# Frameworks Coordinate Scientific Data Management

NASA Technology

Voyager 2 sailing beyond the far boundary of the solar system. The rover Opportunity churning across the red soil of Mars. Cassini-Huygens imaging the moons of Saturn.

Capable of journeying well beyond the reach of human explorers, NASA's robotic missions have probed the distant reaches of space, sending back to Earth streams of unique data and images essential to developing an understanding of our universe. These returns are ultimately housed in NASA's Planetary Data System (PDS), an archive of data products derived from NASA's robotic missions, from Galileo to Pioneer to Stardust and more. Appropriately massive for the information it contains, the PDS is distributed across the Nation and organized in eight nodes in conjunction with a host of NASA partner institutions.

To help researchers draw the information they need from the ever-growing repositories of the PDS, in 1998 Daniel Crichton, program manager and principal computer scientist at NASA's Jet Propulsion Laboratory, designed a unique software framework called the Object Oriented Data Technology (OODT) that transformed the PDS into an accessible virtual knowledge system. "The idea of OODT was to be able to capture all the data, the history of the data, and be able to tie and link all that together into an integrated but distributed system," says Crichton.

OODT primarily functions as a set of building blocks for constructing systems that capture and manage complex parcels of scientific data, Crichton explains. Its cumulative power allows users to connect multiple, distributed databases and other data sources and then to search for and pull together information in varied data formats, building and populating databases with the aggregated results. During the software's development, Crichton was careful to separate software architecture from data architecture, meaning OODT functions as a general-use tool that can
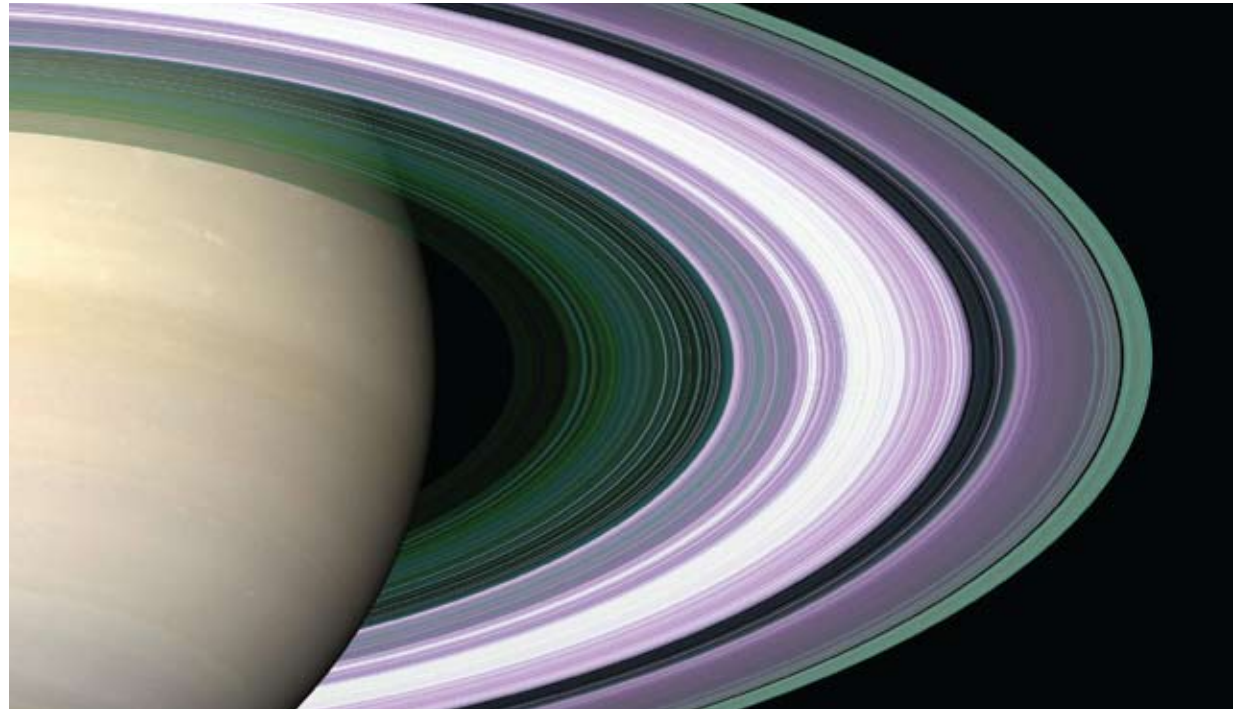
Data from NASA's many spacefaring missions, like this image of Saturn from the Cassini-Huygens spacecraft, is stored in the Planetary Data System, a massive archive of scientific information distributed across the Nation.

plug into existing systems and be tailored and extended for their data. In addition to the PDS, NASA also uses OODT for multiple Earth science missions.

