



by Lori Ann White



HOW THE PUBLIC RELEASE OF DATA FROM THE FERMI GAMMA-RAY SPACE TELESCOPE'S MAIN INSTRUMENT HAS AFFECTED THE HUNDREDS OF RESEARCHERS WHO USE IT— AND RESULTED IN MORE AND BETTER SCIENCE.

For three and a half years, the Fermi Gamma-ray Space Telescope has been circling the planet in low-Earth orbit, scanning the sky for evidence of the most energetic objects in the universe. For three and a half years, data captured by Fermi's instruments have been accumulating on a NASA server at Goddard Space Flight Center in Maryland.

Who accesses those data, who analyzes those data, who owns those data—the very stuff of discovery—is not who you might expect.

The telescope itself is a space-worthy particle detector built to capture some of the most energetic photons in the universe—gamma-ray photons slung out of the magnetic fields of neutron stars or the blazing hearts of active galactic nuclei, or thrown off the spinning accretion disks of black holes.

It's a hybrid project, born of a partnership between astrophysics and high-energy physics, led, not surprisingly, by NASA and the Department of Energy. And when it comes to handling data, most scientists working on the project represent one of two distinct cultures.

In particle physics, researchers tend to keep their data close, holding it within their experimental collaborations. Many of them have worked on specific particle detectors for years, designing and building and tweaking the equipment. They argue that only collaboration members understand the detectors well enough to interpret the data correctly.

In astrophysics, big orbiting observatories like NASA's Hubble Space Telescope are generally at the disposal of independent research groups that are awarded blocks of observation time, based on proposals reviewed by expert panels. Observers typically get sole use of the resulting data for a set time period—often a year—before they go public.

But the scientists working with Fermi's main instrument, the Large Area Telescope or LAT, do neither.

After Fermi's first year in orbit—as much a time to verify that the telescope worked as a data-gathering period—the LAT collaboration posted all of its data online. And it continues to put the data online just as fast as the instruments on the scope can spit it out and crews on the ground can process it.

GETTING THE WORD OUT

Locking up precious information about the whole sky for months so that a guest investigator with an interest in, say, a particular blazar has sole access to the data would not make sense, says Stanford University's Peter Michelson, principal investigator for the LAT.

"Fermi is a survey scope. One of the wonderful things about this mission is that we're looking everywhere," Michelson says. The LAT can see about 20 percent of the sky at any one time, covering the entire visible sky every three hours.

But nothing as complex as the LAT instrument had been flown before, and the team would need time to check it out, verify that it was performing as expected, determine the level of background signals and so on before unleashing it on the wider community. So the collaboration decided to wait a year, then release all the data as soon as they came in.

There are pluses and minuses to this approach, says Seth Digel, an astrophysicist based at SLAC National Accelerator Laboratory who has been with the Fermi LAT collaboration for more than a decade. Among the pluses: "NASA's interest is in maximizing the scientific return from a mission. Also, to their credit, they're interested in archiving data." He sees that as evidence of NASA's commitment to making mission data available not just across space, but across time, too.

Among the minuses: "The data management decisions

were pretty painful," Digel says. Another minus? "We're giving away great stuff." Another plus: "There's lots of it. In a sense, there's plenty to go around."

DATA TRAIL

Just how public is this data windfall?

"Anybody is free to download the data and the software" used to analyze them, says Anders Borgland, a SLAC physicist who leads the data processing team at the laboratory's Instrument Science Operations Center, where raw LAT data are processed. SLAC managed the development and construction of the LAT and is still deeply involved in its day-to-day operations.

The LAT has sorted through a lot of photons in its three and a half years of operation, but the actual amount of data available for download is surprisingly small. The telescope's trigger—a yes-no switch that decides if a charged particle hitting the LAT's detectors is interesting or not—says "yes" to about 2000 of these hits, called events, per second. "But then a software filter reduces that to about 500 events per second," Borgland explains.

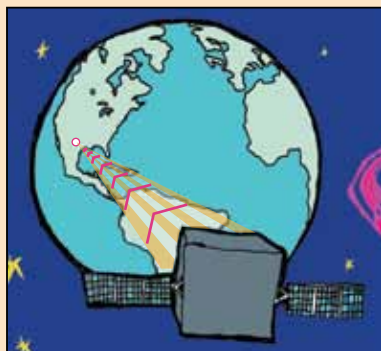
After this quarter-fold data reduction, the information about the 500 events each and every second is transmitted via satellite relay to the White Sands Test Facility in New Mexico, then to Fermi Mission Ops at NASA's Goddard Space Flight Center, thence to SLAC—where the amount of data is reduced another 10-fold—and finally back to Goddard's Fermi Science Support Center (FSSC), where it's put on the Web. Borgland estimates that about eight hours pass between the time the LAT takes the data and the time they're available for download.

