

Returning Lightning Data to the Cloud

Scientists are assembling an online database with decades of low-frequency radio measurements collected worldwide to facilitate modern research about lightning, space weather, and more.



Data collected using the antenna array of the High-frequency Active Auroral Research Program in Alaska are included in the new Worldwide Archive of Low-frequency Data and Observations (WALDO) database, along with data from numerous other sites. Credit: [Secoy, A.](#), [CC BY-SA 4.0](#)

By [Morris Cohen](#) © 24 April 2020

This past December at AGU's Fall Meeting in San Francisco, I presented a [poster](#) (<https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/586335>) with not a shred of new science on it. Yet it

might turn out to be the highest-impact presentation I've made.

With the poster, several colleagues and I introduced WALDO to the world. WALDO (<http://waldo.world/>), or the Worldwide Archive of Low-frequency Data and Observations, is a large—and growing—trove of low-frequency (0.5 to 50 kilohertz) radio data collected over decades at sites around the world. Mark Golkowski of the University of Colorado Denver (CU Denver) and I jointly manage the database.

Such data have all kinds of uses in geophysics, including in lightning detection (<https://eos.org/articles/students-launch-balloon-borne-payloads-into-thunderstorms>) and characterization, remote sensing of ionospheric and magnetospheric phenomena, and detection of solar flares, gamma ray flashes (<https://eos.org/science-updates/understanding-high-energy-physics-in-earths-atmosphere>), and gravity waves (<https://eos.org/editor-highlights/novel-simulations-of-upper-atmosphere-gravity-wave-dynamics>). Until recently, however, the data on WALDO have been amassed and stored mainly on tens of thousands of DVDs—and thus have been largely inaccessible to anyone interested in using them.

Our goal with WALDO is to transfer and organize these historical data, augmented with ongoing data collection, into a single, standardized cloud-based repository so that scientists today and in the future can access them and put them to use in studies of lightning, the ionosphere, the magnetosphere, space weather, and more.

The Science of ELF/VLF

Lightning Science Strikes



(https://eos.org/wp-content/uploads/2020/04/May20_magazine.pdf)

- [Lightning Research Flashes Forward](https://eos.org/features/lightning-research-flashes-forward) (<https://eos.org/features/lightning-research-flashes-forward>)
- [Planetary Lightning: Same Physics, Distant Worlds](https://eos.org/features/planetary-lightning-same-physics-distant-worlds) (<https://eos.org/features/planetary-lightning-same-physics-distant-worlds>)
- [Studying Earth's Double Electrical Heartbeat](https://eos.org/features/studying-earths-double-electrical-heartbeat) (<https://eos.org/features/studying-earths-double-electrical-heartbeat>)
- [Returning Lightning Data to the Cloud](https://eos.org/science-updates/returning-lightning-data-to-the-cloud) (<https://eos.org/science-updates/returning-lightning-data-to-the-cloud>)
- [Understanding High-Energy Physics in Earth's Atmosphere](https://eos.org/science-updates/understanding-high-energy-physics-in-earths-atmosphere) (<https://eos.org/science-updates/understanding-high-energy-physics-in-earths-atmosphere>)
- [Mapping Lightning Strikes from Space](https://eos.org/articles/mapping-lightning-strikes-from-space) (<https://eos.org/articles/mapping-lightning-strikes-from-space>)
- [Investigating the Spark](https://eos.org/agu-news/investigating-the-spark) (<https://eos.org/agu-news/investigating-the-spark>)

Each of the millions of lightning strokes per day on Earth releases an intense, roughly 1-millisecond-long pulse of extremely low frequency to very low frequency (ELF/VLF) radio energy known as a sferic. These sferics reflect from the lower ionosphere (60–90 kilometers altitude) and off the ground, allowing them to travel—and be detected—globally. A handful of VLF receivers scattered around the globe can geolocate most lightning flashes with incredible kilometer-level accuracy [*Said et al.* (<https://doi.org/10.1029/2010.JD013863>), 2010]. Sferic detection can also be used to characterize the electrical properties of the lower ionosphere between the source and a distant receiver.



