CC* Data: An Open, Federated, Sustainable Storage Cloud for the Tennessee Research and Education Community

While cyberinfrastructure providers have long had the software infrastructure (e.g., Unix/Linux and TCP/IP) necessary for sharing computing and networking resources in an interoperable fashion, they have enjoyed no such luxury in the case of storage resources. But after more than a decade of explosive growth in the volume and velocity of data generation in nearly every field of research, the pervasive Balkanization of storage and caching/buffering resources is making problems of *data logistics* (i.e., of "ensuring that data [are] in the right place at the right time and accessible by the right users" [1]) increasingly intractable. Researchers who need to make large amounts of data suitably proximate to their colleagues in a timely way, within a distributed community of collaborators who have disparate access to disparate computing resources, frequently find that their efforts are impeded or frustrated in one way or another, e.g., because of network bottlenecks, non-interoperable data management tools, cumbersome and community-unfriendly security controls, etc. More than fifteen years into the 21st century, we still often find, even at the campus level, that essential, data-intensive research workflows routinely rely on a combination of big portable hard-drives and FedEx or sneakernet as the only practical solution.

*The goal of the **Tennessee Open Research Cloud** (TORC) project, described below, is to create a wide-area, high-performance and interoperable storage infrastructure, one which is designed for scalable, multi-level federation under cooperative management, and which thereby shows how this critical challenge to all data intensive research can be overcome in a sustainable way.* TORC will build on the NSF funded Data Logistics Toolkit [2], a solid package of software components that emerged from more than a decade of research in Logistical Networking [3–5]. Explicitly designed and developed to help distributed and data intensive research communities confront their problems of data logistics, it integrates and packages a set of individually well-tested storage and networking technologies for creating topologically embedded, topology aware storage nodes, or *depots*, to support the design and implementation of WAN-enabled storage fabrics that dramatically improve the ability of data intensive research communities to collaborate.

PI Sheldon and his team at Vanderbilt developed the DLTs production-ready storage depot and L-Store file manager, which they have used to successfully create and deploy a model for a node of the open storage cloud that TORC envisions. Significant portions of this system have been in production at Vanderbilt for several years, enabling physicists who work on the CMS experiment at CERNs Large Hadron Collider, to explore and analyze the more than 4PB of data resident at Vanderbilt. Exploiting the unique ability of this model to facilitate scalable federation at three different levels—for sharing of raw storage, for cross-site data access, and for unified, authenticated and distributed data service—TORCs team of collaborating researchers will work with cyberinfrastructure providers on their respective campuses to build TORC.

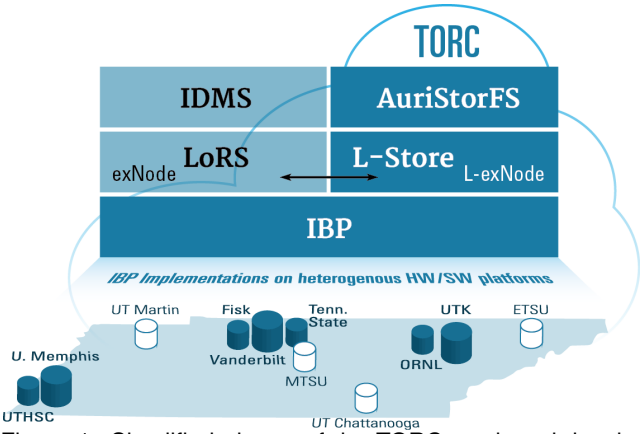

Figure 1: Simplified picture of the TORC stack and the deployment of depots in TORC's core (dark blue) and prospective partnering sites (white).

As shown in Figure 1, as a unified storage facility that spans the state of Tennessee, TORC will enable researchers from across the state to share their data with each other and to transparently access it from scientific computing resources within the state. Because of the unique properties of its simple, generic and limited storage virtualization protocol (IBP, see 1.2), the same depot infrastructure that supports TORC's L-Store/Auristor approach to wide-area storage federation, could also be used by participating campuses to support the LoRS/IDMS components of the DLT, which are designed primarily for building content distribution networks (CDN). CDNs are not part of TORC plan during this cycle. The illustration also shows the initial deployment of core depots to TORC collabo-

rators across Tennessee (dark blue), as well as additional partners we expect to recruit during Phase 2 or Phase 3 of the projects.

The work undertaken to achieve goal this falls into three main categories:

- *Deploy TORC core depots and collaborate with applications partners to create TORC workflows and services*: In order to deliver benefits to participating application groups and attract new users, in phase one we will deploy the core TORC depots in Nashville, Memphis, and Knoxville, and bring up TORC distributed file system services in preproduction mode. Working with our application partners, we will use this preproduction deployment to establish collaborative workflows that will help tune and harden the infrastructure and that demonstrate the campus level and state level benefits of TORC's federation model. During this period AuriStor will be integrated with L-Store using a native plugin whereby all data and metadata flow through the AuriStor file server. This approach is not well suited for HPC applcations.
- *Extend TORC middleware and integrate with distributed file system technology*: Along with enhancements to core TORC components (e.g., improved performance and robustnes of the DLT depot code), in phase two we will develop some key pieces of middleware that will make it significantly easier for our collaborating communities (as well as many other research, education, and non-commercial communities) to design and deploy TORC services that solve their data logistics problems. Additionally, outside the scope of this proposal, the native L-Store plugin for AuriStor will be extended to support third party data transfers. This will allow clients that have access to the depots to directly access the storage, bypassing the AuriStor file server completely for disk I/O. All metadata operations will still flow through AuriStor. This approach is designed for supporting HPC workloads.
- *Integrate TORC services with local campus IT and reach out to federate with wide area partners*: In phase three we will work to catalyze significant new research capabilities by facilitating efficient data transfer and sharing and demonstrating an approach that can be applied nationally. To achieve this objective we will connect with and leverage the best approaches for managing data via the Big South Data Hub and collaborating with DataOne/Allard along with coordinating efforts with XSEDE.

1 Background

In this age of data intensive science, it is now routine to find that when the data that a research effort generates at any stage are large, collaborating researchers will confront structural barriers that impede to this workflow and severely restrict data movement and sharing. Although networking and computing resources tend to be highly interoperable, there are no comparable standards for storage interoperability. Moreover, and perhaps as a consequence of a lack of suitable standards, campus research groups and core facilities purchase and administer their storage infrastructure in isolation: it is often intended for use only by that group. This situation leads to significant heterogeneity: the hardware that is optimal for each group or facility varies, while administrative decisions regarding security and access are primarily made with the local group in mind. Firewalls or network bottlenecks often exist between campus groups who want to share data, which further exacerbates the Balkanization. Data movement tools compatible with the different operating systems and security protocols of each group often require a fair amount of hand-holding and are difficult to automate. It is not uncommon for researchers to resort to "moving data" by carrying hard drives from their lab to a core facility. All these data logistics problems only get worse as the number of administrative boundaries increase at metro, regional and national scales,

To create a storage infrastructure for data logistics that facilitates the automated and transparent sharing and utilization of massive amounts of data on and between campuses and institutions, the TORC architecture supports storage federation at three different levels: at the level of its storage virtualization protocol (IBP), between particular local file managers (L-Store), which form the nodes of the TORC infrastructure, and at the level of its overarching distributed file system (AuriStor file system), which offers the most powerful secure and consistent environment. Below, we explain the motivation for this layered approach to federation and then describe each of the different modes in turn.

1.1 TORC: A Design for Scalable, Interoperable Storage Federation at Multiple Levels

In order to solve the problem of data logistics for collaborative research described above, TORC's unique approach to storage federation, based on Logistical Networking [3] and the Internet Backplane Protocol (IBP) [6, 7], provides opportunities for institutions and departments to federate storage resources at three different levels, with different levels of openness and quality of service at each level. Underlying this tripartite

approach to federation is the familiar tradeoff that usually has to be made between the sophistication and complexity of a given service, and its deployment scalability: the more complex a service is, the less easily it can be expected to scale. Accordingly, to serve as the "spanning layer" and source of interoperability [8], IBP was designed as a simple, generic and limited storage protocol, so as to provide the most degrees of freedom and greatest potential for innovation; consequently, however, it also has the fewest and weakest services. By contrast, the L-Store file managers provide more services but at the cost of a greater degree of commitment to specific structures and algorithms. As we describe in more detail below, the Aristor distributed file system provides an extremely powerful, secure and consistent environment, but at the cost of a corresponding loss of generality, flexibility and performance. The essential point is that, such a layered approach is essential for a data logistics infrastructure that balances the ability of researchers to try new services and address unexpected opportunities, against the desires of administrators for security, close accounting of resources, compatibility with commercial products and so on. So the three forms of federation that TORC will support are characterized as follows:

1.1.1 Federation for sharing raw storage resources: Federation at the storage level enables the sharing of raw storage resources that are virtualized at each site using IBP. Sharing of such resources is possible between users who are not authenticated to the IBP depot and at sites that do not themselves host IBP depots. But features such as user authentication, high security, fault tolerance and synchronization of updates must be implemented end-to-end on top of this service. This level of federation supports those users, who require the greatest scale and highest level of performance in the use of departmental or institutional storage resources, and desire fully local control over such issues as provisioning, access, user identity, security and network utilization without requiring a high degree of cross-domain uniformity and update conherence. Federated raw storage resources are well suited to supporting innovative, unique, or specialized utilities.

