ITS_LIVE: A Cloud-Native Approach to Monitoring Glaciers from Space

Luis A. López[a], Alex S. Gardner[b], Chad A. Greene[b], Mark Fahnestock[c], Joseph H. Kennedy[c], Maria Liukis[b], Theodore A Scambos[e], Jacob R. Fahnestock[f]

[a] *NSIDC, University of Colorado Boulder,*
[b] *NASA, Jet Propulsion Laboratory,*
[c] *UAF, Geophysical Institute,*
[d] *UAF, Alaska Satellite Facility,*
[e] *CIRES, Earth Science and Observation Center,*
[f] *Wentworth Institute of Technology,*

Abstract

NASA's ITS_LIVE project delivers near-real-time information on global glacier dynamics and provides historical context for climate change with a record of how every glacier in the world has evolved over decades of satellite observation. To handle petabytes of data in the satellite archives and a constant influx of new observations, ITS_LIVE has adopted a modern cloud-native approach that is scalable, highly performant, user friendly, and embraces transparent and collaborative code development. As every discipline of Earth science is being transformed by a new era of remote sensing and cloud computing, ITS_LIVE offers a progressive approach to maximizing scientific advancements through open science.

Keywords: glaciology, climate change, big data, cloud computing

—A sea change for Earth Science

Earth Science is being upended by advancements in cloud computing technology, an explosion in remote sensing data collection (Fig. 1), and a cultural shift toward open science. The global satellite data record has grown too large to be downloaded and analyzed on personal computers, machine-learning algorithms have grown too complex to be maintained by any single research group, and there is a growing recognition that public investments in research are compounded when code is open. NASA is embracing the new paradigm, and is moving all Earth observation data archives into Earthdata cloud https://www.earthdata.nasa.gov/eosdis/cloud-evolution as part of the Transform to Open Science initiative https://nasa.github.io/Transform-to-Open-Science/. Here we show how NASA's MEaSUREs program is pushing Earth Science into the cloud computing era with automated production, open code, intelligent cloud-optimized data products, single-line-of-code data access and subsetting, and no-code, user-friendly interfaces for data exploration and visualization as developed for the ITS_LIVE project https://its-live.jpl.nasa.gov.

—Canaries in a Coalmine

The ITS_LIVE project serves a community of researchers who study how glaciers have, and will, respond to our changing climate. By measuring how glaciers interact with Earth's system, scientists can form better predictions on sea level rise and potential impacts on ocean circulation, Earth's energy budget, agriculture, hydropower generation, and water security for large populations.

Glaciers often serve as "canaries in the coalmine", responding to a warming environment well before other impacts of climate change become evident. But to benefit from Earth's natural early-warning systems requires near-real-time monitoring of all the world's glaciers at high spatial and temporal res-

*Corresponding author
Email addresses: `luis.lopezespinosa@colorado.edu` (Luis A. López), `alex.s.gardner@jpl.nasa.gov` (Alex S. Gardner), `chad.a.greene@jpl.nasa.gov` (Chad A. Greene), `mfahnestock@alaska.edu` (Mark Fahnestock), `jhkennedy@alaska.edu` (Joseph H. Kennedy), `maria.liukis@jpl.nasa.gov` (Maria Liukis), `tascambos@Colorado.EDU` (Theodore A Scambos), `jrfz31@gmail.com` (Jacob R. Fahnestock)

olution. ITS_LIVE continuously processes, assimilates, and harmonizes data from multiple satellites and sensors including optical imagers, and synthetic aperture radars to produce a high-resolution, near-real-time picture of ice velocity worldwide. Products are available and ready to be integrated into diverse scientific workflows within days of data acquisition, and scientists have taken note, generating workflows that rely on access to near-real-time data. To date, the project has processed over a Petabyte (250 bytes) of archival satellite imagery using complex algorithms to extract the geophysical variable of interest. Such an effort is made feasible by the scalable architecture of cloud computing that provides access to tens of thousands of computational cores co-located with satellite data archives.

—Measuring glacier flow from space

ITS_LIVE measures glacier flow from space using the project's open-source autoRIFT[1,2] algorithm which can be applied to optical or radar satellite data to measure horizontal displacements of subtle features in repeat images, with sub-pixel accuracy. ITS_LIVE currently uses autoRIFT to process all relevant Landsat and Sentinel-2 optical satellite imagery and all Sentinel-1 radar satellite imagery to measure glacier velocity.

