NSF Cyberinfrastructure Facilities Whitepaper

S. Berukoff, E. Cross

Daniel K. Inouye Solar Telescope

National Solar Observatory

**Introduction**

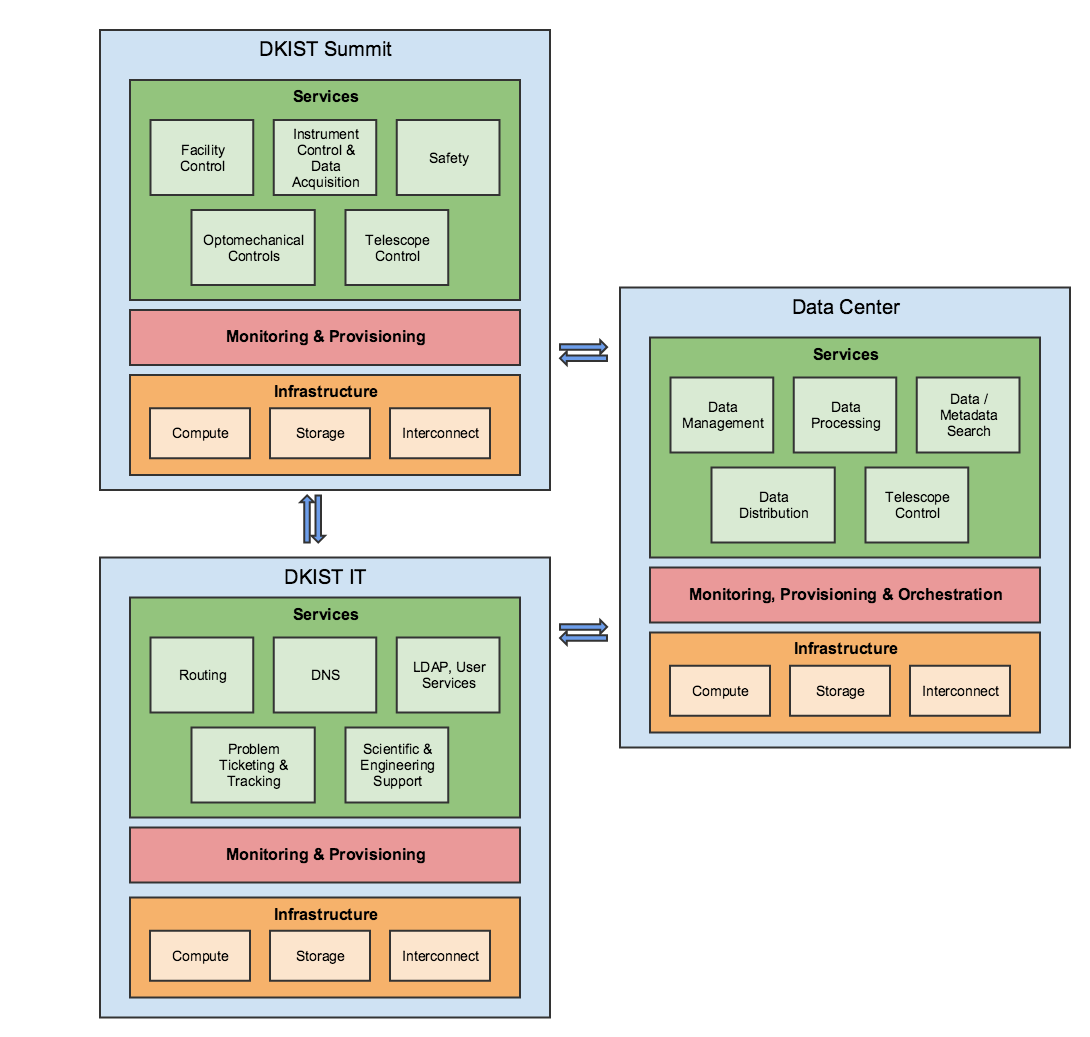
The Daniel K. Inouye Solar Telescope (DKIST) (<http://dkist.nso.edu)> is a four-meter, off-axis Gregorian solar telescope currently under construction by the National Solar Observatory and AURA on Haleakala, Maui, Hawai’i. When complete in 2019, it will be the largest solar telescope in the world, providing facility-class, high-resolution solar observations to a small but growing community of students, researchers, and the general public. In full operations, planned to last fifty years, the DKIST will house five complex instruments and a state-of-the-art adaptive optics system, generating over three petabytes of raw data annually. Key to its success, then, is a cyberinfrastructure providing facility and instrument control, scientific and operational data acquisition, and data management, processing, and distribution services. In this whitepaper, we provide a high-level description of primary components of the cyberinfrastructure.