1.1.2 Federation for cross-site data access: Federation at the site level requires authentication to the local L-Store file manager, which controls institutional storage resources and mediates sharing among local users. L-store supports a number of higher level file services such as security, fault tolerance, optimized access performance and file update synchronization. Federated L-Store servers provide access across sites, but do not implement unified and fully coherent distributed services on a global basis. This level of federation supports researchers who manage large data sets in the wide area and may require the freedom to experiment with and develop new protocols and data structures to achieve the highest levels of openness, scalability, performance, security, availability or other unique characteristics.

1.1.3 Federation for unified, authenticated, distributed data services Federation of AuriStor systems provides a fully unified and synchronized namespace for concurrent access to files through shared directories. Several features promise to simplify data services and enhance researcher productivity: single login can utilize a number of standard user identity services to simplify identity management; client interfaces are supported for all computational and personal computing platforms used by the R&E community; and resources managed by a variety of back end file systems, including but not limited to L-Store, can be seamlessly combined. This level of federation will serve the large class of users who need to share large data sets across sites and institutions and require unified and secure directory administration, fully coherent global caching and replication, support of commodity client platforms, and interoperability with a variety of legacy back-end file management systems currently in use. Federated AuriStor managers provide these benefits, along with a commitment to work to optimize access to L-Store and IBP backend resources to enable scaling and high performance to users of these lower-level service interfaces.

1.2 Software Base

The foundation of TORC software infrastructure is the *Data Logistics Toolkit* (DLT), an NSF CC-NIE funded software package that incorporates and extends the software components produced by the research program in *Logistical Networking*. PI Sheldon and his team helped develop this technology for the expressed purpose of addressing the data logistics problems faced by the large LHC and remote sensing/geospatial communities and by many other data intensive application communities.

The software base TORC builds on is very mature. While the first implementation of IBP's simple, generic and limited storage protocol was 16 years ago, it has seen particular success over the past decade as the basis for Vanderbilt's L-store system. In production service for over 7 years supporting both CMS and the Vanderbilt Television News Archive, it currently has over 6 PB of disk space with more than 4 PB used and routinely sustains 150 Gpbs or more of I/O bandwidth with jobs loads of over 5000 cores. The Vanderbilt

team has used this experience to significantly harden L-Store and IBP, and improve its fault tolerance and resilience. For example, Vanderbilt recently experienced mass coordinated disk drive failures due elevated temperatures from a fire in the data center. Although drives failed at the rate of one per day for several months, these failures were easily handled without any data loss.

Below we describe the main components of the DLT that TORC will utilize, including IBP and L-Store. We also describe AuriStor distributed file system, which, though not part of the DLT, will be integrated with DLT technology to provide the single global namespace that the most powerful forms of secure data service federation require.

1.2.1 The Internet Backplane Protocol (IBP) The IBP Storage Service [9] allocates, reads, writes, and manages storage in variable-sized chunks known as allocations. The server that implements the IBP service is called a *depot*. In order to separate access policy from the mechanism that implements it, access to an allocation is governed by read, write, and manage keys, or *capabilities*. An allocation has no externally visible address or name to which it can be referred by IBP clients other than these three capabilities which are returned by the depot as the result of a successful allocate operation. Designed explicitly to be a better form of storage for data logistics, IBP's leading characteristics can be summarized as follows: 1) *Thin* - IBP is a minimal/primitive interface that includes no features that can be correctly and completely implemented at a higher layer, unless there is a clear benefit such as performance or security; 2) *Open* - IBP can be given policy settings that enable clients to make allocations without any notion of identity or authorization, so that clients can reside anywhere that has connectivity and are not restricted to specific domains (e.g., a LAN or SAN); 3) *Limited* - IBP can enforce policy choices such as maximum allocation size and maximum duration, but one basic notion is that, since no provision is made for a "filesystem check" to verify that all unreferenced allocations are free, allocations are time limited and must be refreshed; 4) *Non-rendezvous* - To rule out the use of IBP as a rendezvous point (e.g., between a content publisher and consumers) without some other service to create a match, directory listings are hidden and the server chooses long, semantically neutral names (e.g., random strings); 5) *Best Effort* - IBP can be thought of as best effort because there are no guarantees of service quality (e.g., correctness or performance); 6) *Generic* - IBP does not include features that restrict the possible implementations; 6) *Third Party Transfer* - IBP allows third-party transfers between nodes as a primitive; 6) *Multiplexing* - The use of a depot by one user should not be directly visible in any way by other users, but at best only indirectly inferable through its effects.

1.2.2 ExNodes IBP is analogous to a block-level storage service and clients must aggregate these low-level storage allocations into larger structures like those found in file systems. The exNode (modeled on the Unix filesystem *inode*) is the data structure that aggregates storage allocations into a file-like unit of storage [10]. The exNode implements many, but not all, of the attributes that are associated with a file, leaving some, such as naming and permissions, to other services. While the exNode omits some file attributes, it also implements file-like functionality with greater generality than a typical file. For example, the exNode can express wide-area distribution of data replicas, whereas file systems are typically restricted to a local or enterprise network.

1.2.3 L-Store The L-Store software stack (see Fig. 2) provides a flexible logistical storage framework for distributed and scalable access to data for a wide spectrum of users. L-Store is designed to provide: virtually unlimited scalability in raw storage; support for arbitrary metadata associated with each file; user controlled fault tolerance and data reliability on a file and directory level; scalable performance in raw data movement; a virtual file system interface with both a native mount in Linux (exportable via NFS and CIFS to other platforms) and a high performance command line interface; and support for the geographical distribution and migration of data to facilitate quick access. These features are accomplished by segregating directory and metadata services from data transfer services. Logistical Networking is used

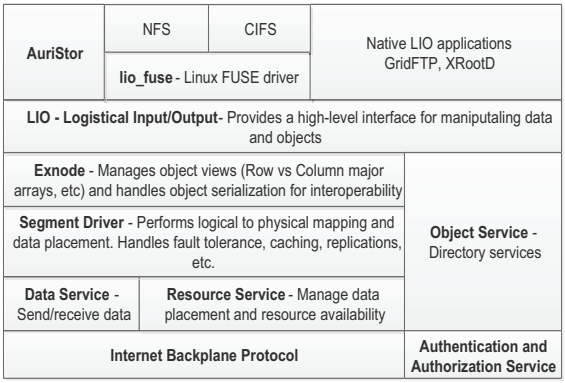

Figure 2: L-Store software stack

as the block-level data abstraction for storage with L-Store clients directly performing all data I/O without going through an intermediate server as is done for NFS and CIFS. This means adding storage adds both capacity and bandwidth.

L-Store is designed from the ground up to be extensible with support for dynamically loaded plugins for each of the architectural building blocks in the diagram. Data integrity becomes a much more significant issue at petabyte scales due to increasing probabilities of data loss and array rebuild times. At these scales, unrecoverable read errors start to become statistically significant. These issues are handled by the Segment Drivers in conjunction with the Data Service for performing I/O operations along with the Resource Service for tracking resource availability and handling data placement needs to satisfy fault tolerance constraints.

Data integrity schemes typically employ either data mirroring or some form of RAID. Data mirroring is nothing more than keeping multiple copies of the data stored on different devices at different locations. Mirroring is great for read dominated workloads since all replicas can be used to service user requests. Write performance is slowed by the making of all the replicas. Mirroring is not very space efficient, becoming quite costly at the petabyte scale.

L-Store has implemented generic Reed-Solomon encoding and several other data integrity schemes. We typically recommend 6+3 Reed-Solomon (RS-6+3) encoding: 6 data disks and 3 parity disks placed on separate storage depots. More generally we recommend the use of the RS-d+(d/2) family of configurations – RS-6+3, RS-8+4, and RS-10+5. This family uses 2/3 of the total space for data with the remainder for parity and has excellent reliability. RS-6+3 has the same reliability as keeping 3 copies of the data but uses only half the space.

