# A Comprehensive Campus-based Approach to Address the Opportunities and Challenges Posed by Data Intensive Research

James C. Wilgenbusch[1], Joshua A. Baller[1], Carla Bates[2], Evan Bollig[1], Lisa Johnston[3], Claudia Neuhauser[4].

[1]Minnesota Supercomputing Institute, [2]Psychology, [3]Libraries, University of Minnesota; [4]University of Houston

INTRODUCTION

Faced with escalating expenses related to data storage needs and a capable set of on campus storage service providers, the University of Minnesota (UMN) developed a comprehensive framework to better address current and emerging challenges and opportunities brought by data intensive research. Elements of this framework and the process used to develop it could be applied at other research institutions to advance their efforts to address the challenges they face supporting data intensive science. While approaches may differ slightly, addressing these challenges within our respective universities is critical, and perhaps a prerequisite to building and sustaining partnerships among providers of advanced research computing and fully realizing the value of our data.
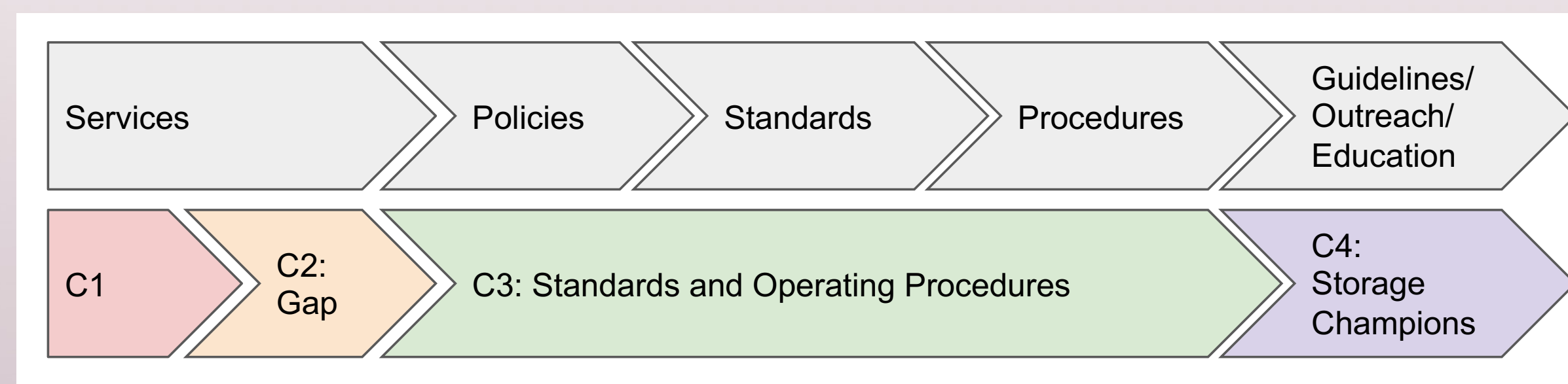
OUR APPROACH

The approach taken at the University of Minnesota aimed broadly at creating a user-centric framework for addressing the current and future data storage needs of our researchers. As part of our process, a data storage working group consisting of faculty and staff members representing our libraries, academic units, and campus service providers was assembled in 2016 and asked to develop the charge and charter for a larger committee that would be responsible for:

1. Inventory current data storage resources
2. Characterize data storage needs and identify gaps
3. Inventory UMN policies and practices
4. Recommend actions to promote best practices
5. Identify ways to better educate storage consumers and providers and champion existing storage services

This working group became known as the Storage Redesign and Restructure Committee (SRRC) and had an equal number of faculty and staff among its 60 members.

KEY FINDINGS

Over the course of six months, four key findings emerged from the SRRC:

1. At least 45 petabytes of data were stored across university data storage solutions, requests for more storage were increasing, and there was little coordination among university service providers.
2. Strategies and technical changes to address this influx were made solely at the local service unit level and were not communicated to or coordinated at a higher level.
3. Lack of coordination was preventing alignment of storage portfolios across providers. Implementation of duplicative and overlapping resources reduced efficiencies and diluted more targeted investment in addressing gaps and service improvements.
4. Ultimately, there was not a continuous holistic review of storage supply and demand to ensure existing and future needs were being addressed efficiently.
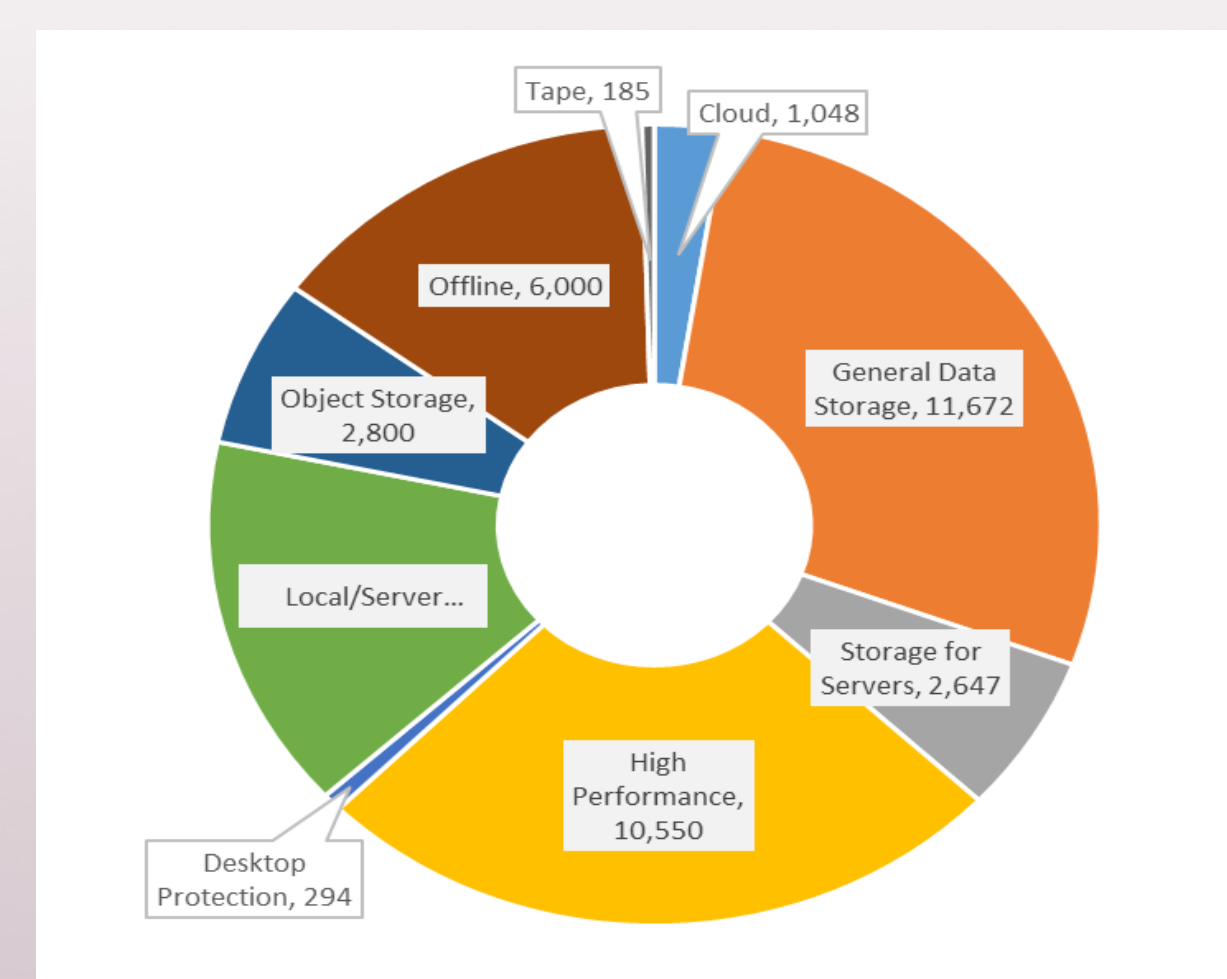
RECOMMENDATIONS