As of writing, ITS_LIVE has processed more than 22 million image pairs and generated an equal number of NetCDF files (granules) that contain high-resolution (120 m) velocity data. Granules are restructured into analysis-ready Zarr [https://zarr.dev/] data cubes that are optimized for time series analysis, providing access to all data from 1984 to present for any 120 m by 120 m point on a glacier within a second or two. In all, ITS_LIVE has created more than 138 TB of data. The ITS_LIVE archive is expected to grow at a rate of about 5-10 TB per month, as it is a living dataset that will grow with time, as new data are acquired, and new sensors are assimilated.

—Scaling to 10s of thousands of cores

ITS_LIVE data processing uses a cloud-first approach to minimize downloading of data and to optimize for rapid data access and large-scale analysis. Dataset generation is carried out in Amazon AWS using HyP3[3] , scaling to >10,000 cores. HyP3 is an open-source, cloud-native batch processing pipeline developed by the Alaska Space Facility for processing optical and Synthetic Aperture Radar (SAR) imagery that addresses many common issues for users of SAR data. Most analysis code is run in the US-West-2 region where the majority of the imagery is archived while the Sentinel-2 processing is run in EU-Central-1.

—The magic of user aligned cloud optimized data

Analyzing and visualizing millions of NetCDF files is not an easy task and historically has required prohibitive computing resources. To address data accessibility challenges, ITS_LIVE is one of the first NASA projects to produce cloud optimized data products from the start. To facilitate efficient access data needs to be cloud optimized so that the stored objects can be efficiently retrieved and the objects themselves need to be structured in a way that minimizes the number of objects that need to be located and the amount of data (or chunks) that need to be uncompressed. When done properly, well structured data eliminates the need for complex and computationally expensive web services to subset and transform the data. Development of such a product requires a deep understanding of user access patterns.

ITS_LIVE users, like most Earth Scientists[4,5,6], primarily access the raw data in two ways: (1) spatial exploration and (2) temporal exploration. Spatial exploration is supported through library access to the millions of NetCDF granules that can be identified programmatically using an OpenAPI search or interactively using the ITS_LIVE webtool. Temporal exploration is accommodated by a restructuring of the NetCDF granules into Zarr datacubes that are composed of 10 x 10 x 20,000 chunked data, where 10 x 10 is in the spatial domain and 20,000 is in the temporal domain. This chunking scheme minimizes the amount of data that needs to be uncompressed for a time series at a single location, providing near-zero latency for temporal exploration, which is critical for process understanding and change detection.

ITS_LIVE also provides annual regional mosaics of glacier velocities for those users that require less dense temporal sampling. Mosaics are posted at the native 120 m spatial resolution and provide additional gridded metrics, such as measurement error and data count. The annual mosaics are provided as both metadata-complete NetCDFs and cloud optimized geotiffs.

The intelligently structured, cloud-optimized ITS_LIVE data formatting, and user-friendly analytical tools developed for the project, together enable rapid development of code that can be run on a personal computer or in the cloud. By structuring the data in a way that supports low-latency access, ITS_LIVE makes it painless for users to develop code that streams the data directly from the cloud, supplanting the need for local archives. Such code is immensely more reproducible as it can be run on any computer with an internet connection, without the need to recreate local data archives and update local paths.

—Serverless access, exploration and visualization

ITS_LIVE provides users with intuitive no-code options to explore the rich dataset through a Voila-based dashboard and a serverless application [Fig 5].. In the following pages we present some of the data visualizations we can obtain from using ITS_LIVE data without costly pre-processing steps using open-source GIS tools like QGIS[8] and the Pangeo stack in Python (Xarray[7], Zarr[8] , HoloViews[9]). A companion GitHub repository contains all the notebooks and QGIs projects used to generate the figures in this article. The focus areas include Malaspina Glacier in Alaska (Fig. 4,7), Pine Island Glacier in Antarctica and Antarctic ice shelves (Fig. 6)

—**Keys to success**

Earth Science is undergoing a major transition, from being data-poor to data-rich. The old paradigm of navigating to a webpage, downloading data and working with it locally on one's computer is quickly dying. We are now entering an era where we will expect shared code to link to low-latency cloud-accessible datasets so the analysis code can run out-of-the-box and be truly reproducible. By breaking down traditional disciplinary barriers of engineering, technology, and Earth science, ITS_LIVE has developed optimized technical solutions to obstacles that might otherwise stand in the way of creating, accessing, exploring, and visualizing such novel and complex datasets. ITS_LIVE was built on Findable, Accessible, Interoperable, and Reusable (FAIR) data principles, and has mitigated many of the challenges surrounding rapid time series analysis of massive archives. By providing intelligently structured cloud-optimized glacier velocity data, ITS_LIVE has reduced the demand for expertise in remote sensing and geospatial analysis, making the data more accessible to a broader range of researchers. By designing to meet user needs and enlisting the help of experts in modern cloud computing, ITS_LIVE has built an active and growing community of scientists that are generating new insights into why and how glaciers have responded to changes in their environment, giving clues as to what our future might hold.