The L-Store web management interface is designed to give a helpful overview of the storage pool to technical staff and data users/owners. The web interface offers a hierarchical (global, site, and server) real-time overview of the storage pool, and gives easy access to filesystem operations such as failing over a drive or monitoring rebuild progress, hardware management such as powering a depot or hard drive on/off or turning on the slot locate LED, performance information such as disk transfer speed and network bandwidth, and diagnostic information such as hardware status and SMART errors.

1.2.4 AuriStor The AuriStor File System provides a global namespace that allows secure, platform-independent, and high-performance access to data in a transparent manner. AuriStor is an enterprise grade extension of the Andrew File System (AFS) but with modern security, performance, scalability, and functionality. The AuriStor product development began in 2008 under a Department of Energy SBIR to create a High Performance Global File System to meet modern performance, security, management, and operational requirements [11]. As with the original AFS, AuriStor is designed for the Wide Area Network (WAN) to transparently and uniformly access data regardless of where it is physically or geographically stored with zero local configuration. By providing a global namespace, institutions are able to break down data silos by having an easily deployed mechanism for independent laboratories or departments to expose their data and storage for access and collaboration in a secure, high-performance, and auditable manner.

Data in AuriStor is stored in volumes which are rooted directory trees that may be mounted anywhere in the global namespace. Volumes may correspond to any logical granularity such as the data for a single or group of experiments. AuriStor Volumes are policy containers with fine-grained policy-based administration for physical data location, access controls, redundancy, audit, snapshots, and data replication onto any storage backplane whether in local or remote data centers with Flash, SAN, NAS, in the Cloud, or by L-Store. Policy enforcement also includes data residency restrictions and requirements to address business continuity, data access, and physical locality to achieve compliance to regulatory data residency restrictions.

Authorized users or processes across the WAN and on any platform, device, VM, or container may access data with policy enforcement built into the AuriStor compute fabric, including multi-factor access controls on combinations of user/group, device, and network.

Unlike NFS or CIFS, the AuriStor protocol is highly optimized for the WAN while being resilient to network uncertainties. Volumes may be replicated onto any file server in the TORC cloud to ensure resiliency and reduce latencies. AuriStor clients are available for most platforms (Windows, OSx, Linux, Android, iOS). Clients will automatically access data from the most appropriate server. AuriStor clients maintain a local cache for meta-data and data that is kept coherent with server callbacks, thus reducing the need to access data over the WAN.

The global namespace, security model, and design for use in geographically distributed data-centers over the WAN found in AFS provides the optimal complement to the L-Store Stack in creating a federated

"self-service" storage cloud to be used by the widest audience of users and applications. The AuriStor product development branched off from the OpenAFS open source AFS implementation in 2008 as a commercial product and has diverged significantly since the fork. The decision to have TORC based upon the commercial AuriStor product over the open source OpenAFS was a clear choice based on net financial benefit, functional capabilities and reduced risk. OpenAFS is lacking in some critical areas including: modern security, IPv6 support, and scalability. The AuriStor product addresses all of these issues and others. Since its availability as an enterprise grade commercial product AuriStor is replacing OpenAFS deployments including at a major global bank, major research universities and national security research labs. In addition to the significant improved capabilities of AuriStor, there is concern about the shrinking OpenAFS core contributor base and about availability of support and expeditiousness of security patches from the open source community. Additionally, due to inferior performance and scalability to AuriStor, the free OpenAFS would incur higher hardware and operating costs offsetting the cost of the commercial AuriStor license. TORC is acting as the primary licensor for AuriStor so organizations outside of this grant that wish to stand up additional AuriStor or L-Store servers within TORC network may inexpensively obtain AuriStor licenses from TORC under the bulk TORC license.

1.2.5 Synergies Each of the core technologies proposed solve a different aspect of the data sharing and logistics problem. Logisticial Networking with its generic, best effort, interoperable block-level abstraction IBP provides a solid foundation for storing, moving, and managing data. L-Store builds on this by providing a resilient method of storing and retrieving data using data placement policies determined by the end user. L-Store uses IBPs support of depot-to-depot transfers allowing for bulk data movement. This allows the client to act as a flow control manager instead of the traditional approach of having the data flow directly through the client. AuriStor's sophisticated authentication and authorization components provide system level security via rich support of cross-platform ACLs, delivering consistent access tools independent of the host operating system. AuriStor's ability to replicate data sets using data volumes supporting quotas and max ACLs protects against accidental permission mistakes and provides a powerful framework for creating data workflows. Combining these tools together with TORC significantly lowers the barriers for sharing and moving data, thus enabling more effective scientific research.

1.3 Related Work

Data classification and compartmentalization into silos is the industry best practice to address security, organizational, and compliance issues. This leads independent departments, labs, and organizations to create parallel storage infrastructures with separate hardware, authentication, access, and administration. The result is multiple underutilized silos that must be over-provisioned to meet their individual needs. Each of these systems also incurs the expense of independent administration. Sharing data between these silos is complex and introduces friction to collaboration, becoming even more complex when multiple institutions are involved.

To address this, organizations craft multi-layer solutions at the application, operations, and communication levels and on top of an incoherent set of building blocks, many of which are decades old and never intended for the purpose for which they are being applied. These complex, multi-level solutions are difficult to build, maintain, and manage and result in attack surfaces more difficult to defend. This unnecessary complexity in securing, accessing, and using data affects the ability to exploit collaboration opportunities and handle challenges.

- **GridFTP:** GridFTP is an extension of FTP designed for high performance, reliable transfer of very large files across the WAN. Since GridFTP simply targets file transfer and not the underlying storage, it does not face cross platform compatibility issues regarding storage or access. Because GridFTP does not provide file system semantics, users must create their own scripting, tooling, and protocols for exchanging data in a point-to-point manner.
- **OpenAFS:** OpenAFS is an open source implementation of the Andrew File System. OpenAFS is poorly supported and has significant performance, scale and security limitations at the core architectural level.
- **CIFS/NFS:** The CIFS/SMB or NFS file systems work best for Windows or Linux environments, respectively. They are both designed with an assumption of constant connectivity which is not consistent with multiple data centers over the WAN. Additionally their access control models do not match, introducing dissonance for cross platform collaboration.

- **Lustre:** Lustre is designed for high I/O rates for both single and multiple file access within a cluster. It was not designed for WAN use or cross-platform accessibility, although these can be achieved with CIFS/NFS exporting of the underlying Lustre file system leading to the same issues with CIFS/NFS mounts.
- **IBM's Spectrum Scale (formerly GPFS):** Spectrum Scale is a common parallel file system for HPC environments and also the most full-featured and general purpose solution. It has integrated CIFS/NFS support for WAN access but suffers from similar problems as with CIFS/NFS systems because Spectrum Scale expects constant network connectivity to all systems mounting the file system. Any network oddities or issues will lead to hiccups and service outages.
- **Enterprise File Sync and Share (DropBox, Box, Google Drive):** This class of solutions works well for ad hoc sharing of small data sets but is not architected for efficient access in HPC environments.
- **Commercial Cloud (Amazon AWS, Microsoft Azure, etc):** Cloud services are designed to allow data to easily and cheaply go into the cloud where they stay for analysis. Moving data out of the cloud back to local storage is expensive.

2 TORC Vision

Our vision for TORC is driven by the needs of research communities. The problems they face and that we propose to solve are present at multiple scales. Below, we describe three demonstrative use cases at the campus, metro, and regional scales. We then list the researchers who plan to use TORC once it is deployed - we expect this number to grow substantially during the life of the project and part of our plan of work is to foster this growth. Finally, we describe the services and hardware infrastructure TORC will deploy to meet the needs of these application communities.

2.1 Research Use Cases

2.1.1 Example Campus-Centered Use Case Vanderbilt professor Bennett Landman (Electrical Engineering) demonstrates that TORC is needed to overcome data sharing issues within a campus. Landman is director of the Center for Computational Imaging (CCI). The mission of the CCI is to support collaborative image processing research by: creating infrastructure to interface medical image processing with high performance computing; supporting implementation and standardization of medical image processing; and facilitating collaboration on medical image processing. The CCI uses XNAT, an open source imaging informatics platform, to manage its imaging data. As of August 2016, the CCI XNAT contains 296 projects, 38712 subjects, and 66934 imaging sessions. The XNAT system has completed 238,749 medical image processing jobs on with a total cluster usage time of 1,848,569 hours with more than 70TB of compressed magnetic resonance imaging (MRI), computed tomography (CT), near infrared spectroscopy (NIRS), optical coherence tomography (OCT), and light microscopy imaging modalities. These processing jobs were run on the Vanderbilt Advanced Computing Center for Research and Education (ACCRE), a campus-wide facility. The MRI, CT, etc., data is generated by instruments housed in distributed campus core facilities. Getting this data into XNAT currently requires each piece of information to be transferred individually using HTTP, which is slow and extremely cumbersome.