Together, the members of the USC developed the following overarching goals to address immediate storage related needs and to recommend longer term strategies related to data intensive research.
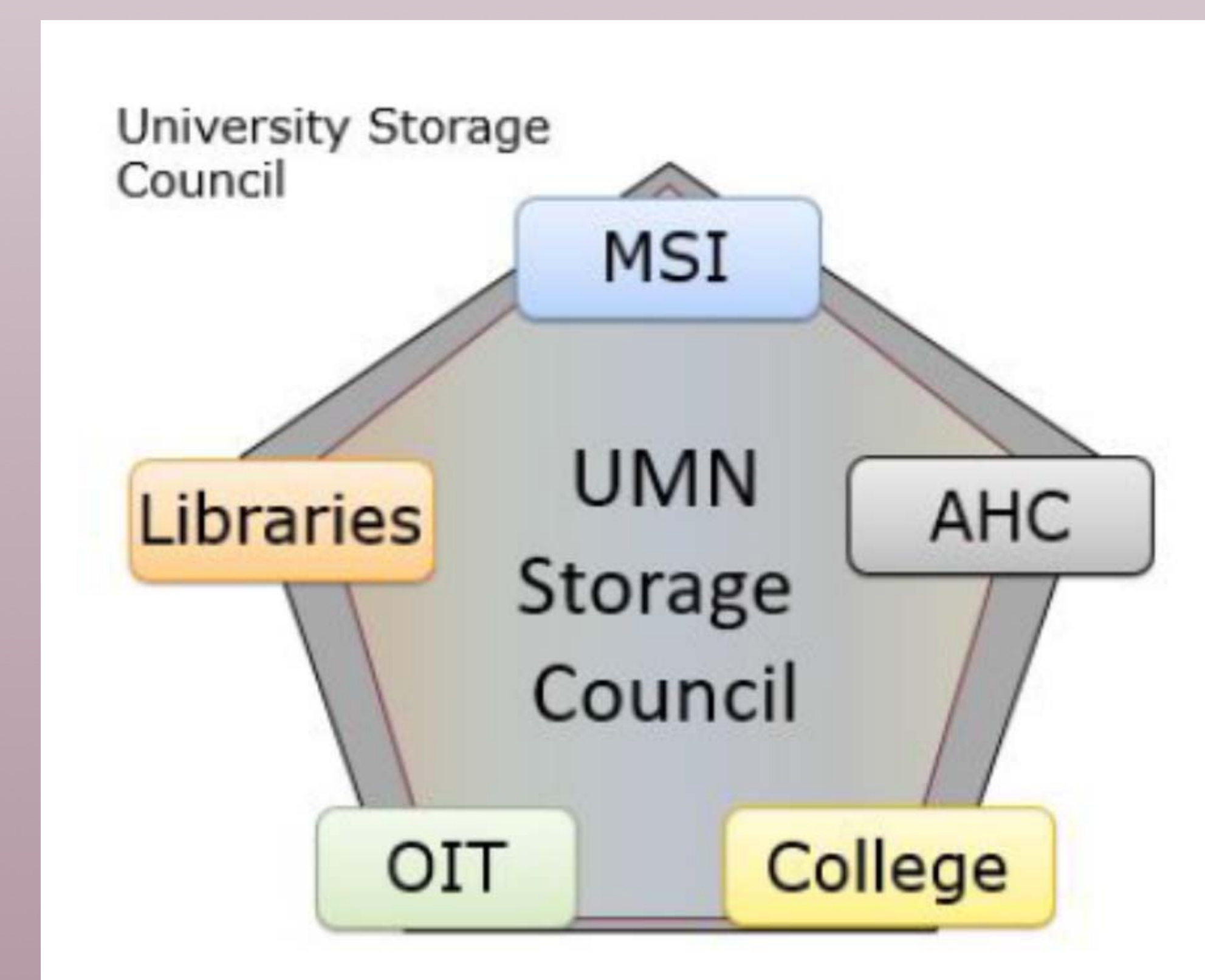
- **Establish a Storage Champion Program:** Mobilize a storage champion program to provide the institution with a distributed team of knowledgeable employees whose focus is to deepen relationships and to provide a bridge of expertise between the requestor and the provider.
- **Enhance Website Infrastructure:** Provide a single-entry point for all storage requests, reports, (e.g. usage and show back) to improve the delivery of storage resources in the most cost- effective manner and to inform consumers of cost and usage of the resources.
- **Enhance User Training and Onboarding:** Incorporate and/or offer additional training at the point of the request to educate and inform the user at the most propitious moment.
- **Collaborate and Share Internal Knowledge Articles**: Ensure that all Knowledge Base (KB) articles from across all providers can be located in a singular accessible source, thereby allowing more cross-information sharing and providing the storage champions as well as the multiple support units a unified source in which to access all storage related information.
- **Promote, Market, and Communicate:** Create a comprehensive marketing and communication campaign to help promote the newly coordinated University storage service model. Users will be made more aware of storage best practices, storage service selection tools, and storage champions.

CONCLUSION

Data intensive research is pushing the limits of what we thought possible. The volume, rate, and diversity of data being acquired on our campuses now is introducing a new set of challenges to sustaining university-based research cyberinfrastructure. These new and constantly evolving challenges are unlikely to be solved by one-off technology-based solutions, designating a single campus resource provider, or "moving everything to the cloud." Successfully sustaining campus resources for data intensive research requires coordination among the following set of campus stakeholders.

1. Upper level administration
2. Senior and junior faculty members
3. Central research computing and IT directors and managers
4. Departmental IT staff
5. University Libraries

A coordinated and comprehensive approach will not only greatly facilitate campus-level collaborations, but in many cases will also be needed to effectively collaborate and form effective research partnerships at regional, national, and global scales. In addition, campus-level coordination efforts can clarify the value proposition for data, so that the data on our respective campuses are seen as an asset rather than a liability. Given the critical role that academic institutions play in research and discovery and in training the next generation of researchers, it is imperative that we take a more holistic approach to address these challenges.