Towards a more data-centric operation

A white paper, by S. H. Kwok, 2022-01-02

The Keck Observatory has been transitioning from operating mostly in the classical observing mode to a mixture of classical and service-oriented observing modes. This trend will most likely continue in the next decade. While in the past, the only data products were the raw data that was collected from the instruments, nowadays observers demand data products that are more structured, processed, and annotated with meta-data. The demand will likely increase in form of higher data quality, more meta-data, and more reduced data. This paradigm shift has profound implications for the Observatory and at the same time offers new opportunities.

With foresight, the Observatory has collaborated with NASA to create the Keck Observatory Archive (KOA) and worked with other partners to implement data reduction tools (DRPs) for all current and new instruments. The KOA and the DRPs are being integrated into operations and in five years, they are expected to be part of routine operations. What other data-centric services can the Observatory provide that can help experienced and novice observers alike?

Data-centric services comprise software applications and tools, computing resources, and data warehousing, as well as interfacing with services from third parties. Data services perform tasks on behalf of the observers or users in a more efficient and standardized way than what observers can do individually. As the amount and the complexity of data increase, data services provided by the Observatory may be the only available resource for many observers to obtain scientific results.

Besides KOA and DRPs, the Observatory has developed standalone tools that can be seen as precursors for full-fledged data services. As the fields of computing and software advance, it is predictable that observers will expect more elaborate and integrated data services in the next decade. Here are a few possible data services:

- An integrated observing planning service that includes a history of previous observations and results, as well as possible theoretical results and simulations. This can lead to the elimination of unnecessary observations.
- Near real-time data quality assessment through continuous monitoring of the data collected by the instruments during observations. Implementation of error detection and anomaly detection.
- Automated in- and post-observing data reduction and analysis to determine the quality of the data and validate the results. Remote display and advanced visualization of raw and processed data.
- Computing resources and storage (the Keck cloud) for observers to run their self-developed/customized DPR, the default DPR, or any tool provided by the Observatory.
- Curation of existing data and creation of new data sets using existing data (virtual observing). Extraction of specific features from existing data (virtual survey).
- Integration of Keck data services into the worldwide network of astronomical data services to allow for example coordinated observing scheduling (TOO), multi-messenger astronomy, and data blending using various sources from different observatories.

Prerequisites to successfully implement and support data services are a healthy computing platform and a strong software team. While it is hard to predict what the general computing landscape will look like in the future, it is safe to say that a good amount of legacy software will remain until they are replaced or upgraded. Hence, the future computing platform will feature a mix of legacy systems and state-of-the-art systems. These systems are isolated from those systems used in operations. Cloud computing seems to be a good candidate for future data services. The Observatory can also provide technologies, such as virtualization and containerization, to help bridge the gap between the old and the new. This obviates the need for observers to set up and maintain these tools.

A strong software team is essential for the realization of data services. However, it does not mean that all the software is required to be developed in-house. On the contrary, collaboration with partner institutions will be crucial for the future of the Observatory. Here are a few ideas on how to achieve a strong software team:

- Partner with one or more universities and work with their software or astronomy departments to solve existing and new software problems. This can be done through internships, periodic lectures, or summer programs. This will not only educate the next generation of software engineers and astronomers but also the Observatory will benefit from advances in academic research.
- Implement regular hands-on workshops with software engineers and astronomers to share their experiences and invite the community to participate.
- Invest in testing equipment such as simulators and encourage the team to experiment and prototype new ideas.
- Institute task or role rotation to promote knowledge exchange and cross-training.
- Implement a safe working environment to enable change.

Building a strong software team takes a long time, requires a culture of self-improvement, and demands the participation of everyone.