The CCI will integrate with TORC so that investigators and core facilities can securely and efficiently access their data across campus. The Pathologies of the Human-eye, Orbit, and The Optic Nerve (PHOTON) project will be used as a specific case study to validate the performance of TORC and quantify its benefits to research. PHOTON seeks to map normal and abnormal variability of the optic nerve to identify targets for biomarker design (i.e., where effect size is large relative to normal variation), identify targets for improved imaging (i.e., where imaging variability is relatively high, but there is significant group separation), enable accurate power analyses for subsequent study designs of the optic nerve, and provide evidence to enhance understanding of the processes underlying optic nerve pathology.

This project involves medical (ophthalmology, surgery, and radiology) faculty, residents, and staff and engineering (electrical engineering, biomedical engineering, computer science) faculty, students, and staff who are accessing the data through the Vanderbilt University (VU) and Vanderbilt University Medical Center (VUMC) technology systems.

2.1.2 Example Use Case spanning a Metropolitan Area Advances in massively parallel sequencing technologies (Next-Generation Sequencing, NGS) have profoundly impacted life science research and its broad applications across biology and medicine, but at the same time have raised new big data challenges in the field [12]. NGS technology enables researchers to, within hours, sequence an entire genome or to

obtain genome-wide transcriptome and epigenetic profiles of an organism. Each run on such sequencing platforms can generate up to 1TB of raw sequence data, which then have to be processed and analyzed using downstream computational methods. As the adoption of such technologies increase in the research community, the need for large scale storage and processing of data rises accordingly. NGS projects require collaboration between researchers who collect the samples, the sequencing cores which generate the data, and the bioinformatics teams who perform the analyses. In the Memphis area, there are three sequencing cores located at different institutions: (1) Feinstone Center for Genomics at the University of Memphis; (2) Molecular Resource Center at the University of Tennessee Health Science Center; (3) Hartwell Center for Genomics at St. Jude Childrens Research Hospital. These sequencing cores serve a large number of basic science and applied researchers, many of whom collaborate with bioinformatics groups at other locations. Currently, the NGS data generated at sequencing cores are transported via external hard drives or transferred via the internet from in-house storage devices or commercial cloud storage facilities. These mechanisms are costly and suffer from slow file transfer rates.

To demonstrate the utility of TORC, we have engaged several investigators at the University of Memphis as well as the sequencing core at the University of Tennessee Health Science Center, which provides services for a large number of local researchers (see support letters). Specifically, the following projects will be tested on TORC: **(1)** Insect 5000 Genomes (i5k) Pilot Project (Dr. Duane McKenna, University of Memphis), which is an international effort to sequence and analyze the genomes of 5000 arthropod species [13]; **(2)** 1KITE 1000 insect transcriptome evolution project (Dr. Duane McKenna, University of Memphis) which aims to study the transcriptomes of more than 1,000 insect species encompassing all recognized insect orders. **(3)** Chloroplast sequencing of 192 wild carrot, Daucus carota, species to investigate the effects of heteroplasmy in plants (Dr. Jennifer Mandel, University of Memphis). **(4)** Targeted capture and NGS sequencing of more than 200 taxa across the Compositae (sunflower family) to investigate the patterns of diversification and gene duplication throughout evolution in plants (Dr. Jennifer Mandel, University of Memphis). **(5)** Transcriptome sequencing and Chromatin-Immunoprecipitation NGS sequencing of cell lines in order to investigate the signaling mechanisms during epidermal-to-mesenchymal (EMT) transition with respect to normal development and tumor formation (Drs. Ramin Homayouni and Amy Abell, University of Memphis).

2.1.3 Example Regional Use Case In 2011, the White House announced the establishment of the Materials Genome Initiative (MGI), a a multi-agency initiative ($250M to date) designed to support the development and application of high accuracy validated computational materials methods to enable the discovery, manufacture, and deployment of advanced materials twice as fast, at a fraction of the cost. [14] One of the fundamental tenets of the MGI is the concept of computational screening of materials, in which thousands of candidate materials are computationally evaluated for specific properties. For this work complex molecular simulations are needed on large petascale computing platforms and in some cases performed at multiple levels of detail. These simulations have intricate workflows that experts understand and perform routinely, through a series of steps (building an initial configuration of the system, gathering – and if necessary deriving from *ab initio* calculations – the forcefields which model interactions between atoms and molecules, equilibrating the simulation, performing production runs, and analyzing results) largely performed manually. Supported by three NSF grants [15–17], we are developing an integrated open-source environment, called the Molecular Simulation and Design Framework (MoSDeF), based on the software engineering paradigm of model-integrated computing (MIC) [18] to translate these workflows into a set of tasks that can be automated, combined with other tasks and embedded within MGI screening or targeted design.

Peter Cummings and Clare McCabe and their research groups at Vanderbilt are applying MoSDeF to two complex soft materials systems: modeling nanoscale lubrication systems (needed for MEMS/NEMS and hard disk drives, for example), and understanding/controlling self-assembly in biological structures such as the lipids that provide the skin's barrier function, respectively. In the screening calculations, $10^1 - 10^3$ molecular dynamics (MD) simulations are performed simultaneously with each simulation generating about 25 GB of data with the entire set being on the order of 10 TB. The transfer of this data back to VU and collaborators becomes the bottleneck.

Using the proposed TORC interfaced to the leadership-class Titan in the Oak Ridge Leadership Class Facility (OLCF, see support letter from Jack Wells), the turn-around time for screening calculations will be reduced by orders of magnitude. The apparent immediate availability of data as it is generated will enable "quick and dirty" on-the-fly analyses using short subsets of the trajectories to ensure that the calculations are running as planned, and to give previews of full results. Both Cummings and McCabe have allocations on

Titan large enough to demonstrate the feasibility of TORC-enabled MGI screening for their respective problem areas. Additionally, MoSDeF is configured to work with and create multiple public domain databases (of simulation results, of auto-generated forcefields, and of *ab initio* calculations used in forcefield generation) which will be shared via TORC with collaborators at Oak Ridge National Laboratory and beyond.

2.1.4 Other Initial TORC Use Cases In addition to the three use cases given above, the following researchers will use TORC:

1. Vanderbilt Astronomers Andreas Berlind and Kelly Holley-Bockelmann use ACCRE and XSEDE resources to run massively parallel dark matter simulations.
2. TORC co-PI Abhishek Dubey will use TORC as a part of his work on "Robust Analysis of Dispersed Data Sources For Smart and Connected Communities."
3. Fisk faculty and students will use TORC through the Fisk-Vanderbilt Masters-to-PhD Bridge Program.
4. UTK Civil and Environmental Engineering Professor Joshua Fu seeks to further understanding at the nexus of energy, climate change, and air quality by using computational modeling. He has authored and contributed to reports for the United Nations, IPCC, Department of Energy, and other governmental entities.
5. UTK Materials Scientist David Keffer is working with collaborators at UTK and ORNL to convert data-rich information generated by atomic probe tomography (APT), into a description of the distributions of local structure along spatial and compositional axes.
6. TORC PI Paul Sheldon uses distributed computing resources for his work on CMS.
7. Vanderbilt Astronomer Keivan Stassun needs to collaborate with researchers at Fisk University, Ohio State University, and University of Cape Town on the Kilodegree Extremely Little Telescope (KELT), used to discover and analyze planets orbiting bright, solar-type stars.
8. Tennessee State engineering faculty and students will use TORC in collaboration with Vanderbilt engineering faculty.

2.2 TORC Services

2.2.1 AuriStor AuriStor provides the file system, security, and data volume management for TORC. Client machines/users access files and directories in the TORC Cloud using standard local file system semantics. Free client software can be downloaded from TORC or the AuriStor website regardless of whether the student/researcher has a TORC identity.