TWO CULTURES

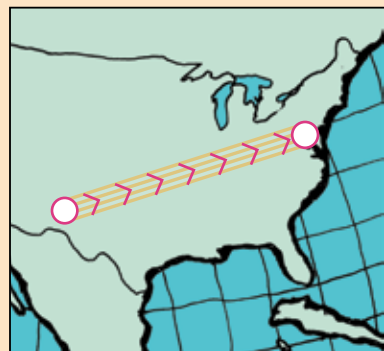
Once uploaded, the data aren't frozen. Taking a page from high-energy physics experiments, the Fermi data analysis teams continue to learn more about their instruments, using that knowledge to

HOW THE DATA TRAVELS

FROM THE LAT
TO YOUR COMPUTER
(IN ABOUT 8 HOURS)



1 Systems on board the Fermi space telescope sort through collected LAT data and preserve only the most relevant hits—about 500 events per second—which are transmitted via satellite to the White Sands Test Facility in New Mexico.



2 From White Sands, data travel to Fermi Mission Ops at NASA's Goddard Space Flight Center in Maryland.

upgrade the analysis software. Once a new version of the software is released all the Fermi data are reanalyzed. The team released data from Pass 7 this summer, after which researchers the world over checked the new data against their old papers to see if revisions were called for.

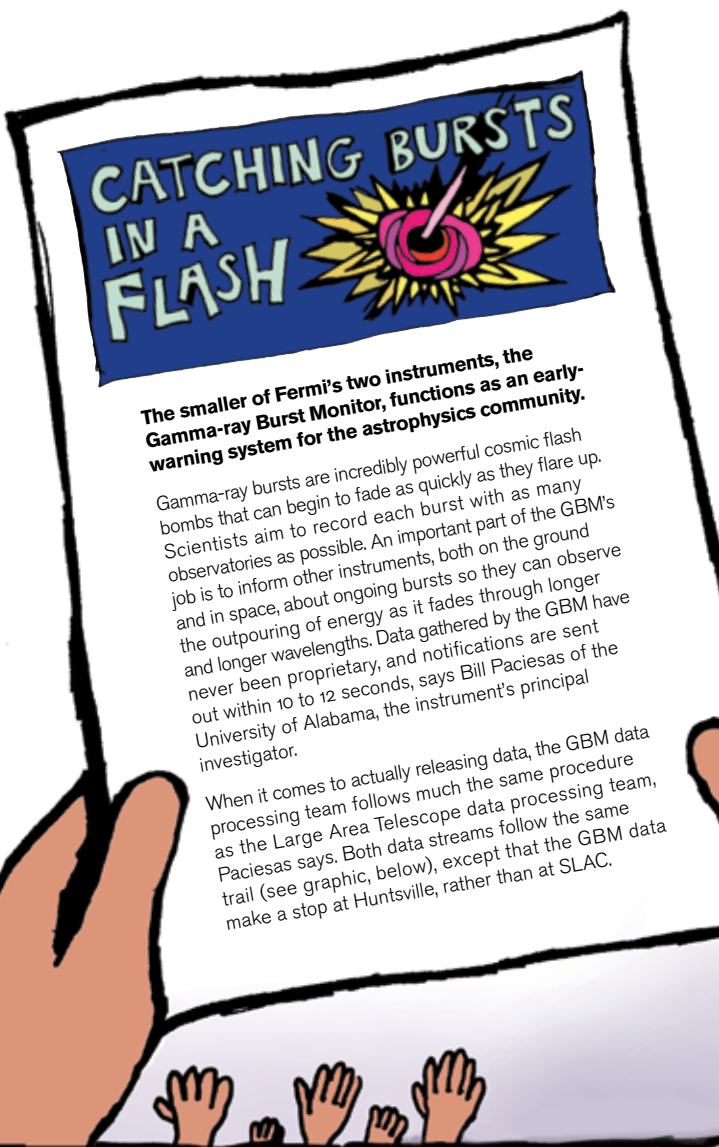
"That's not so common in astrophysics," says Borgland, who hails from the high-energy physics side of the collaboration. There's a trace of glee in his words.