ACKNOWLEDGMENTS

The ITS_LIVE user community for their helpful feedback. Pangeo community. Copernicus Sentinel data processed by ESA. Landsat data courtesy of the U.S. Geological Survey. Funding for ITS_LIVE provided by the NASA MEaSUREs program.

Luis A. López is a Research Software Engineer currently working with the National Snow and Ice Data Center in Boulder, Colorado. His research interests include remote sensing data workflows, cloud computing and open science. Contact him at luis.lopez@nsidc.org

Alex S. Gardner is the ITS_LIVE Principle Investigator and a NASA Research Scientist at the Jet Propulsion Laboratory in Pasadena, California. He studies the Earth's cryosphere (frozen Earth) with a particular focus on glaciers and ice sheets and their impacts on sea level rise and water resources.

Chad A. Greene is a NASA Research Scientist at Jet Propulsion Laboratory, California Institute of Technology. He uses satellite and airborne remote sensing to study the Earth's cryosphere, and is a former editor of the Proceedings of the National Academy of Sciences.

Joseph H. Kennedy is a Computational Scientist at the Alaska Satellite Facility in Fairbanks, AK. He specializes in building Big Data processing and analysis platforms for Earth data and studies the Earth system using HPC models.

Maria Liukis is an Engineering Applications Software Engineer at the Jet Propulsion Laboratory, Pasadena CA. With a specialization in scientific programming, she assists scientists in addressing computational challenges, while also contributing to software development for various missions and tools.

Mark A. Fahnestock is a Research Professor in the Snow, Ice, and Permafrost group at the Geophysical Institute of the University of Alaska Fairbanks. He studies glaciers and ice sheets with a focus on observational constraints on change

References

[1] Y. Lei, A. Gardner, P. Agram, Autonomous repeat image feature tracking (autoRIFT) and its application for tracking ice displacement, Remote Sens. (Basel) 13 (4) (2021) 749.

[2] A. S. Gardner, G. Moholdt, T. Scambos, M. Fahnstock, S. Ligtenberg, M. van den Broeke, J. Nilsson, Increased west antarctic and unchanged east antarctic ice discharge over the last 7 years, Cryosphere 12 (2) (2018) 521–547.

[3] K. Hogenson, H. Kristenson, J. Kennedy, A. Johnston, J. Rine, T. Logan, J. Zhu, F. Williams, J. Herrmann, J. Smale, F. Meyer, Hybrid pluggable processing pipeline (HyP3): A cloud-native infrastructure for generic processing of SAR data (2023).

[4] A. Fouilloux, annefou/metos_python: Working with spatio-temporal data in python (2018).

[5] E. Marshall, S. Henderson, D. Cherian, e-marshall/itslive: itslive tutorial updates (2022).

[6] L. Lopez, A. Gardner, M. Fahnestock, T. Scambos, M. Liukis, C. Greene, Y. Lei, J. Kennedy, B. Wallin, ITS_LIVE global glacier velocity exploration and analysis (2021).

[7] S. Hoyer, J. Hamman, xarray: N-D labeled arrays and datasets in python, J. Open Res. Softw. 5 (1) (2017) 10.

[8] A. Miles, jakirkham, M. Bussonnier, J. Moore, D. P. Orfanos, J. Bourbeau, A. Fulton, D. Bennett, G. Lee, Z. Patel, R. Abernathey, M. R. B. Kristensen, M. Rocklin, A. B. Awa, S. Chopra, E. S. de Andrade, J. Hamman, M. Durant, V. Schut, R. Dussin, S. Verma, S. Chaudhary, C. Barnes, J. Nunez-Iglesias, A. Sansal, B. Williams, B. Mohar, C. Noyes, zarr-developers/zarr-python: v2.16.1 (2023).

