

Positions for Fiscal Year 2023

Institutional CyberInfrastructure Group

Final Draft: Jul 22, 2022

Endorsed by ICIG Members and RCC co-Chairs:

- Cody Hanson
- Colby Reese
- Daniel Stocker
- James Wilgenbusch
- Charles Nguyen -- RCC co-Chair
- Dan McDonald -- RCC co-Chair

Purpose

Why These Positions Matter:

- **Develop** consensus on the priority areas where improvements to our research cyberinfrastructure are needed and where tangible improvements can be made in the coming fiscal year.
- **Focus** ICIG's operational activities to be in the service of these strategic positions.
- **Communicate** both up (to our leaders) and down (to those implementing changes) our top priorities, promoting alignment of goals and eliciting new ideas from everyone involved.
- **Promote** a culture of transparency and shared university-wide mission.
- **Identify** redundancies and minimize duplication of effort.

Cross Cutting Themes:

- **Consult Peers, Lead Where Practical:** Look outside of our institution for ways that other institutions are addressing the challenges described below. Where there are gaps in processes, decide whether that gap represents an opportunity to lead, to be first.
- **Align and Collaborate:** Where possible, make a strong effort to align and collaborate on shared mission efforts with other UMN technology providers.
- **Promote Diversity Equity and Inclusion:** Seek opportunities to advance DEI in our activities by sharing experience that work and cross listing opportunities.

Positions

Position 1: Improve Support for Research CI Security and Compliance

Why It Matters

The integrity, privacy, and availability of critical research CyberInfrastructure (CI) and data are increasingly under attack. Additionally, our interaction and partnerships with industry and the

private sector requires a degree of privacy and systems security assurance that open and public data do not afford, in part because of concerns regarding intellectual property and reputation. This phenomenon is not unique to the University of Minnesota - it is increasing the case that large R1 academic research institutions are having to face this. The federal government and private research sponsors are keenly aware of the emerging threats and have codified their expectations in the form of laws and regulations (e.g., HIPAA and ITAR) and more general agreements (e.g., NDAs). The threat landscape is constantly changing, and so too are the requirements to support small and large research activities. Better addressing emerging cybersecurity threats is essential to advancing our ability to compete successfully for externally sponsored research and fulfilling a foundational obligation of our public trust.

Current Challenges

- Lack processes for bringing new technologies into production under various compliance frameworks
- Fail to leverage existing UMN talent to address emerging opportunities
- Lack a holistic view of some sensitive workflows results in duplication of effort to address compliance requirements
- UMN's Data Classification policy is difficult to navigate and does not offer the degree of granularity needed to address the broad spectrum of research requirements faced by our researchers.

Vision of Success

- A data classification system that is easy to use and addresses the broad spectrum data types that our researchers face
- Well defined processes for bringing new technologies into production to address research and development opportunities.
- Collaborative approaches that leverage existing IT talent to address scale with emerging challenges.
- Engagement of Research Computing, Data, and IT professionals at all stages of the research lifecycle -- from grant development to result archiving.

Recommended Next Step(s)

- Charge an ICIG led working group to convene relevant stakeholders to identify bottlenecks and propose tangible solutions across the entire spectrum of protected and sensitive data.

Position 2: Improve Support for Team Science and End-to-end Workflows

Why It Matters

The evidence for the significant impact that Team Science has had in advancing knowledge in all areas of scientific inquiry is overwhelmingly strong¹. Team Science matters to the University of Minnesota because it has become the prevalent approach to conducting research. Therefore, the purpose of this section is to highlight its importance and the needs for this community to address gaps in our approach to supporting Team Science, not to make a case for its importance. When viewed through a Team Science lens, OIT, MSI, and Libraries operate as nodes in a larger network of multidisciplinary collaborations. Where entire experiments were once conducted from beginning to end on a campus High Performance Computing (HPC) systems, it is now common for the HPC system to serve as a part of a larger workflow, which might include remote instruments, such as DNA sequencers, microscopes, or a repository of remote data, such as multispectral satellite imagery. Viewed more broadly, the workflows that we facilitate at the University of Minnesota in some cases serve as inputs to larger, national or global research collaborations, giving us a platform to broaden the impact of our research and attract additional sponsorship. In short, better integrating research computing and data systems and processes into the broader research fabric of our campuses and of the nation is essential to sustaining the advances made possible by Team Science.

Current Challenges

- Lack a coherent view of what resources (e.g., on premise HPC, national HPC resources, archive storage, compliant storage, cloud resources, etc.) can be made available to address research workflows.
- Lack a coherent strategy to support the data and data processing needs of our Shared Research Facilities (SRFs)
- Lack a holistic view of some research workflows, which can result in duplication of effort

Vision of Success

- Researchers get a coherent view of a wide-range of cyberinfrastructure platforms by looking on-line or talking to any UMN research computing or IT professional.
- Data moves seamlessly from instruments into the next step of the data lifecycle.
- Engagement of Research Computing, Data, and IT professionals at all stages of the research lifecycle -- from grant development to result archiving.

Recommended Next Step(s)

- Charge ICIG to advance and coordinate activities specifically designed to support Team Science, such as, SURFs (<https://www.msi.umn.edu/surfs>). Part of this would include coordinating with RCC to develop metrics and create accessible user documentation.

¹ Wuchty, S., Jones, B.F. and Uzzi, B. 2007. The increasing dominance of teams in production of knowledge. *Science*. 316, 5827 (May 2007), 1036–1039. DOI:<https://doi.org/10.1126/SCIENCE.1136099>.

- Hire an RC project coordinator who will interface with ICIG, RCC, and SRFs.
- Support RCC in growing the network of champions and developing tools and resources for champions to connect researchers with cyberinfrastructure platforms.

Position 3: Improve Support and Awareness for Data Management Lifecycle: Develop a culture that embraces Data as an Asset

Why It Matters

Data are at the heart of modern research. They are the evidence, the facts, the truth on which assumptions are tested and new ideas formed. They have become an essential part of scholarly publication, and are crucial to ensuring reproducibility. The University of Minnesota values data as a strategic asset and as a mechanism for impacting the region and the world through our research. We value our physical assets like buildings and equipment and spend millions a year keeping these up to date. Our data require similar investment to ensure that we can sustainably gather, maintain and share this important scholarship.

Current Challenges

- In general, the University does not have a sustainable storage model that keeps pace with existing data growth/demands.
- Compliance context switching and interpretation generates long lead times for project implementation.
- Need for more robust and consistent governance, tools and support structures for data management
- [2023 NIH data sharing policy](#) will require many more researchers to have a data management plan in which researchers will have to describe how they plan to share their data.

Vision of Success

- Broader consensus and understanding of how to address the need for data privacy and data openness.
- Increase visibility and appreciation for the cost of data services so that researchers can understand and make informed decisions
- Broad access to tools, training and support structures for data management including those that address Options that simplify compliance with legal protections of data as well as more general Data Use Agreements

Recommended Next Step(s)

- Support RCC in propagating good stewardship practices to researchers and engage with RCC champions to migrate static data from active storage locations to new Static Storage platforms