On the other hand, the publicly available Fermi data are formatted as FITS files. FITS stands for Flexible Image Transport System, the standard for astronomical data. "The HEP people expected to use ROOT," a standard set of tools for analyzing particle physics data, says astrophysicist Digel. "I'm really not sure why." Take that, HEP.

Such little digs are mild, but Michelson, the LAT principal investigator, recalls a time in the early days of the collaboration when these cultural differences were more pronounced. "In meetings, people would say, 'Well, the way we do things in particle physics...'" or "The way we do things in astrophysics...". I finally said, "Let's stop saying that." He smiles. "That was a long time ago. Fights break out, but not along cultural lines."

DATA DRIVEN

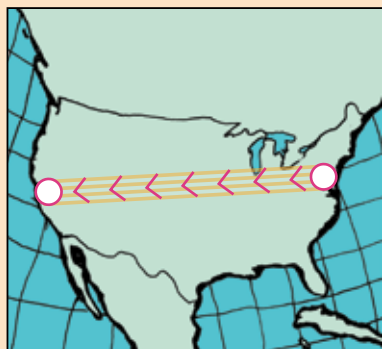
Before the LAT data are sent to the FSSC at Goddard, they are processed. Processing amounts to much more than boiling down each second's worth of events to the one or two really interesting ones. It also involves creating datasets that are of the greatest use to the greatest number of researchers: a list of photons—with the directions they came from, the times they arrived and their energies, of course. Also sent along are data that describe the properties of the LAT needed for analyzing the photon data, but not much more.



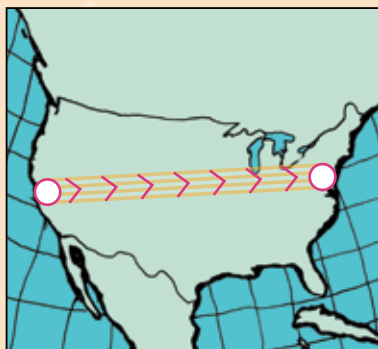
The smaller of Fermi's two instruments, the Gamma-ray Burst Monitor, functions as an early-warning system for the astrophysics community.

Gamma-ray bursts are incredibly powerful cosmic flash bombs that can begin to fade as quickly as they flare up. Scientists aim to record each burst with as many observatories as possible. An important part of the GBM's job is to inform other instruments, both on the ground and in space, about ongoing bursts so they can observe the outpouring of energy as it fades through longer and longer wavelengths. Data gathered by the GBM have never been proprietary, and notifications are sent out within 10 to 12 seconds, says Bill Paciesas of the University of Alabama, the instrument's principal investigator.

When it comes to actually releasing data, the GBM data processing team follows much the same procedure as the Large Area Telescope data processing team, Paciesas says. Both data streams follow the same trail (see graphic, below), except that the GBM data make a stop at Huntsville, rather than at SLAC.



3 From Goddard, LAT data go to SLAC's Instrument Science Operations Center for processing.



4 They return to Goddard and the Fermi Science Support Center, which puts the data on the Web.



5 About 8 hours after Fermi detects an event, the resulting data are available to researchers all over the world.

Scott Ransom knows all about that. Ransom, an astronomer with the National Radio Astronomy Observatory and the University of Virginia, has been a Fermi guest investigator several times over. That means he pores through Fermi data with NASA's blessing—and funding—to find previously unknown pulsars.

He's becoming adept at writing his own software tools, but he's done so with a lot of input from colleagues on the collaboration. "It's been a back-and-forth process," Ransom says. "Almost everything I do with Fermi data involves the Fermi team," which often means the Fermi Support Team at the FSSC, Fermi's main point of contact for researchers outside of the collaborations. "They're the experts."

What he's learned: Fermi may see the entire sky, but not at a very high resolution. When looking at a pulsar, Ransom says, "Fermi sees a blob. But something about the gamma rays is pulsar-like. So we turn the radio telescope to that spot to check."

Those blobs have been like nuggets of gold for pulsar miners like Ransom. "We've been very lucky," he says. "We've found a huge number of pulsars by looking at the Fermi blobs."