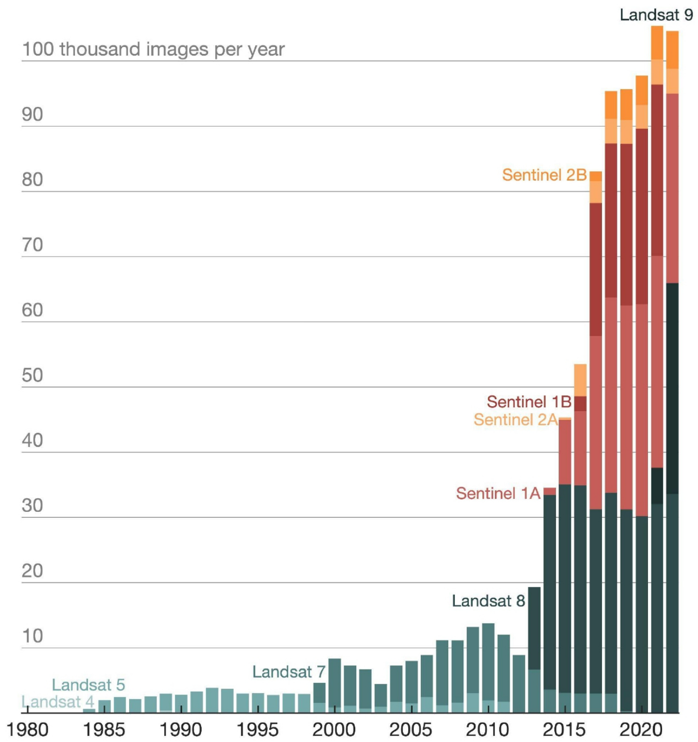

Figure 1. The satellite images processed by ITS_LIVE produce ice velocity measurements for all of the world's glaciated regions. Recent satellite launches have markedly increased the number of images that must be processed each year, and as image resolution increases, so do the computational demands of processing each image. Remote sensing data collection will likely continue to accelerate over the coming years, meaning cloud-native solutions will need to be adopted across the disciplines of Earth science to handle the deluge of new data.

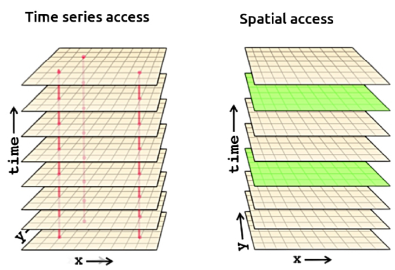

Figure 2. For fast time series data retrieval, we need our data chunks to be aligned and continuously stored along the time dimension. ITS_LIVE provides both regional mosaics optimized for spatial access (represented in green in the figure above) and time series data cubes using Zarr format.[10],[3]

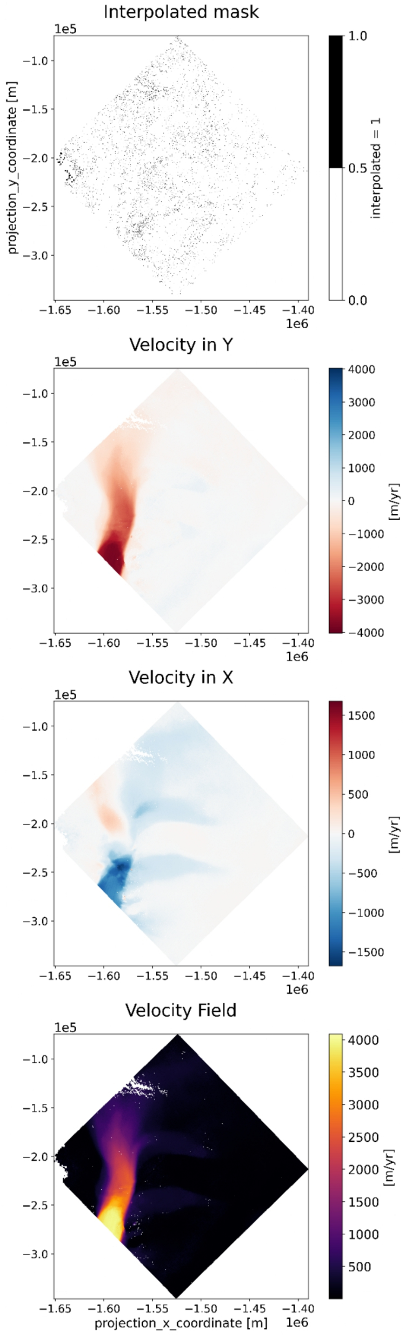

Figure 3. Velocity components of a single file generated with autoRIFT. Ice velocity is determined by measuring the displacement of features between two satellite images acquired days to years apart[11]