- **Global Namespace:** A globally accessible universal file system namespace provides WAN client access using local file semantics. No local configuration of logical to physical mapping is required.
- **Federated Authentication:** Authentication is not necessary for open data, but is required for controlled-data. Authentication into the TORC network can be done via TORC identities. Federated access can be achieved with direct TORC Federation to individual organizations or via Shiboleth single sign-on.
- **Platform Independent:** Native Operating system client file access across Windows, Windows Server, Apple OSx, Apple iOS, Linux (Red Hat, Debian, Fedura, Ubuntu, others on request), Solaris, IBM AIX.
- **Volume Management:** In AuriStor, the unit of management is a 'volume' which is a rooted directory that may be 'mounted' by TORC administrative tooling anywhere in the TORC global namespace. Volumes provide a policy container for security and other functionality. Volumes may be created at a granularity to meet data management requirements (e.g., for an experiment or group of experiments, individual user home directories, or a class web-site).
  - **Volume Replication:** Volumes may be administratively replicated onto multiple file servers within a single data center or across data centers for high availability. For example a volume with an experimental dataset may be replicated onto multiple Tier-1 sites or to a Tier-2 site to eliminate inter-site access latency.
  - **Reliability, High Availability, Performance, Uninterrupted Maintenance:** AuriStor clients automatically use the most appropriate file server for that volume for performance. If that file server becomes unavailable the client will transparently attach to another server without application interruption. Similarly it is possible to take a file server down for maintenance without interrupting long running applications because its volumes may be replicated on the fly onto another server, allowing applications to automatically switch over when the file server is then taken offline.
  - **Read Only Snapshots:** A 'read-only' snapshot of any volume may be created for independent mounting anywhere in the namespace.

  - **Automated Atomic Data Replication:** Read-only snapshots of volumes can be atomically published for replication across the TORC (e.g., to provide application or dataset updates).
  - **Wire Level Security:** Wire Level Security encryption policy is specified on the volume level and enforced by both the file server and clients.
  - **Maximal Access Control:** To prevent users from inadvertently exposing sensitive data, an administrative maximal ACL is placed on every volume that supersedes user settings on files/directories in that volume regardless of where the volume is mounted in the namespace.
  - **Remote Volume Management:** Volume and File Server management can be accomplished remotely across TORC by authenticated administrators, greatly simplifying maintainability.
- **Consistent Access Control Model:** The AuriStor Access Control Model for files and directories is uniform across all platforms and is more robust than NFS or CIFS/SMB. Having a consistent ACL model avoids the inevitable dissonance between NFS and CIFS/SMB access control models that invariably occurs with cross platform data collaboration. Additionally the access control model is also more robust than either NFS or CIFS with read, write, insert, delete, list and ACL modification restrictions.
  - **Multi-Factor Access Control:** Unlike NFS/CIFS which only provides user/group restrictions AuriStor provides multi-factor access controls on combinations of user/group, device/machine, and network.
  - **Universal Anonymous Accessibility:** Data may be published with open accessibility to anyone without their needing a TORC account or identity.
  - **User Defined Groups:** Users may define their own groups without the need for administrators.
- **Auditability:** Auristor provides full auditability on administrative actions, ACL changes, and access.
- **Local Disk Caching:** Unlike other file systems which only maintain in-memory RAM caches, AuriStor clients also maintain a local disk cache which is kept coherent with file server side callbacks (for data, meta-data, and access control). This allows for zero network access local file system performance for recently read files and directories. For example, it is possible to locally configure the AuriStor cache disk to be larger than the data files used in a series of computations. In this case, re-running computations over that data results in locally reading the AuriStor disk cache without additional network traffic. The size of the AuriStor client disk cache is locally reconfigurable by the user and can be set to an appropriate size for their workflow based on the availability of local disk space. It could be configured to be as large as several terabytes if the local desktop computer, workstation, or compute node has sufficient local disk space.
- **Caching vs File Sync and Share:** AuriStor disk caching provides significant advantages over Enterprise File Sync and Share (e.g., Dropbox, Box, or Google Drive) in which the unit of 'caching' is an entire file. With AuriStor, only the parts of the file that are read are brought over the network and into the cache. For example with AuriStor, an application can process data immediately as bytes arrives over the network rather than having to wait for the entire file to be downloaded nor does it requires local storage to mirror all the files on the share. With Sync and Share any change to a file (contents or metadata) forces the entire file to be downloaded again. With AuriStor only the modified portion of a file will be brought over. For example, an application reading from a live log file over the network only requires the tail of the file to be retrieved.

2.2.2 L-Store Services L-Store privides data fault-tolerance and performance scalability in both raw storage and I/O bandwidth.

- **Fast data movement:** ACCRE routinely moves data at 100 Gbps speeds with each storage depot easily sustaining 20Gbps.
- **Data resilience:** L-Stores support of many erasure encoding schemes, with full support of generic Reed-Solomon encodingd, means that each user can tailor the data redundancy as needed.
- **Data Management Policy:** Users can specify how and where their data gets stored. They could choose to stripe their data across multiple drives in the same depot or across multiple depots for better reliability or performance, at a remote site, only on SSDs, or move the data between data centers. All of these options are available using L-Store's rich data policies.
- **Heterogeneous Hardware:** IBP provides block level abstraction to data. This means no complex RAID controllers or SAN arrays are required. Instead, each drive can be exported as an individual

resource with its own arbitrary collection of attributes to facilitate data management and mixing drives of different sizes and types (e.g., spinning disk vs. SSD) is perfectly natural posing no isues.

- **Easy provisioning of storage** Adding storage simply requires setting up a depot and registering with the Resource Service which publishes the new resources to all clients automatically.
- **Add/Remove Drives:** Drives can be added or removed from running IBP servers with the changes automatically propagated to all clients. The IBP server will automatically detect problem drives and eject them from the running server and turn on the fault light making repair trivial.
- **Life cycle management of hardware free (toss if failure):** Using the L-Store management interface, drives and depots can be life cycled/retired and replaced by draining them of data.
- **Rebalancing disk usage:** Having different size drives can lead to smaller drives being completely full among nearly empty large drives. Same goes for adding new storage. These are easily handled using native rebalancing tools so that each drive uses the same percentage of disk space. For example moving 1PB of data can be done in the background in about a week with no impact to end users.
- **Petabyte scale:** Vanderbilt has a decade of production experience at the Petabyte scale withL-Store.
- **Backup and Disaster Recovery Service ready:** While out of funding scope of this proposed CC* Grant, TORC can provide service contract based remote backup/disaster recovery for constituents for any data housed on file servers under TORC, regardless of where it is located.

2.3 TORC Hardware

The TORC hardware will be configured in a spine and leaf configuration with 3 primary backbone (Tier 1 or T1) sites acting as the spine and 6 satellite sites acting as leaves. The T1 sites are connected via high speed (40 or 100 Gbps) networks to facilitate data movement between sites. All storage depots will be deployed partially populated with drives, allowing sites to inexpensively expand capacity by purchasing additional drives with their own funds.

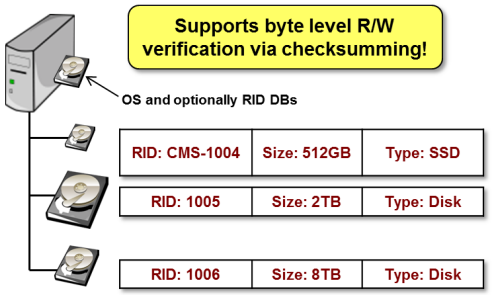

Figure 3: TORC Storage Depot example

The storage depot, shown in Fig. 3, is the fundamental hardware building block for storage using Logistical Networking's IBP for block-level storage. Because of the IBP block-level abstraction, drive sizes and types can be mixed within the same depot. Each drive is assigned a unique resource ID or RID and can be given an arbitrary collection of attributes used to create an unlimited number of storage pools and selection criteria to facilitate storage tiering and data placement. Resources can be added and removed without taking the server down, making lifecycle management and replacement easy. The IBP server can detect dead or failing drives and will automatically eject them from the IBP server process to be handled by an administrator. The IBP server also has a data scrubbing process that runs at regular intervals to detect problematic spots on the drive and alerts the administrator.

Each of the proposed TORC depots will have dual 1/10/40/56 Gbps network interface cards and support up to 36 12Gbps SAS drives. We are proposing each depot be initially configured with 16 8TB drives funded by this project, with the remaining 20 slots available for future expansion. Each T1 site will have a slightly different number of depots tailored to their use case. The Univeristy of Memphis (UM) will have 16 depots (2 PB of native disk space) and the National Institute for Computational Sciences (NICS) will have 20 (2.56 PB). Vanderbilt University (VU) already has 60 depots in use by the CMS experiment, and they will export these depots to TORC by purchasing additional network switches through this project. These depots have available slots and TORC will purchase additional disks to add capacity (1.28 PB). TORC will also purchase 10 new depots (1.28 PB) for Vanderbilt. At the end of the grant the aggregate available native space would be 7.2 PB expandable to over 20 PB. Having many depots at each site makes it possible to satisfy RS-6+3 by striping data across 9 depots allowing any 3 depots to be down for maintenance without a service disruption. A satellite would have a single TORC depot with 128 TB of native space and expandable to 288 TB. The data would be striped across 9 separate drives, again via RS-6+3, to protect against local data loss from drive failures.