**Cyberinfrastructure**

The DKIST cyberinfrastructure is comprised of three primary components: the systems and infrastructure providing services to operate the telescope and its supporting subsystems (“Summit”), the core services and infrastructure needed to support science and engineering activities related to observatory operations and network services (“DKIST IT”), and the services and infrastructure performing long-term data management, processing, discovery, and distribution (“Data Center”). These components are highlighted in Figure 1, and discussed in more detail below.

*Summit*

The DKIST Summit cyberinfrastructure comprises integrated facility, instrument control and safety systems, enabling telescope and dome control, optical alignment and routing, mechanical controls, observation execution and monitoring, instrument data acquisition, management, and distribution, and environmental monitoring and control. These systems are comprised of a High Level Software suite written primarily in Java and Python, utilizing CORBA. They are deployed through configuration-controlled provisioning stacks, including SaltStack, and sit atop an HPC architecture comprising many dedicated nodes interconnected through 10 Gb Ethernet and FDR InfiniBand. The Summit cyberinfrastructure is currently being readied for integration testing as a prelude to observatory integration efforts coming in the next 12-18 months.

Figure 1: DKIST Cyberinfrastructure

*DKIST IT*

The DKIST IT supports the observatory through deployment of core services such as routing, DNS, LDAP, and network maintenance and monitoring for the summit and a remote support building, as well ensuring SLAs and/or contracts with partner organizations (U. Hawai’I in Maui and U. Colorado in Boulder at the NSO Headquarters) are met and maintained. In addition, the DKIST IT provides operational support for physical infrastructure (optical fiber, Ethernet and InfiniBand networking, and routing hardware) on the Summit and the remote support building. Services are deployed through configuration-controlled provisioning stacks, sitting atop commodity equipment including Cisco switching. The DKIST IT is ramping its efforts, particularly with regard to network buildout on the Summit and the remote support facility.

*Data Center*

The DKIST Data Center will provide long-term data management, scientific processing, search, and distribution services for the observatory. It will manage 3.2 PB of data per year, comprised of hundreds of millions of observations and tens of billions of metadata, exported by the Summit and, after calibration, intended for end-user consumption. Thus, data management and processing services must scale effectively with little rework, while data search depends on appropriate data modeling and well-developed use cases to allow end-users to effectively target data of interest. Key aspects of the architecture include a combined microservices and virtual machine deployment, provisioned through SaltStack and managed with Elastic and related tooling. While it is planned for the Data Center to reside at the NSO Headquarters, economies of scale are shifting, indicating a need to ensure “deploy-anywhere” (e.g., commercial cloud providers) can be supported effectively. The Data Center is currently completing its design phase, with development expected to occur in 2018-2020, with phased delivery of critical services occurring as DKIST comes online.

When combined with a rigorous systems-engineering approach, including detailed requirements and interface controls, these three primary components will support DKIST use and scientific data exploitation. Despite the bespoke nature of the Summit CI, there is a significant focus on leveraging open source technologies in the DKIST, rather than relying on integration of commercial products. This is partly due to the long-term nature of the program and tight budgetary constraints. However, there are no free lunches – significant open source adoption without proactive forward replacement planning can leave obsolesced components underpinning critical systems. Given the long development timeline for the DKIST – the first CI work began in 2005 – these issues are already creeping into a yet-to-operate facility. Yet, the state of system development shows significant progress forward, and a bright future, for the DKIST CI.

**Summary**

This whitepaper briefly discusses the DKIST end-to-end cyberinfrastructure, focusing on the three primary entities and their roles. Each is in a different developmental state, emphasizing the importance of clear requirements and interfaces, effective team communication strategies, and stakeholder management.