Figure 1: See[8,4,10]

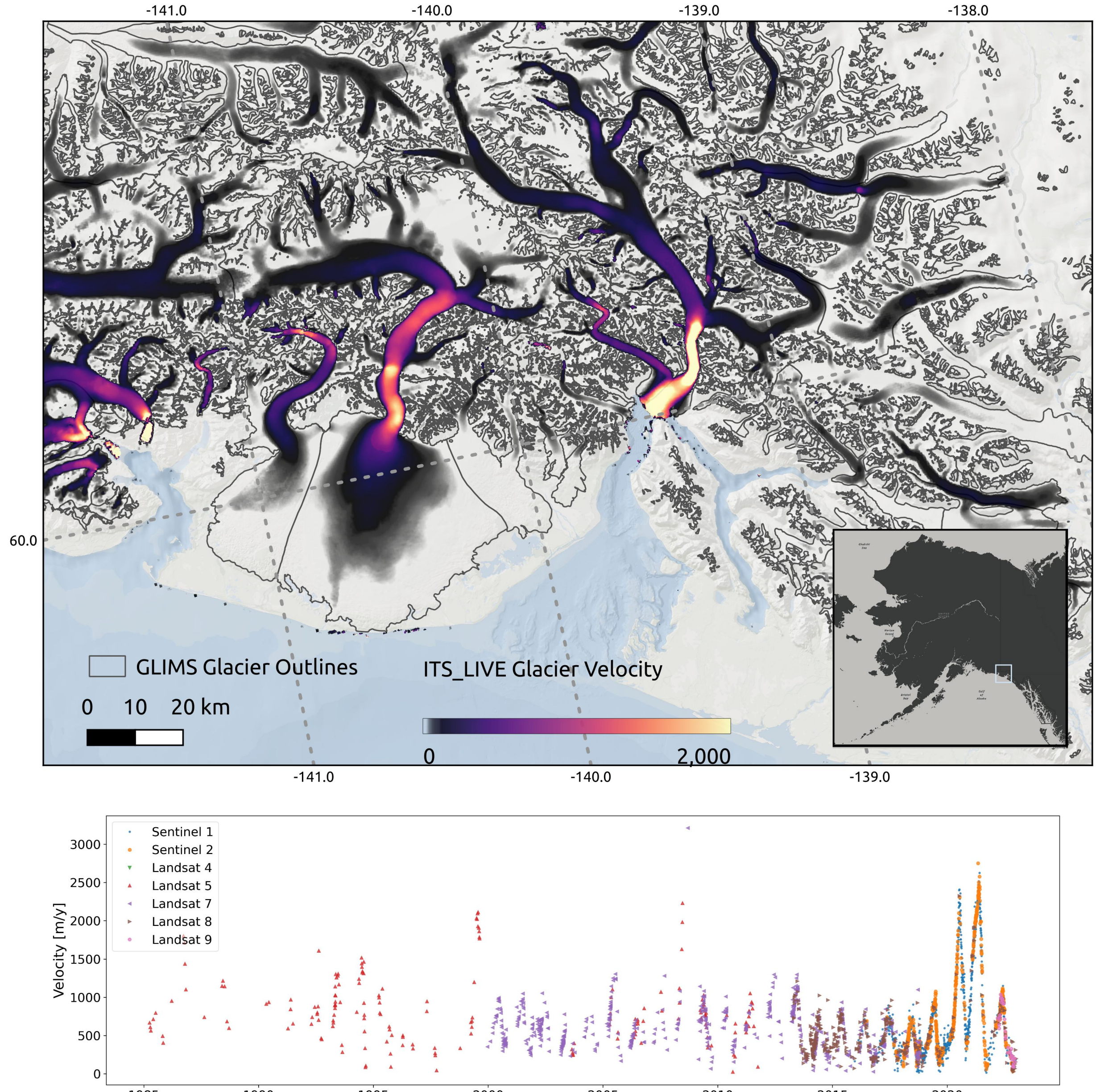

Figure 2: The Jupyter-based dashboard https://itslive-dashboard.labs.nsidc.org/ and web based widget https://nasa-jpl.github.io/itslive-web/ let us query time series of glacier velocity and export the data into familiar formats like NetCDF or CSV files. A single time series query can contain more than 200k data points. Despite the size of the dataset rendering these time series usually take less than 3 seconds to be completed thanks to the time-aligned chunking. Malaspina glacier (center) in Alaska is one of the glaciers that have shown a recent surge in activity (rapid acceleration). Near real time data processed by ITS_LIVE has helped scientists plan field campaigns to place in situ instruments to study the surge in detail.

