

Managing the deluge of 'big data' from space

October 18 2013



The center of the Milky Way galaxy imaged by NASA's Spitzer Space Telescope is displayed on a quarter-of-a-billion-pixel, high-definition 23-foot-wide (7-meter) LCD science visualization screen at NASA's Ames Research Center in Moffett Field, Calif. Credit: NASA/Ames/JPL-Caltech

(Phys.org) —For NASA and its dozens of missions, data pour in every day like rushing rivers. Spacecraft monitor everything from our home planet to faraway galaxies, beaming back images and information to Earth. All those digital records need to be stored, indexed and processed so that spacecraft engineers, scientists and people across the globe can

use the data to understand Earth and the universe beyond.

At NASA's Jet Propulsion Laboratory in Pasadena, Calif., mission planners and software engineers are coming up with new strategies for managing the ever-increasing flow of such large and complex [data](#) streams, referred to in the information technology community as "big data."

How big is big data? For NASA missions, hundreds of terabytes are gathered every hour. Just one terabyte is equivalent to the information printed on 50,000 trees worth of paper.

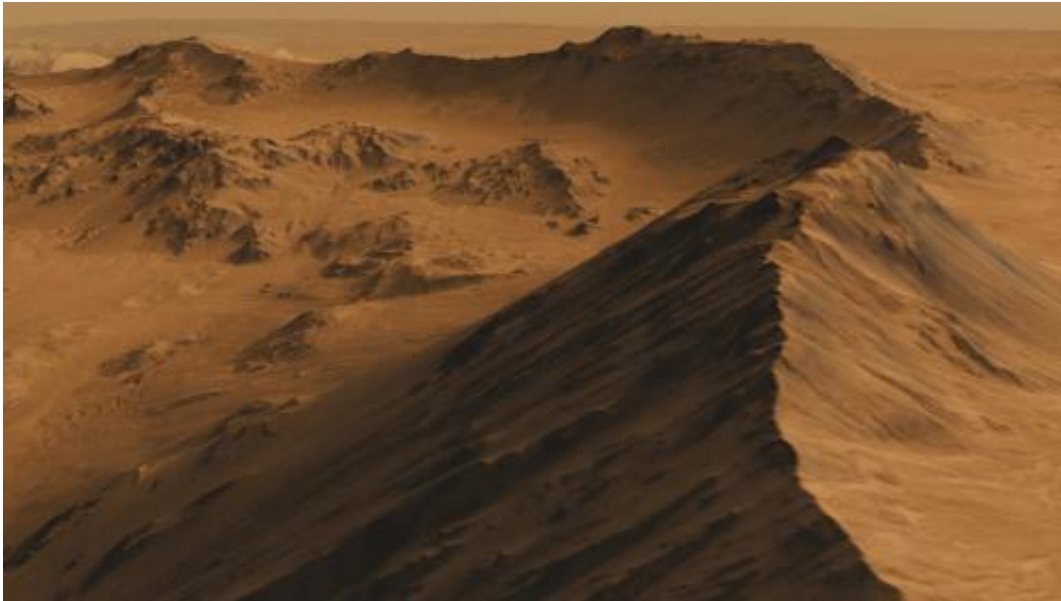
"Scientists use big data for everything from predicting weather on Earth to monitoring ice caps on Mars to searching for distant galaxies," said Eric De Jong of JPL, principal investigator for NASA's Solar System Visualization project, which converts NASA mission science into visualization products that researchers can use. "We are the keepers of the data, and the users are the astronomers and [scientists](#) who need images, mosaics, maps and movies to find patterns and verify theories."

Building castles of data

De Jong explains that there are three aspects to wrangling data from space missions: storage, processing and access. The first task, to store or archive the data, is naturally more challenging for larger volumes of data. The Square Kilometer Array (SKA), a planned array of thousands of telescopes in South Africa and Australia, illustrates this problem. Led by the SKA Organization based in England and scheduled to begin construction in 2016, the array will scan the skies for radio waves coming from the earliest galaxies known.

JPL is involved with archiving the array's torrents of images: 700 terabytes of data are expected to rush in every day. That's equivalent to

all the data flowing on the Internet every two days. Rather than build more hardware, engineers are busy developing creative software tools to better store the information, such as "cloud computing" techniques and automated programs for extracting data.



This is a screen shot from a high-definition simulated movie of Mojave Crater on Mars, based on images taken by the High Resolution Imaging Science Experiment (HiRISE) camera on NASA's Mars Reconnaissance Orbiter. Credit: NASA/JPL-Caltech/Univ. of Arizona

"We don't need to reinvent the wheel," said Chris Mattmann, a principal investigator for JPL's big-data initiative. "We can modify open-source computer codes to create faster, cheaper solutions." Software that is shared and free for all to build upon is called open source or open code. JPL has been increasingly bringing open-source software into its fold, creating improved data processing tools for [space missions](#). The JPL tools then go back out into the world for others to use for different applications.

"It's a win-win solution for everybody," said Mattmann.

In living color

Archiving isn't the only challenge in working with big data. De Jong and his team develop new ways to visualize the information. Each image from one of the cameras on NASA's Mars Reconnaissance Orbiter, for example, contains 120 megapixels. His team creates movies from data sets like these, in addition to computer graphics and animations that enable scientists and the public to get up close with the Red Planet.

"Data are not just getting bigger but more complex," said De Jong. "We are constantly working on ways to automate the process of creating visualization products, so that scientists and engineers can easily use the data."

Data served up to go

Another big job in the field of big data is making it easy for users to grab what they need from the data archives.

"If you have a giant bookcase of books, you still have to know how to find the book you're looking for," said Steve Groom, manager of NASA's Infrared Processing and Analysis Center at the California Institute of Technology, Pasadena. The center archives data for public use from a number of NASA astronomy missions, including the Spitzer Space Telescope, the Wide-field Infrared Survey Explorer (WISE) and the U.S. portion of the European Space Agency's Planck mission.

Sometimes users want to access all the data at once to look for global patterns, a benefit of big data archives. "Astronomers can also browse all the 'books' in our library simultaneously, something that can't be done on

their own computers," said Groom.

"No human can sort through that much data," said Andrea Donnellan of JPL, who is charged with a similarly mountainous task for the NASA-funded QuakeSim project, which brings together massive data sets—space- and Earth-based—to study earthquake processes.

QuakeSim's images and plots allow researchers to understand how earthquakes occur and develop long-term preventative strategies. The data sets include GPS data for hundreds of locations in California, where thousands of measurements are taken, resulting in millions of data points. Donnellan and her team develop software tools to help users sift through the flood of data.

Ultimately, the tide of big data will continue to swell, and NASA will develop new strategies to manage the flow. As new tools evolve, so will our ability to make sense of our universe and the world.

Provided by JPL/NASA

Citation: Managing the deluge of 'big data' from space (2013, October 18) retrieved 24 April 2024 from <https://phys.org/news/2013-10-deluge-big-space.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.