities
- **#1 (A1)**
- **#2 (F1)**
- **#3 (D1)**
- **#4 (F3)**
- **#5 (E6)**
- **#6 (D2)**
- **#7 (C5)**
- **#8 (A2)**
- **#9 (B2)**
- **#10 (E1)**
- **#11 (D3)**
- **#12 (D8)**

### Consideration
- **Consider at School-Level:** High Rating w/in Specific Scholarly Domain → Candidate for Future Pilot / Service Collaboration
- **Deferred for Further Study**

### Legend
- **Bubble Size - Est. Annual Cost:**
  - Exceeds $750K
  - $150K-750K
  - Less than $150K

- **Bubble Color - Working Group:**
  - Orange: Group A Services - Personnel
  - Brown: Group B Services - Two Infrastructure
  - Yellow: Group C Services - Security/Compliance
  - Blue: Group D Services - OIT Services
  - Red: Group E Services - Service Navigation
  - Green: Group F Services - Storage Service

### Note
- Eight lower priority services not in faculty survey were placed on X-axis based on poster session ratings.
**People: Expand and Improve Support**

**A.** Duke lacks sufficient personnel to support domain specific research

- Add 15-20 FTEs spanning Libraries, ORI, OIT & Schools plus 1-3 FTEs for research program and to support
- Build cross-department virtual teams; develop specific job expectations for research support professionals

**Technology: Enhance and Simplify IT Offerings**

- Make more powerful and flexible DCC capacity available more broadly
- Outfit shared/scavenge GPUs in DCC
- Provide secure DCC services
- Support faculty startups and semi-autonomous sub-clusters

**B.** Separate research infrastructures hinder research and collaboration

- Institute protected enclaves to encapsulate individual projects/data with the necessary security protections as easy but secure way to move data in/out

**Technology: Enhance and Simplify IT Offerings**

- Develop training programs for faculty and students; ensure IT personnel are well trained on research support services
- Establish a Self-service Option like the Cornell "Finder Tool"

**C.** Current security / compliance approaches seem “one size fits all”

- Use a risk-based approach on a project-by-project basis to review and set security and compliance requirements, with clear guidance on exceptions and risk acceptance, based on magnitude

**Technology: Enhance and Simplify IT Offerings**

- Devise tools to a) manage data over its life cycle; b) understand storage cost; and c) identify where data reside
- Enable storage flexibility to meet varied research needs (secure + public access) and compliant w/ regulations

**Process: Reduce Structural Barriers**

**A.** Duke lacks sufficient personnel to support domain specific research

**Technology: Enhance and Simplify IT Offerings**

- Develop training programs for faculty and students; ensure IT personnel are well trained on research support services

**B.** Separate research infrastructures hinder research and collaboration

- Institute protected enclaves to encapsulate individual projects/data with the necessary security protections as easy but secure way to move data in/out

**Technology: Enhance and Simplify IT Offerings**

- Devise tools to a) manage data over its life cycle; b) understand storage cost; and c) identify where data reside
- Enable storage flexibility to meet varied research needs (secure + public access) and compliant w/ regulations

**C.** Current security / compliance approaches seem “one size fits all”

- Use a risk-based approach on a project-by-project basis to review and set security and compliance requirements, with clear guidance on exceptions and risk acceptance, based on magnitude

**Technology: Enhance and Simplify IT Offerings**

- Devise tools to a) manage data over its life cycle; b) understand storage cost; and c) identify where data reside
- Enable storage flexibility to meet varied research needs (secure + public access) and compliant w/ regulations
Twelve Service Proposals Form 3 Service Clusters

Better Support Researchers by Adding Personnel, Improving Coordination, and Easing Service Discovery

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.
5. Develop training programs for faculty and students (grad and undergrad) and ensure IT personnel are well trained on research support services.
8. Build cross-department virtual teams, to better support personnel across Schools as well as ORI, OIT, and Libraries, with 1-3 FTEs to manage, develop and support the personnel.
10. Develop a self-service tool to guide service selection based on data classification, access attributes, etc. (like Cornell’s “Finder tool”).

Enhance Computational Services and Build Capacity for Data Intensive Research

2. Devise tools to manage data over its life cycle, understand storage cost, and identify where data reside. Provide storage capacity to meet 80% of active research project need.
3. Enhance processing and memory VM provisions in the DCC that are available to all researchers and extend access to graduate (PhD) students and postdocs.
6. Better support AI/ML and other research through GPU capacity like DCC’s on-demand CPUs access (shared and scavenged).
12. Support faculty startup packages / semi-autonomous sub-clusters, supporting direct and immediate access while also expanding the DCC and leveraging spare cycles.