In addition to the depots, each T1 site will also have 4 Auristor servers to handle client load and metadata replication and failover. A satellite installation would have a single Auristor/L-Store server. All sites could add more Auristor servers if needed and still be supported under the project at no additional cost.

3 Proposed Work

3.1 Final Design

In the first few months of thie project, we will convene a meeting with interested participants and stakeholders to finalize design details, deployment plans, and coordinate with application groups. This meeting will be hosted at Vanderbilt. We will invite participants from Tennessee State, Fisk, the University of Tennessee Health Science Center, Middle Tennesse State University, Eastern Tennesse State University, the University of Tennessee at Chattanooga, the South Big Data Hub, XSEDE, and DataONE.

3.2 Implementation/Deployment

The hardware for TORC will be deployed over a three year period. Hardware will be shipped to the sites. Local IT support personnel will rack the hardware and Vanderbilt TORC personnel will configure it remotely.

3.3 Software

Vanderbilt and AuriStor are currently implementing a native L-Store plugin, bypassing the existing POSIX file system implmentations, that will be available when TORC is initially deployed. This initial implementation will still have all data and metadata flowing through the AuriStor file server, which is not ideal for HPC usage. This initial implementation will be used during the first year of the grant. At the same time, work will be done completing the integration in order to overcome this limitation. By the end of the grant's first year clients will be able to directly access data on depots, bypassing AuriStor. Auristor will still be responsible for all metadata operations, control cache coherency, and depot access using IBP capabilities, but data will flow directly between the clients and depots, which is much better for HPC environments. Clients that don't support native L-Store integration will fall back to having both data and metadata going through the AuriStor file server.

Additionally during the first year of the grant AuriStor will extend their existing Kerboros authentication system to support Internet2's Shibboleth project enabling single sign-on at all participating sites. This will allow local users to share data with collaborators by just providing their home organization's identifier thereby federating all Internet2 Shibboleth institutions. We will also investigate using Globus/XSEDE identity management capabilities along with InCommon.

3.4 Configuration and Specification of Analytical Workflows

The integration of TORC storage into applications and their associated workflows promises to simplify data services for scientists and engineers, enabling better collaboration and exploration. We will provide a variety of resources to assist researchers with this integration. NICS scientific computing staff will provide consulting support for application groups seeking to tune their codes and workflows. One of the hallmarks of the XSEDE project is its use of Extended Collaborative Support Services (ECSS) to help users to tune their codes and workflows to better use one or more of the Service Providers' HPC systems, often with dramatic improvements in productivity. As key participants in XSEDE and its ECSS, NICS staff have significant expertise and experience with such consulting tasks.

TORC team members at Vanderbilt will also assist researchers with their workflows. by integrating the CHARIOT [22] ecosystem into TORC. An analysis campaign is made up of several types of components: analytical algorithms, data mining steps and data sources. Some of these components can themselves be an assembly of other simpler components. This hierarchical component view can be realized by creating a *Domain Specific Language* that supports the specification of analysis components along different dimensions: functional, execution platform, deployment, and performance, among others and supports their composition to make different campaigns. Atthe top of CHARIOTs stack is a design layer implemented via a generic system description language. This design layer captures system specifications in terms of different kinds of available hardware resources, software applications, and the resource provided/required relationship between them.

3.4.1 Resilient Deployment Deployment and resilient management of remotely executing workflow services requires the ability to reconfigure the analytical workflow upon failures. The bottom of CHARIOTs stack is a management layer that comprises monitoring and deployment infrastructures, as well as a novel management engine that facilitates application (re)configuration as a mechanism to support autonomous resilience. This management engine uses system information stored in the persistent storage to formulate Satisfiability Modulo Theories (SMT) constraints that encode system properties and requirements, enabling the use of SMT solvers (such as Z3) to dynamically compute optimal system (re)configuration at runtime. A

Zookeeper based deployment mechanism is then used to apply the required configuration changes to the system.

3.5 Management

TORC will be managed by a steering committee headed by the lead Principle Investigator and consisting of all of the co-PIs plus three additional members chosen by the PIs from the application and networking communities. The three additional members will be asked to serve for the entire three-year duration of the project. The entire committee will meet face-to-face twice a year and will have at least two additional teleconference meetings. Responsibility for management of all operational elements of the project will fall to the lead PI or his designee. The project will maintain a Web site with statistics on resource availability and usage and that details all important aspects of its development. The main criterion for the success of the project will be the level of use by the research applications. If the storage servers are full of data and there are a lot of transfers in and out of them, then the infrastructure can be said to be well-used. We anticipate that these original application communities will decide to add resources to further support their activity, and that new research groups will decide to contribute resources in order to become partners. The amount of added resources will provide a useful benchmark.

3.5.1 Operations, maintenance and support TORC hardware will be centrally managed by Vanderbilt. Vanderbilt has extensive experience maintaining remote hardware from the previous NSF funded REDDnet grant. To facilitate this all hardware for the grant have incorporated baseboard management controllers for supporting the intelligent platform management interface (IPMI) standard. IPMI supports power cycling hardware remotely as well as running system diagnostics via a dedicated IPMI management port. This will be used in conjunction with the existing command line and web management tools provided with L-Store. The existing L-Store tools support remote disk drive repair including enabling the location LED for the local site administrator to replace a drive along with power cycling the depot and doing firmware upgrades as needed.

3.5.2 System Monitoring AuriStor supports full auditability on all administrative actions, access control list changes and client access. L-Store supports depot disk and network monitoring. This includes data transfers between depots and clients and depot-depot transfers. L-Store's web management interface will be used for monitoring the depot hardware for drive failures and collecting data transfer statistics. It will also be used to perform drive repairs and life-cycling of hardware.

3.5.3 Project Milestones and Deliverables Table 1 lists our project milestones and deliverables.

4 Broader Impacts

TORC will collaborate with the South Big Data Hub consortium, XSEDE, DataONE, and the Fisk-Vanderbilt Masters-to-PhD Bridge Program (FVBP) to broaden its reach and impact.

The Fisk-Vanderbilt Masters-to-PhD Bridge Program (FVBP) is committed to increasing the number of underrepresented minorities earning PhDs in astronomy, biology, chemistry, materials science and physics. After 12 years and 110 students, there have been 55 Masters degrees awarded, 43 in Physics, 9 in Biology, and 3 in Chemistry. To date, there are 30 students in the Masters phase with 11 graduating this summer and 12 more joining in Fall 2016. The students are a diverse group: 59% are African-American, 23% Hispanic, 4% Native American or Pacific Islander, and 14% are white or other non-minority. Approximately half of the students are women. Most are from traditionally underserved populations, first generation, low-income, or have physical or learning disabilities. With an overall retention rate of 95% and a PhD retention rate of 87%, Fisk has achieved the distinction of being the top awarder of Masters degrees in physics to African Americans, and Vanderbilt is one of the top awarders of PhDs to URMs in astronomy, physics, and materials science. Twenty-one students have earned PhDs, and 100% of them are employed as faculty, postdocs, staff scientists, or scientists in industry.

Vanderbilt professor Kelly Holley-Bockelmann heads this program and will work with TORC to integrate its students into TORC research and work. The FVBP is already working with XSEDE to increase the number of underrepresented minority XSEDE users – XSEDE hosted the first Regional XSEDE Users Workshop at Vanderbilt, and returned in 2015. Through those efforts, the number of Bridge students doing computational work has increased and there's an emerging need for data storage, one that's not available at all at Fisk. TORC will give them accounts, access, and hands-on tutorials on how to use the facility through this proposal.

Table 1: Milestones and Deliverables

Year 1

1.1 Host Final Design Meeting at Vanderbilt.
1.2 Deploy year 1 depot and server hardware at T1 sites.
1.3 Make exisiting ACCRE depots available to TORC by adding necessary network hardware.
1.4 Characterize and tune networks between T1 sites.
1.5 Setup L-Store servers at each T1 and configure them to use their local depots.
1.6 Setup AuriStor file and database servers at T1 sites.
1.7 Configure T1 AuriStor servers to use the native L-Store plugin.
1.8 Perform transfer benchmarks using the full software stack to diagnose and correct bottlenecks.
1.9 Establish tools to connect TORC and third party systems (such as those at ORNL).
1.10 Establish requirements for and design researcher workflows using CHARIOT.
1.11 Start using TORC for production work.

Year 2

2.1 Implement and evaluate CHARIOT workflows for an initial set of researcher use cases.
2.2 Select the initial 3 satellite sites and provision hardware and assist with initial configuration.
2.3 Expand T1 sites with additional year 2 hardware.
2.4 Mount TORC on ACCREs compute cluster.
2.5 Contribute to the South Big Data Hub infrastructure working group.
2.6 Start testing of AuriStor direct client-depot transfers in HPC environments.
2.7 Work with South Big Data Hub to develop additional TORC user communities.