A VLF/LF antenna stands mounted on the National Oceanic and Atmospheric Administration's R/V Ronald H. Brown, while the ship was docked in Puerto Rico. Credit: Morris Cohen

Narrowband beacons used by the U.S. Navy, nominally for submarine communications, also transmit in the ELF/VLF frequency band, providing another means of ionospheric remote sensing. Although these messages are encrypted for security, the radio signals themselves are a useful ionospheric diagnostic that can be picked up anywhere on Earth. Changes in ionospheric conditions, namely, the electron density, manifest as changes to either the amplitude or the phase of received signals. In turn, the ionosphere can be used as a sensor to monitor all kinds of geophysical phenomena, including solar

flares, electron precipitation from the magnetosphere, solar eclipses, lightning-related heating, cosmic gamma rays, gravity waves, and much more. Each of these phenomena disturbs VLF signals propagating under the ionosphere in different ways—affecting how quickly a disturbance begins and ends, for example—and these signatures allow them to be distinguished from one other. Some ionospheric disturbances are very reliable and repeatable, like the effect of the Sun rising and setting.

Studying ELF/VLF radio waves allows us to piece together mysteries of what happens during space weather events and geomagnetic storms.

Some ELF/VLF energy also escapes into the magnetosphere (as lightning-generated plasma waves called whistlers (<https://eos.org/research-spotlights/how-lightning-creates-killer-electrons-in-earths-radiation-belts>)), where it can interact with trapped energetic electrons in Earth's radiation belt and trigger precipitation of electrons into the atmosphere. ELF/VLF waves are also generated and accelerated in the magnetosphere (as waves called chorus and hiss ([https://en.wikipedia.org/wiki/Hiss_\(electromagnetic\)](https://en.wikipedia.org/wiki/Hiss_(electromagnetic)))) as a result of wave-particle interactions and thus play a role in the dynamics of space weather at Earth. Studying ELF/VLF radio waves allows us both to study and better understand these processes and to piece together mysteries of what happens during space weather events and geomagnetic storms.

These uses of ELF/VLF data, reviewed by, for example, *Barr et al.* ([https://doi.org/10.1016/S1364-6826\(00\)00121-8](https://doi.org/10.1016/S1364-6826(00)00121-8)) [2000], *Inan et al.* (<https://doi.org/10.1029/2009JA014775>) [2010], and *Silber and Price* (<https://doi.org/10.1007/s10712-016-9396-9>) [2017], have been developed since the late 1800s, when natural ELF/VLF signals could be heard coupling into long telegraph lines. But a number of other applications outside the traditional uses of ELF/VLF data have also popped up recently. For example, detection of objects inside metal boxes using ELF/VLF waves [*Harid et al.* (<https://doi.org/10.1109/TEM.2019.2952560>), 2019] could be used to discover a cache of guns hidden inside a shipping container.

In partnership with a cybersecurity research group at the Georgia Institute of Technology (Georgia Tech), colleagues and I are also using ELF/VLF data to boost the security of the power grid against cyberattacks, such as the major attack (<https://theconversation.com/cyberattack-on-ukraine-grid-heres-how-it-worked-and-perhaps-why-it-was-done-52802>) in Ukraine in December 2015 in which hackers disabled multiple electrical substations. ELF/VLF data detected by radio receivers can be used to monitor power grid signals for irregularities. These data are also littered with sferics from lightning flashes around the world, which arrive at receivers at quasi-random times as lightning occurs. Nature thus provides an effective and detectable random number generator that because lightning flashes cannot be predicted in advance, allows us to validate the integrity of other data detected by the receivers [*Shekari et al.* (<https://doi.org/10.14722/ndss.2019.23462>), 2019].

Developing WALDO

The WALDO database—currently about 200 terabytes and growing daily—already contains or will soon contain data that could enrich studies of all of the above phenomena and applications. Much of the data were collected by Stanford University ELF/VLF receivers and, more recently, by new sites deployed by Georgia Tech and CU Denver.



This map shows a selection of VLF antenna sites that have, or still are, collecting data included in the WALDO database. Credit: Morris Cohen

WALDO also includes ELF/VLF recordings from experiments carried out as part of the High-frequency Active Auroral Research Program (HAARP (<https://www.gi.alaska.edu/facilities/haarp>)) in Alaska [*Cohen and Golkowski* (<https://doi.org/10.1002/jgra.50558>), 2013], which has been running experiments to study the high-latitude ionosphere since the mid-1990s. It includes many years of data from Palmer Station (<https://antarcticsun.usap.gov/aroundthecontinent/1603/>) on the Antarctic Peninsula. And it will eventually include a lot of data from the famous Siple Station ELF experiment (<https://vlf.stanford.edu/research/vlf-transmitter-siple-station-antarctica>), which ran from 1973 to 1988 to study the amplification and triggering of ELF signals in the magnetosphere using a 42-kilometer antenna in Antarctica. By the end of the year, we anticipate having 500–1,000 terabytes of data available.