Partnership

While developing OODT, Crichton was already thinking about applications for the software beyond NASA's missions.

"We saw the unification and integration of science data as a real national need," he says. Crichton and his colleagues looked into ways of better engaging the open-source software community to transfer the benefits of NASA software innovations to the public. With this in mind, Chris Mattmann, a senior computer scientist at

JPL who worked with Crichton on OODT, cultivated connections at the Apache Software Foundation (ASF), based in Forest Hill, Maryland. An all-volunteer, nonprofit organization supported by major information technology companies like Google, Microsoft, and Yahoo!, the ASF manages almost 150 open-source software projects, including the Apache HTTP Server—a key technology in the development of the World Wide Web and the world's most widely used Web server—and other popular developer software. Mattmann believed Apache was the ideal partner for transferring OODT for public use.

"Apache is different from other open-source communities," he says. The organization follows a unique vetting

> "The more we can share software, the more benefit we're going to see in our scientific community."
> —Daniel Crichton,
> Jet Propulsion Laboratory

process, he explains, that includes an incubation period to ensure that the candidate software is not only sound, but is also supported by a diverse community that will grow the software. It also provides infrastructure and leadership for housing, distributing, and managing the continued development of the technology.

"The ASF has been well known as having the 'secret sauce' for how to create successful, long-term, healthy open-source projects," says ASF president Jim Jagielski. "We worry about the mailing lists, infrastructure, resources, and fundraising, and the projects can focus on what they do best, which is building great code and great communities."

The ASF placed OODT in the Apache incubator program, a 1-year process during which Mattmann, Apache mentors, and committed collaborators from institutions as diverse as AOL, the University of Southern California, and Children's Hospital Los Angeles thoroughly reviewed the software for open-source use and expanded OODT's community. In November 2010, OODT graduated as an Apache Top-Level Project—the first NASA-developed software to gain the distinction—and now benefits from the full resources of the ASF, including an Apache Project Management Committee that guides day-to-day operations, product releases, and community development for the technology.

Benefits

Apache OODT version 0.1 is currently available for public use under the Apache Software License. The software is generating significant worldwide interest and

NASA's robotic missions, such as the Mars Reconnaissance Orbiter pictured here in an artist's rendering, beam massive amounts of data back to Earth for scientific review. The challenge of accessing that data effectively led to the development of the Object Oriented Data Technology software now in use by hospitals and medical researchers through the Apache Software Foundation.

contributions while supporting a number of research networks outside of NASA. "If you look at planetary science, Earth science, even cancer research, there's actually a lot of consistency or similarity in the kinds of software capabilities needed," Crichton says.

Even before the release of version 0.1, OODT had found users who have employed the NASA-developed software to forward medical research. The National Cancer Institute uses OODT as the foundation of its Early Detection Research Network, unifying multiple laboratories to capture and share research into the early detection of cancer biomarkers. Children's Hospital Los Angeles is establishing a virtual infrastructure for joining pediatric intensive care units across the country, allowing doctors to examine the outcomes of various interventions and make better informed treatment decisions.

These organizations and others will benefit as OODT improves as a result of open-source community contributions, says Mattmann, who is also vice president of Apache OODT. "The dissemination of information happens out in the clear, where others can contribute and weigh in. We get lots of feedback," he says.

Sharing innovative technology with the public is an impulse that is common to both NASA and the ASF, according to Jagielski, a former Goddard Space Flight Center engineer of 19 years. "It's all about developing technology that you can then distribute to whoever needs it for the public good," he says. "It's a win-win for everyone, because from the taxpayer money that was invested, we now have many groups that are able to use this technology," says Crichton. He says Apache OODT is already benefitting from the Apache partnership and the contributions of open-source developers. These contributions, he notes, will pay dividends for scientific research in the future.

"The more we can share software, the more benefit we're going to see in our scientific community." ❖

Apache™, Apache HTTP Server™, and Apache OODT™ are trademarks of the Apache Software Foundation.