Year 3

3.1 Refine CHARIOT workflows and roll out workflows for all use cases.
3.2 Select the final 3 satellite sites and provision hardware and assist with initial configuration.
3.3 Expand T1 sites with additional year 3 hardware.
3.4 Contribute to the South Big Data Hub infrastructure working group.
3.5 Work with South Big Data Hub to develop additional TORC user communities.

The South Big Data Hub is part of a network of four regional Big Data Hubs, launched by the National Science Foundation. The South Hub serves 16 states and the District of Columbia from Texas to Delaware with more than 500 members from universities, corporations, foundations, and cities committing their support. TORCs highly scalable and extensible design will enable it to serve South BD Hub consortium members, who would immediately be able to access the facility via the Auristor AFS client software (which is available free). Members who wanted fully participate for even more enhanced access and performance could stand up their own storage and AFS servers and become "nodes" in the TORC cloud. The South BD Hub will help with the "matchmaking" process and to assist Hub member institutions to join and/or use TORC resources. This TORC/South BD Hub collaboration will dramatically improve the ability of researchers in the South BD Hub consortium to share data and compute resources. TORC personnel will join and contribute to the South BD Hub infrastructure working group and con-tribute to other activities and working groups of the Hub.

The Data Observation Network for Earth (DataONE) seeks to aid researchers in preserving, accessing, using, and reusing data from disparate scales, disciplines, and organizations to further scientific inquiry. A distributed network of data centers can be accessed as Member Nodes, with Coordinating Nodes providing associated data catalogs. An Investigator Toolkit can then be used by researchers to manage all aspects of the data life cycle in conjunction with these data resources. As part of the DataONE leadership team and as a member of the TORC project team, Suzie Allard brings highly relevant experience to help leverage the proposed TORC cyberinfrastructure.

5 Results from Prior Support

Peter Cummings: OCI-1047828 $2,594.000, 10/1/2011-9/30/2017 "Collaborative Research: SI2-SSI: Development of an Integrated Molecular Design Environment for Lubrication Systems (iMoDELS)." This project focuses on developing, deploying and distributing the Integrated Molecular Design Environment for Lubrication Systems (iMoDELS), an open-source simulation and design environment that encapsulates the expertise of specialists in first principles, forcefields and molecular simulation related to nanoscale lubrication

in a simple web-based interface. We have developed a suite of tools that enable the efficient execution of molecular dynamics simulations of interfacial systems relevant to lubrication on various platforms and using a variety of open-source codes. **Itellectual merit**: The accomplished research developed a meta-language description of molecular simulation that can be instanced into syntactically correct simulations using a number of public domain simulation codes, opening up the world of complex molecular simulations to a very broad audience. **Broader impacts**: This project included the training of undergraduate and graduate students from Vanderbilt and beyond (e.g., REU participants). These tools have also been used in classroom contexts to teach molecular simulation. Ultimately, our computational environment for molecular simulations will enable greater impact of molecular simulation in solving societal problems that ultimately relate to molecular phenomena, including many problems in energy, biology, medicine and the environment. This project, along with collaborating project OCI-1047857 at Penn State University, has resulted thus far in 7 refereed journal publications [21, 23–28], 2 refereed conference papers [20, 29], 8 conference presentations [30–37] (including a keynote presentation at a leading computational science conference [33] and three invited conference presentations [34, 35, 38]), and one invited seminar [39].

Abhishek Dubey has been supported by the CPS award (CNS-1329803, $399,951, 10/13-09/16), entitled Diagnostics and Prognostics Using Temporal Causal Models for Cyber Physical Systems—A Case of Smart Electric Grid is looking into model-based diagnostics and prognostic techniques to improve the effectiveness of isolating failures in large systems. **Intellectual Merit**: The work completed so far is focused on developing the theory behind modeling failure cascades using temporal causal diagrams. The approach requires the identification of impending failure propagations and determining the time to critical failures that will increase system reliability and reduce the losses accrued due to failures. Key papers were published in reputable conferences and journals, such as Prognostics and Health Management [40, 41] and IEEE Autotestcon [42] and IEEE Instrumentation [43]. **Broader Impact**: The modeling approach of this project, after being validated on more complex examples, can become a standard paradigm for modeling failure dynamics in CPS, where failure modes and their impact on the controller are explicitly modeled.

Ramin Homayouni has not had any NSF funding in the last five years.

Greg Peterson is co-PI and Director of Operations for the $125M NSF ACI-1053575, "XSEDE: eXtreme Science and Engineering Discovery Environment" project. **Intellectual merit**: To date, the XSEDE project has created and operated the most advanced digital cyberinfrastructure in the world, supported with an expert and experienced team of CI professionals, which has enabled researchers across the nation to conduct transformational research efforts in science, engineering, and the humanities. Of particular note, XSEDE operational excellence resulted in the project having no security incidents, a significant improvement over the TeraGrid project, handled around 1,000 tickets per month, deployed central services with over 99% availability, and demonstrated the potential for employing Software Defined Networking to support data sharing across HPC centers. **Broader impact**: XSEDE has performed training, education, and outreach for the scientific community, reaching thousands of students and faculty at hundreds of institutions. Publications resulting from the XSEDE project to date include more than 14,000 publications from users supported by XSEDE and dozens of staff publications (listed in XSEDEs quarterly reports), including [44]. XSEDE reports are available from the XSEDE website [45], and more recent user and staff publications are available from the user portal [46].

Paul Sheldon: NSF ACI-1541443, $500,000, 8/5/2015; entitled "CC*DNI Networking Infrastructure: Enabling International Data Intensive Scientific Collaboration for Vanderbilt Researchers." This project is acquiring the hardware necessary to upgrade the Vanderbilt external network connection to 100 gigabits per second and upgrading the managed circuit between the Vanderbilt campus and Southern Crossroads point of presence in Atlanta. **Intellectual Merit:** This project is fostering and enabling national research programs in a wide variety of disciplines. Vanderbilt University is home to a highly visible research community, ranked 21st in federal research funding. This community is engaged in research with globally distributed collaborators. Improved Vanderbilt network connectivity enhances discovery in their respective disciplines. No publications have yet been produced under this award. **Broader Impacts:** The Vanderbilt research community provides immersive experiences for undergraduates. This includes capstone projects in undergraduate courses and a summer research program hosted by ACCRE. By fostering international, collaborative data intensive research this project is enhancing these experiences.

References

[1] Kyle Chard, Simon Caton, Omer Rana, and Daniel S Katz. A social content delivery network for scientific cooperation: Vision, design, and architecture. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1058–1067. IEEE, 2012.

[2] Data logitstics toolkit. `http://data-logistics.org`.

[3] M. Beck, T. Moore, J. S. Plank, and M. Swany. Logistical Networking: Sharing More Than the Wires. In C. A. Lee S. Hariri and C. S. Raghavendra, editors, *Active Middleware Services*, volume 583. Kluwer Academic, Norwell, MA, 2000.

[4] D Martin Swany and Rich Wolski. Data logistics in network computing: The logistical session layer. In *Network Computing and Applications, 2001. NCA 2001. IEEE International Symposium on*, pages 174–185. IEEE, 2001.

[5] J. S. Plank, A. Bassi, M. Beck, T. Moore, D. M. Swany, and R. Wolski. Managing data storage in the network. *IEEE Internet Computing*, 5(5):50–58, September/October 2001.

[6] Alessandro Bassi, Micah Beck, Terry Moore, James S. Plank, Martin Swany, Rich Wolski, and Graham Fagg. The Internet Backplane Protocol: a study in resource sharing. *Future Generation Computer Systems*, 19(4):551 – 561, 2003.

[7] Micah Beck, Terry Moore, and James S. Plank. An End-to-end Approach to Globally Scalable Network Storage. In *Proceedings of ACM SIGCOMM workshop on Future Directions in Network Architecture*, pages 339–346, 2002.

[8] David D Clark. Interoperation, Open Interfaces, and Protocol Architecture. In Applications National Research Council NII Steering Committee Commission on Physical Sciences, Mathematics, editor, *The Unpredictable Certainty:White Papers*. The National Academies Press, Washington, DC, 1997.

[9] J. S. Plank, M. Beck, W. Elwasif, T. Moore, M. Swany, and R. Wolski. The Internet Backplane Protocol: Storage in the network. In *NetStore'99: Network Storage Symposium*. Internet2, `http://dsi.internet2.edu/netstore99`, October 1999.

[10] A. Bassi, M. Beck, and T. Moore. Mobile management of network files. In *Third Annual International Workshop on Active Middleware Services*, pages 106 – 114, August 2001.