GETTING IT RIGHT

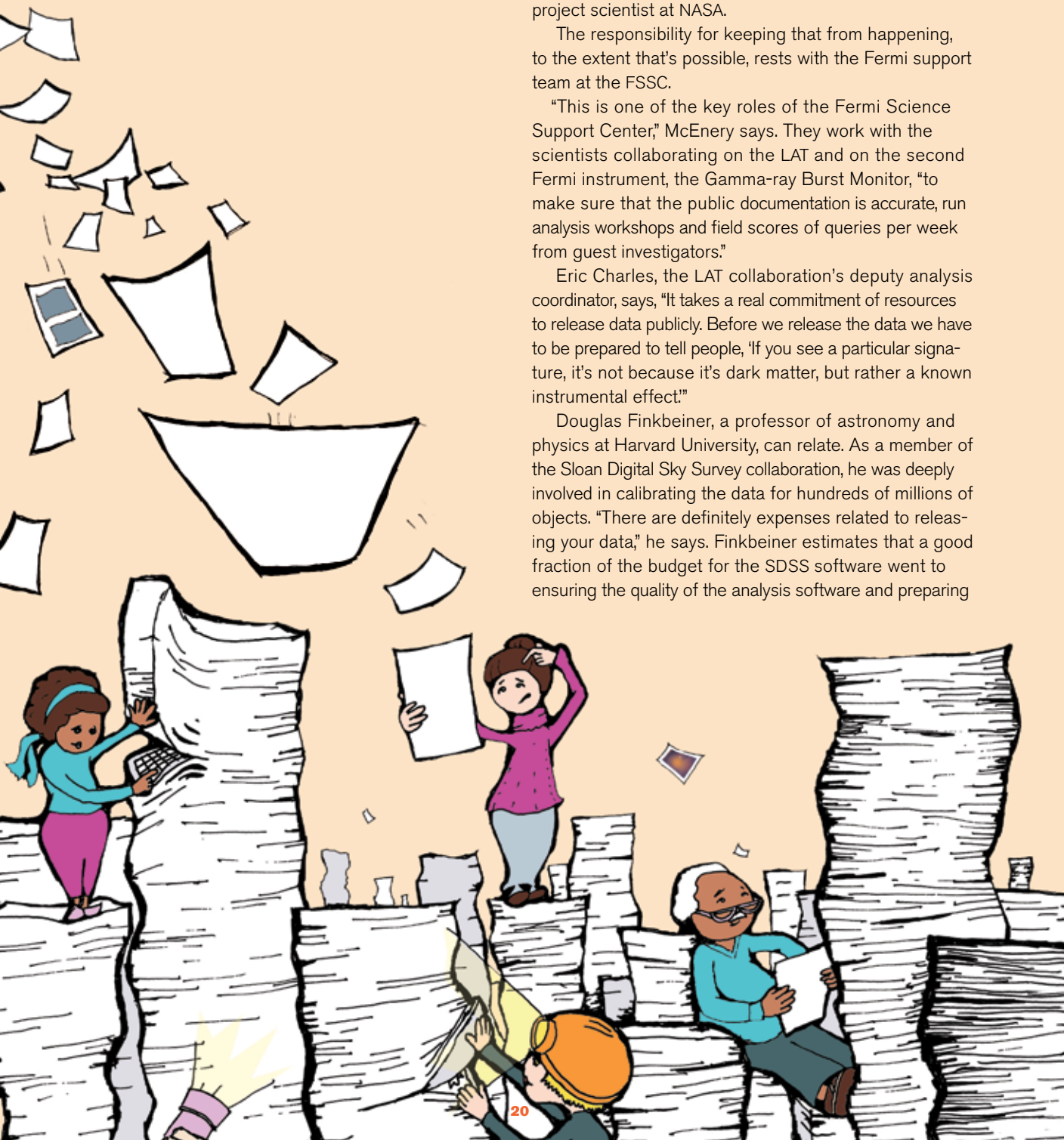
Not everyone who downloads Fermi data does so with a collaboration member on speed dial, a situation that raises its own concerns. "There were real fears that data would be misused" by people who didn't fully understand them, says Julie McEnergy, Fermi project scientist at NASA.

The responsibility for keeping that from happening, to the extent that's possible, rests with the Fermi support team at the FSSC.

"This is one of the key roles of the Fermi Science Support Center," McEnergy says. They work with the scientists collaborating on the LAT and on the second Fermi instrument, the Gamma-ray Burst Monitor, "to make sure that the public documentation is accurate, run analysis workshops and field scores of queries per week from guest investigators."