We have developed an online interface that allows easy access to the data. Through the website, users can view automatically generated quick-look plots to make it easy to find out what's available.

The effort to compile these disparate data sets into a single database began in fall 2018, when the space at Stanford University where these data were physically stored—on roughly 80,000 DVDs and CDs and on one badly corrupted server—had to be cleared. The disks, some of which were damaged after decades of storage, were packed and shipped to either Georgia Tech or CU Denver, where DVD-reading robots that can rip a stack of 300 disks at a time are used to move the data onto hard drives. Meanwhile, John DeSilva at Stanford has slowly extracted the contents of the old server and placed those data into temporary cloud storage for us to retrieve.

After retrieval, the data are passed through a digital sorting scheme that updates the formatting so it is all consistent and then places the data into sorted folders. We have developed an online interface (<http://waldo.world/>) that allows easy access to the data, which can also be shared with anyone with a Google account upon request. Through the website, users can view automatically generated quick-look plots to make it easy to find out what's available, for example, maps of receiver sites from which data from a given day are available, annual calendars showing data availability, and summary charts of the data on a day-by-day basis.

The Value of Dusty Data

The work of preserving data is hard and time-consuming but also rewarding. We have seen evidence of this in many fields. Historical and long-term data sets have been critical in studies of climate and ecosystems, for example, shedding light not only on past conditions but also on the present and future. And thanks to preservation efforts, we are fortunate to have sunspot data extending back more than 400 years—data that underlie critical early discoveries of space weather dynamics.

As a junior at Stanford in January 2002, I approached one of my professors, Umran Inan, and asked whether I could get involved in research. I suspect he wasn't anticipating much from a student who had just gotten a C in his class. Days later I found myself in a dusty, nearly abandoned warehouse near the Stanford Dish (<https://dish.stanford.edu/>), rummaging through 15-year-old Betamax and Ampex magnetic tapes filled with ELF/VLF radio data. The tapes were still stuffed in their original cardboard boxes and were lined up on shelves stacked 5 meters high in several rows, each probably 30 meters long. Why was I there?



A very low frequency radio antenna sits atop a glacier in 2006 (top) near Palmer Station on the Antarctic Peninsula (bottom). Credit: Morris Cohen (top); [Christopher Michel](#)

([https://commons.wikimedia.org/wiki/File:Palmer_station_antarctica_\(15662102160\).jpg](https://commons.wikimedia.org/wiki/File:Palmer_station_antarctica_(15662102160).jpg)) (bottom), [CC BY 2.0](#) (<https://creativecommons.org/licenses/by/2.0/legalcode>).

In 1994, bursts of high-energy gamma rays called terrestrial gamma ray flashes (TGFs) were discovered serendipitously from space [*Fishman et al.* (<https://doi.org/10.1126/science.264.5163.1313>), 1994]. It appeared that TGFs originated with lightning, but that was pretty much all we knew about them. ELF/VLF data can be used to characterize the lightning that caused the phenomenon, but scientists had only two examples in hand of TGFs that could be directly linked with lightning via ELF/VLF data. My job was to find more examples hidden in the data on all those tapes.

As I coughed away the cobwebs, I thought about all the trouble people had gone through to keep these Betamax tapes (long an obsolete format even by then) flowing. The data I was looking through were recorded at Palmer Station, Antarctica, by a receiver mounted on a shifting glacier that was carefully

watched by a full-time science technician and serviced every year by a student in the group. With each boat trip from the station, the tapes were shipped out in large boxes, then stacked and stored in this rodent-infested space—all funded by American taxpayer dollars via the National Science Foundation. And this sort of data collection had been going on for decades at sites all over the world maintained by [this research group](https://vlf.stanford.edu/people/umran-inan) (<https://vlf.stanford.edu/people/umran-inan>).

Living Data Sets

I learned that geophysical data sets are living and that their intellectual value shifts as our scientific priorities do.

