

# Automation offers big solution to big data in astronomy

March 24 2015, by David Tenenbaum

---



Hubble telescope image of stars forming inside a cloud of cold hydrogen gas and dust in the Carina Nebula, 7,500 light-years away. Credit: Space Telescope Science Institute

It's almost a rite of passage in physics and astronomy. Scientists spend years scrounging up money to build a fantastic new instrument. Then, when the long-awaited device finally approaches completion, the panic begins: How will they handle the torrent of data?

That's the situation now, at least, with the Square Kilometer Array (SKA), a radio telescope planned for Africa and Australia that will have an unprecedented ability to deliver [data](#)—lots of data points, with lots of details—on the location and properties of stars, galaxies and giant clouds of [hydrogen gas](#).

In a study published in *The Astronomical Journal*, a team of scientists at the University of Wisconsin-Madison has developed a new, faster approach to analyzing all that data.

Hydrogen clouds may seem less flashy than other radio telescope targets, like exploding galaxies. But [hydrogen](#) is fundamental to understanding the cosmos, as it is the most common substance in existence and also the "stuff" of stars and galaxies.

As astronomers get ready for SKA, which is expected to be fully operational in the mid-2020s, "there are all these discussions about what we are going to do with the data," says Robert Lindner, who performed the research as a [postdoctoral fellow](#) in astronomy and now works as a data scientist in the private sector. "We don't have enough servers to store the data. We don't even have enough electricity to power the servers. And nobody has a clear idea how to process this tidal wave of data so we can make sense out of it."

Lindner worked in the lab of Associate Professor Snezana Stanimirovic, who studies how hydrogen clouds form and morph into stars, in turn shaping the evolution of galaxies like our own Milky Way.

In many respects, the hydrogen data from SKA will resemble the vastly slower stream coming from existing radio telescopes. The smallest unit, or pixel, will store every bit of information about all hydrogen directly behind a tiny square in the sky. At first, it is not clear if that pixel registers one cloud of hydrogen or many—but answering that question is

the basis for knowing the actual location of all that hydrogen.

People are visually oriented and talented in making this interpretation, but interpreting each pixel requires 20 to 30 minutes of concentration using the best existing models and software. So, Lindner asks, how will astronomers interpret hydrogen data from the millions of pixels that SKA will spew? "SKA is so much more sensitive than today's radio telescopes, and so we are making it impossible to do what we have done in the past."

In the new study, Lindner and colleagues present a computational approach that solves the hydrogen location problem with just a second of computer time.

For the study, UW-Madison postdoctoral fellow Carlos Vera-Ciro helped write software that could be trained to interpret the "how many clouds behind the pixel?" problem. The software ran on a high-capacity computer network at UW-Madison called HTCondor. And "graduate student Claire Murray was our 'human,'" Lindner says. "She provided the hand-analysis for comparison."

Those comparisons showed that as the new system swallows SKA's data deluge, it will be accurate enough to replace manual processing.

Ultimately, the goal is to explore the formation of stars and galaxies, Lindner says. "We're trying to understand the initial conditions of star formation—how, where, when do they start? How do you know a star is going to form here and not there?"

To calculate the overall evolution of the universe, cosmologists rely on crude estimates of initial conditions, Lindner says. By correlating data on hydrogen clouds in the Milky Way with ongoing star formation, data from the new [radio telescopes](#) will support real numbers that can be

entered into the cosmological models.

"We are looking at the Milky Way, because that's what we can study in the greatest detail," Lindner says, "but when astronomers study extremely distant parts of the universe, they need to assume certain things about gas and [star formation](#), and the Milky Way is the only place we can get good numbers on that."

With automated data processing, "suddenly we are not time-limited," Lindner says. "Let's take the whole survey from SKA. Even if each pixel is not quite as precise, maybe, as a human calculation, we can do a thousand or a million times more pixels, and so that averages out in our favor."

Provided by University of Wisconsin-Madison

Citation: Automation offers big solution to big data in astronomy (2015, March 24) retrieved 22 May 2024 from <https://phys.org/news/2015-03-automation-big-solution-astronomy.html>

<p>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.</p>
--