[11] YFS, a high performance global file system that is backward compatible with AFS. `https://www.sbir.gov/sbirsearch/detail/354030`. Accessed: 2016-08-23.

[12] Devesh Kumar Srivastava. Big challenges in big data research. *Data Mining and Knowledge Engineering*, 6(7):282–286, 2014.

[13] Bernhard Misof, Shanlin Liu, Karen Meusemann, Ralph S Peters, Alexander Donath, Christoph Mayer, Paul B Frandsen, Jessica Ware, Tomáš Flouri, Rolf G Beutel, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210):763–767, 2014.

[14] National Science and Technology Council. Materials Genome Initiative for Global Competitiveness. Technical report, National Science and Technology Council, 2011.

[15] Development of an integrated molecular design environment for lubrication systems (imodels). `http://tinyurl.com/z6g7yyq`. Accessed: 2016-08-12.

[16] Cyber-enabled design of functional nanomaterials. `http://tinyurl.com/hwrck2t`. Accessed: 2016-08-12.

[17] Development of a software framework for formalizing forcefield atom-typing for molecular simulation. `http://tinyurl.com/jzxkgp9`. Accessed: 2016-08-12.

[18] J Sztipanovits and G Karsai. Model-integrated computing. *Computer*, 30(4):110–111, April 1997.

[19] Janos Sallai, Gergely Varga, Sara Toth, Christopher Iacovella, Christoph Klein, Clare McCabe, Akos Ledeczi, and Peter T. Cummings. Web- and Cloud-based Software Infrastructure for Materials Design. *Procedia Computer Science*, 29:2034–2044, 2014.

[20] Janos Sallai, Gergely Varga, Sara Toth, Christopher Iacovella, Christoph Klein, Clare McCabe, Akos Ledeczi, and Peter T. Cummings. Web- and cloud-based software infrastructure for materials design. *Procedia Computer Science*, 29(0):2034 – 2044, 2014. 2014 International Conference on Computational Science.

[21] Christoph Klein, János Sallai, Trevor J. Jones, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. *A Hierarchical, Component Based Approach to Screening Properties of Soft Matter*, pages 79–92. Springer Singapore, Singapore, 2016.

[22] Subhav Pradhan, Abhishek Dubey, Aniruddha Gokhale, and Martin Lehofer. CHARIOT: A Domain Specific Language for Extensible Cyber-Physical Systems. In *The 15th Workshop on Domain-Specific Modeling*, Pittsburgh, Pennsylvania, United States, october 2015.

[23] Pedro Morgado, Jana Black, J. Ben Lewis, Christopher R. Iacovella, Clare McCabe, Luís F.G. Martins, and Eduardo J.M. Filipe. Viscosity of liquid systems involving hydrogenated and fluorinated substances: Liquid mixtures of (hexane and perfluorohexane). *Fluid Phase Equilibria*, 358(0):161 – 165, 2013.

[24] Henrik R. Larsson, Adri C. T. van Duin, and Bernd Hartke. Global optimization of parameters in the reactive force field reaxff for sioh. *Journal of Computational Chemistry*, 34(25):2178–2189, 2013.

[25] Kaushik L. Joshi, Sumathy Raman, and Adri C. T. van Duin. Connectivity-based parallel replica dynamics for chemically reactive systems: From femtoseconds to microseconds. *The Journal of Physical Chemistry Letters*, 4(21):3792–3797, 2013.

[26] Christoph Klein, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. Tunable transition from hydration to monomer-supported lubrication in zwitterionic monolayers revealed by molecular dynamics simulation. *Soft Matter*, 11:3340–3346, 2015.

[27] Jana E. Black, Christopher R. Iacovella, Peter T. Cummings, and Clare McCabe. Molecular dynamics study of alkylsilane monolayers on realistic amorphous silica surfaces. *Langmuir*, 31(10):3086–3093, 2015. PMID: 25720502.

[28] Andrew Z. Summers, Christopher R. Iacovella, Matthew R. Billingsley, Steven T. Arnold, Peter T. Cummings, and Clare McCabe. Influence of surface morphology on the shear-induced wear of alkylsilane monolayers: Molecular dynamics study. *Langmuir*, 32(10):2348–2359, 2016. PMID: 26885941.

[29] Gergely Varga, Janos Sallai, Akos Ledeczi, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. Enabling cross-domain collaboration in molecular dynamics workflows. In *The Fourth International Conference on Advanced Collaborative Networks, Systems and Applications (COLLA 2014)*, Seville, Spain, 06/2014 2014. IARIA, IARIA.

[30] J. Yeon and A. C. T. van Duin. A reactive molecular dynamics research for the effect of strain energy on hydrolysis reaction of the silica-water interface. AIChE Annual Meeting, San Francisco, November 3-8, 2013, 2013.

[31] C. Junkermeier and A. C. T. van Duin. Using reactive force fields to model adatom domains in fluorinated graphenes. AIChE Annual Meeting, San Francisco, November 3-8, 2013, 2013.

[32] Jana E. Black, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. Reactive molecular dynamics study of alkylsilane monolayers on realistic amorphous silica substrates. AIChE Annual Meeting, San Francisco, November 3-8, 2013, 2013.

[33] P. T. Cummings, C. R. Iacovella, C. McCabe, A. Ledeczi, and G. Karsai. Automating computational materials discovery through model-integrated computing. Keynote Lecture, International Conference on Computational Science (ICCS) 2014, Cairns, Queensland, Australia, 2014.

[34] Adri C. T. van Duin, Murali Raju, Sriram Srinivasan, Jejoon Yeon, Sung-Yup Kim, and Jim Kubicki. Reaxff studies on reactions at the metal oxide surface/water interface: strain/reactivity relations and crystal growth. ACS Spring 2013 meeting (New Orleans), 2013.

[35] Adri C. T. van Duin, Murali Raju, Sriram Srinivasan, Jejoon Yeon, Sung-Yup Kim, Thomas Senftle, and Kaushik L. Joshi. Aps spring 2013 meeting (baltimore). ReaxFF-based molecular dynamics studies on reactions at complex material surfaces, 2013.

[36] Gergely Varga, Sara Toth, Christopher R. Iacovella, Janos Sallai, Peter Volgyesi, Akos Ledeczi, and Peter T. Cummings. Web-based metaprogrammable frontend for molecular dynamics simulations. In *3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH)*, Reykjavik, Iceland, 07/2013 2013.

[37] Christopher R. Iacovella, Hugh Docherty, Matthew A Gebbie, Markus Valtiner, Xavier Banquy, Jacob N. Israelachvili, and Peter T. Cummings. The structural ordering of ionic liquids nanoconfined between charged walls. American Institute of Chemical Engineers Annual Meeting, Pittsburgh, PA, November 2012.

[38] Christopher R. Iacovella, Gergely Varga, Janos Sallai, Clare McCabe, and Peter T. Cummings. Application of concepts from modeling integrated computing for the design of soft materials. The Minerals, Metals and Materials Society (TMS) Annual Meeting and Exhibition, Orlando FL, 2015.

[39] C. R. Iacovella. Improved computational models via synthesis mimetic simulation: Applications to nano confined systems. I3MS Seminar Series, Aachen Institute for Advanced Study in Computational Engineering Science, RWTH Aachen, Germany, January 13, 2014, 2014.

[40] Nagabhushan Mahadevan, Abhishek Dubey, Gabor Karsai, Anurag Srivastava, and Chen-Ching Liu. Temporal causal diagrams for diagnosing failures in cyber-physical systems. *Annual Conference of the Prognostics and Health Management Society*, 2014.

[41] Saideep Nannapaneni, Abhishek Dubey, Sherif Abdelwahed, Sankaran Mahadevan, and Sandeep Neema. A model-based approach for reliability assessment in component-based systems. *Annual Conference of the Prognostics and Health Management Society*, 2014.

[42] Nagabhushan Mahadevan, Abhishek Dubey, Huangcheng Guo, and Gabor Karsai. Using temporal causal models to isolate failures in power system protection devices. In *AUTOTESTCON, 2014 IEEE*, pages 270–279. IEEE, 2014.

[43] N. Mahadevan, A. Dubey, A. Chhokra, H. Guo, and G. Karsai. Using temporal causal models to isolate failures in power system protection devices. *Instrumentation Measurement Magazine, IEEE*, 18(4):28–39, 2015.

[44] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkens-Diehr. Xsede: Accelerating scientific discovery. *Computing in Science and Engineering*, 16(5):62–74, 2014.

[45] XSEDE project documents. `http://www.xsede.org/web/guest/project-documents-archive`. Accessed: 2016-08-23.

[46] XSEDE publication portal. `http://portal.xsede.org/publications/`. Accessed: 2016-08-23.