“Was it worth it?” I thought while slogging away in that warehouse. The answer, as I came to find out, is an unequivocal yes (and not just because these data led to my first peer-reviewed research papers and helped me get my foot in the door of research). I learned that geophysical data sets are living and that their intellectual value shifts as our scientific priorities do.

When the measurements recorded on those Betamax tapes were obtained, no one envisioned eventually needing them to study TGFs; the measurements were originally collected for other reasons. It would have been easy to throw the data away before they proved useful for studying TGFs—or even after that too. Following the use of Betamax tapes, we shifted to recording digital data on CDs, then on DVDs, then on external hard drives, then onto a large data server—and now we’re moving them into the cloud. At every step, we had to drag all the accumulated data from old media into the present day. But because these data haven’t been discarded, they are still available today for studying numerous natural phenomena and processes.

It’s fair to ask whether it’s worth it given the expense and effort. I think it is. You never know how these data might be used. I would have never expected geophysical lightning data to make an impact in the cybersecurity world, for example. Today we are seeing high-performance computing and machine learning reveal new insights from old data, and interdisciplinary projects often find surprising uses for historical data sets. In the not-too-distant future, I suspect someone will think of a new way to look at ELF/VLF data collected a decade ago. But will the data still be available?

We owe it to future scientists—and to U.S. taxpayers, who have funded much of this work—to ensure that they are available. Since announcing WALDO in December, we’ve gotten several inquiries and notifications from people using the database. Our hope is that by preserving these data in WALDO, we will open doors for surprising and unexpected discoveries.

References

Barr, R., D. Llanwyn Jones, and C. J. Rodger (2000), ELF and VLF radio waves, *J. Atmos. Sol. Terr. Phys.*, **62**, 1,689–1,718, [https://doi.org/10.1016/S1364-6826\(00\)00121-8](https://doi.org/10.1016/S1364-6826(00)00121-8) ([https://doi.org/10.1016/S1364-6826\(00\)00121-8](https://doi.org/10.1016/S1364-6826(00)00121-8)).

Cohen, M. B., and M. Golkowski (2013), 100 days of ELF/VLF generation via HF heating with HAARP, *J. Geophys. Res. Space Phys.*, **118**, 6,597–6,607, <https://doi.org/10.1002/jgra.50558> (<https://doi.org/10.1002/jgra.50558>).

Fishman, G., et al. (1994), Discovery of intense gamma-ray flashes of atmospheric origin, *Science*, **264**, 1,313–1,316, <https://doi.org/10.1126/science.264.5163.1313> (<https://doi.org/10.1126/science.264.5163.1313>).

Harid, V., et al. (2019), Magnetic field penetration into a metal enclosure using an ELF/VLF loop antenna, *IEEE Trans. Electromagn. Compatibility*, <https://doi.org/10.1109/TEM.2019.2952560> (<https://doi.org/10.1109/TEM.2019.2952560>).

Inan, U. S., S. A. Cummer, and R. A. Marshall (2010), A survey of ELF and VLF research on lightning–ionosphere interactions and causative discharges, *J. Geophys. Res.*, **115**, A00E36, <https://doi.org/10.1029/2009JA014775> (<https://doi.org/10.1029/2009JA014775>).

Said, R. K., U. S. Inan, and K. L. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, *J. Geophys. Res.*, **115**, D23108, <https://doi.org/10.1029/2010JD013863> (<https://doi.org/10.1029/2010JD013863>).

Shekari, T., et al. (2019), RFDIDS: Radio frequency–based distributed intrusion detection system for the power grid, in *Network and Distributed Systems Security (NDSS) Symposium 2019*, Internet Soc., Reston, Va., <https://doi.org/10.14722/ndss.2019.23462> (<https://doi.org/10.14722/ndss.2019.23462>).

Silber, I., and C. Price (2017), On the use of VLF narrowband measurements to study the lower ionosphere and the mesosphere–lower thermosphere, *Surv. Geophys.*, **38**, 407–441, <https://doi.org/10.1007/s10712-016-9396-9> (<https://doi.org/10.1007/s10712-016-9396-9>).

Author Information

Morris Cohen (mcohen@gatech.edu (<mailto:mcohen@gatech.edu>)), Georgia Institute of Technology, Atlanta

Citation: Cohen, M. (2020), Returning lightning data to the cloud, *Eos*, **101**, <https://doi.org/10.1029/2020EO142801>. Published on 24 April 2020.

Text not subject to copyright.

Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.