Eric Charles, the LAT collaboration's deputy analysis coordinator, says, "It takes a real commitment of resources to release data publicly. Before we release the data we have to be prepared to tell people, 'If you see a particular signature, it's not because it's dark matter, but rather a known instrumental effect!'"

Douglas Finkbeiner, a professor of astronomy and physics at Harvard University, can relate. As a member of the Sloan Digital Sky Survey collaboration, he was deeply involved in calibrating the data for hundreds of millions of objects. "There are definitely expenses related to releasing your data," he says. Finkbeiner estimates that a good fraction of the budget for the SDSS software went to ensuring the quality of the analysis software and preparing



the data for release. "I really do think it's worth it," he says, "but don't pretend you can do it for free."

Finkbeiner's support for the Fermi data release might have a bit of an ulterior motive: He knew precisely what he wanted to look for in the initial dataset, and he didn't need help to do it.

In a 2004 paper, Finkbeiner had described a mysterious microwave feature in the galactic center dubbed the "WMAP haze" for the instrument that first saw it, the Wilkinson Microwave Anisotropy Probe. He wanted to look at the same area through gamma rays, and marked the date on his calendar when the Fermi data would be released: "We were looking forward to the data for years."

Finkbeiner and his colleagues made their own maps of gamma-ray distributions in the galactic center and refined them as the LAT collaboration's own data analysis tools improved and the data got better.

"What we definitely saw was a blob of excess gamma rays" that correlated with the WMAP haze, Finkbeiner says—two 25,000-light-year-long gamma-ray "bubbles" that burp out from the galactic plane in both directions. The resulting paper, which appeared a year ago in the *Astrophysics Journal*, caused quite a fuss.

It also had only three authors, with nary a Fermi member among them, but Finkbeiner, who knows how hard it is to provide quality data to outside investigators, wants to emphasize why: "We didn't need much help from the Fermi team, but only because they did such an excellent job with the data release and documentation," he says.

DATA VALUES

McEnergy is in a very good position for a project scientist right now. When doubts arise about Fermi's data policy, she has three years of results to refute concerns. Data misinterpretation? McEnergy points to Finkbeiner's independent, solid analysis that led to the Fermi bubbles result. Outsiders scoop the collaboration too often? She points to the fact that while more and more papers are being published by outsiders, the LAT collaboration's publication numbers have held steady. The collaboration is weakening? She points out that researchers are still joining.

"The collaboration isn't held together just by the data," says McEnergy, who also belongs to the LAT collaboration. "We've achieved a balance. We've succeeded in creating

within the collaboration a very vibrant scientific environment, and people outside are also doing some very nice things."

For McEnergy, the bottom line is once again the science: "Getting the most and the best science out of the instrument—of course we want the best minds on it."

No matter where those minds may be.

POST-SCRIPT

Strictly speaking, the LAT does not detect gamma rays directly. Instead, it detects charged particles, including paired electrons and positrons that are created when gamma rays hit thin sheets of foil in the detector.

Other charged particles of all types whiz through the not-so-empty space at the telescope's altitude, some 340 miles up. Protons, electrons, atomic nuclei, and even more exotic particles are flung into space by just about every energetic body: the sun, other stars, and especially such highly energetic phenomena as pulsars, black-hole accretion disks, and supernovae.

The LAT picks up this random bounty of exotic charged particles, too, but the instrument is a gamma-ray detector, not a cosmic-ray detector. The extra data are more difficult to analyze and are not included in the packets sent to the FSSC. "Only the photon data are made public," Borgland explains.

But data are data, and LAT collaboration members couldn't let them go to waste. They used the data to confirm a 2009 discovery by an instrument called PAMELA of an excess of positrons in cosmic rays, resulting in what's currently the LAT collaboration's most oft-cited paper: "Measurement of the Cosmic Ray e^+ plus e^- Spectrum from 20 GeV to 1 TeV with the Fermi Large Area Telescope," published in *Physical Review Letters* in May 2009.

LAT researchers didn't stop there. Theorists had conjectured that the positrons could be a sign of dark matter, the mysterious substance that accounts for about 80 percent of the mass of the universe. In an experimental tour de force, the collaboration members used the Earth's magnetic field to split the electrons and positrons, improving their analysis enough to enable them to weigh in on whether the positrons suggested the presence of dark matter. (The answer: Probably not.)

In a universe in which so much is unknown, even the cast-off bits can be a precious commodity.