Balance Security and Compliance Requirements with Flexibility Needed to Support Different Types of Research

4. Provide storage flexibility to meet differing research needs (secure + public access) that are compliant w/ regulations for storage retention.
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.
9. Institute protected enclaves to encapsulate individual projects / data with the requisite security protections that enable authorized access and data movement based on the project circumstances.
11. Provide secure DCC (Duke Compute Cluster) services that are functionally equivalent to OIT’s existing virtual machine (VM) and other offerings.
PHASE 3

Determine Structures
Summer 2023
Seeking Funding through Partnership

- Convene planning team of financial leads & sponsor reps
- Identify workable ongoing funding options for each prioritized service
  - (e.g., Charge to grants, funded via philanthropy, allocations-funded, etc.)
- Develop multi-year funding / walkup plans including, as needed, bridge funding
- Ensure sponsor backing
Financial Proforma

- Co-Sponsors + Co-Funders developed a 5-year proforma
- Categories:
  - Research Support
  - Storage Services
  - Service Navigation
  - IT Services and Infrastructure
  - Security and Compliance
- Initial $1.5M investment to increase to $5M+ over five years
- Exploring philanthropic strategies with Duke Development
Types of Positions

- Office of Research
  - Data security
  - Research Data Navigators

- Libraries
  - Data visualization
  - Research Data Management
  - Data Licensing
  - Metadata
  - Repository Developer
  - Digital Humanities / GIS

- OIT
  - Technical Personnel
  - Website Consultation
  - Software Developer
  - Business Analyst
  - Application and DB Maintenance
  - Client Support
  - Security Analyst

- Domain based Experts
  - Social Science / Data
  - Digital Humanities / GIS
  - Natural and Basic Sciences

- Cross Departmental Leadership
Types of Services

- Cloud infrastructure and storage to connect Active to Published research
- Building protected enclaves for research data
- Collaborative licensing strategies to enable campus-wide and context-limited data sets
- Semi-autonomous and reusable sub-clusters to support Faculty Startup Packages
- Enhanced Duke Compute Cluster
PHASE 4

Implementation
Fall 2023 - present
First Steps

- Implementation Working Group: OIT, Libraries, and Research
- Begin planning around ‘virtual teams’ to engage school-level resources + central units
- Prioritize hiring; Develop job descriptions
- Explore how / where to connect other services from providers, not explicitly included in this study but relevant to research
- Determine metrics: demand, capacity, and success
- Determine added ‘governance structures’ as may be needed
5 Services Already Being Actively Developed

People: Expand and Improve (IT) Support

A. Duke lacks sufficient personnel to support domain specific research
1. Add 15-20 FTEs spanning Libraries, ORI, OIT (& Schools) plus 1-3 FTEs for research program and to support (2.)
2. Build cross-department virtual teams and develop specific job expectations for research support professionals

D. OIT Services are valuable but not as expansive as faculty require
5. Make more powerful and flexible DCC capacity available more broadly
6. Outfit shared/scavenge GPUs in DCC
7. Provide secure DCC services
3. Support faculty startups and semi-autonomous sub-clusters

Technological: Enhance and Simplify IT Offerings

B. Separate research infrastructures hinder research and collaboration
3. Institute protected enclaves to encapsulate individual projects/data with the necessary security protections as easy but secure way to move data in/out

C. Current security/compliance approaches seem “one size fits all”
4. Use a risk-based approach on a project-by-project basis to review and set security and compliance requirements, with clear guidance on exceptions and risk acceptance, based on magnitude

E. Plethora technical solutions and use constraints create confusion
9. Establish a Self-service Option like the Cornell "Finder Tool"
10. Develop training programs for faculty and students; ensure IT personnel are well trained on research support services

F. Current storage services don’t span research lifecycle or university
11. Devise tools to a) manage data over its life cycle; b) understand storage cost; and c) identify where data reside
12. Enable storage flexibility to meet varied research needs (secure + public access) and compliant w/ regulations
Next Steps in 2024

- Provost authorized launch of first wave of hiring and service implementation
- Implementation Working Group finalizing priority order to match demand and current gaps
- Virtual teams being formed to begin collaboration
Takeaways

- Some of our issues are likely common to other institutions of higher education
- Some of our issues are unique to schools with closely coupled academic medical centers
- Faculty input coupled with leadership engagement and service provider expertise is critical to enhancing research support services
Questions?

https://duke.is/research-support-needs

john.board@duke.edu
tim.mcgeary@duke.edu
rebecca.brouwer@duke.edu