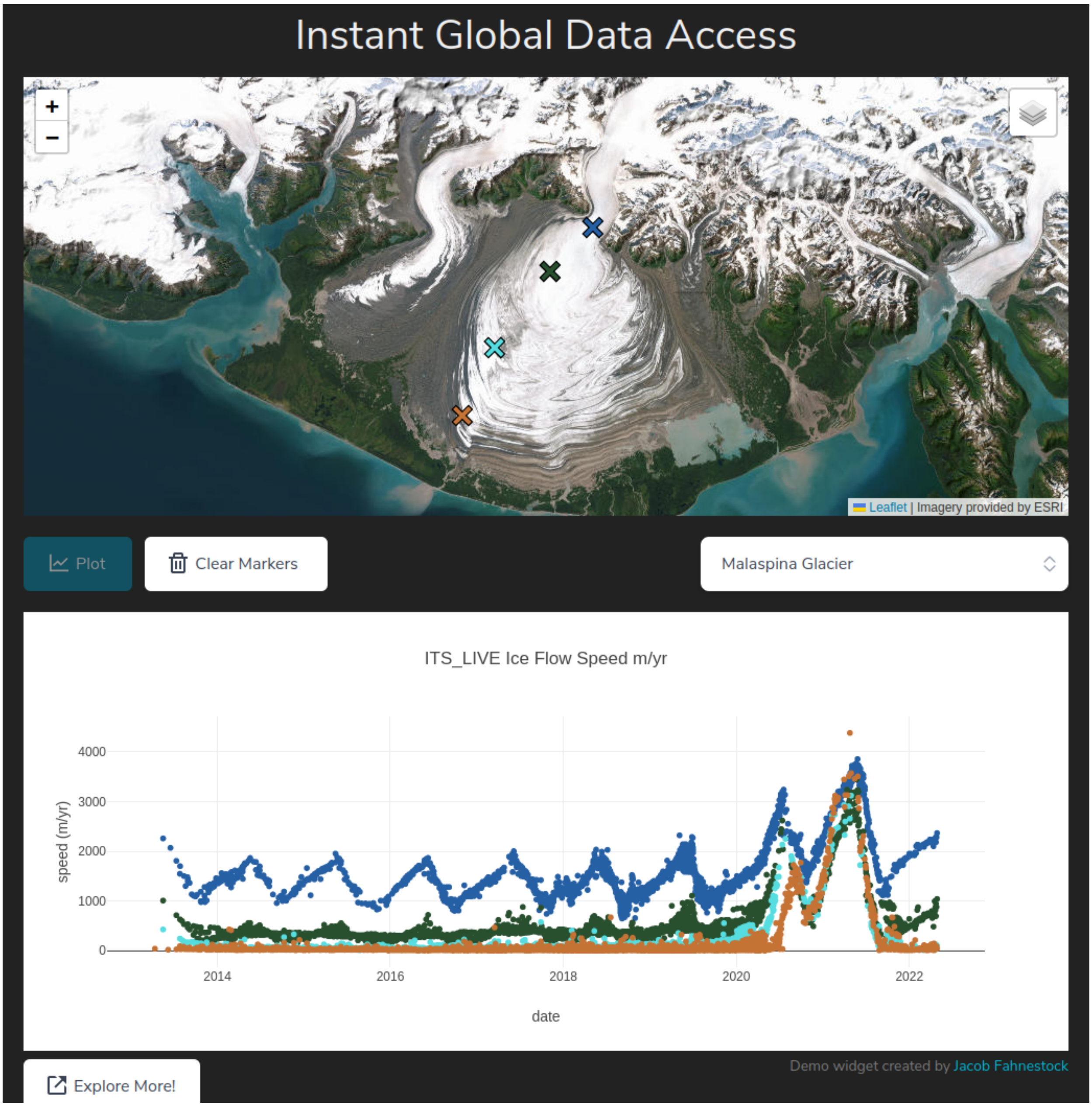

Figure 3: ITS_LIVE gives users who are not familiar with APIs and programming languages a way to explore the rich dataset and to begin analysis, regardless of whether they wish to work locally from a laptop or in the cloud. ITS_LIVE dashboards allow users to search, visualize, and download data without the need for a single line of code, and a client-side only prototype now provides most of the functionality of the Jupyter-based dashboard but without the need of any back-end server[11]

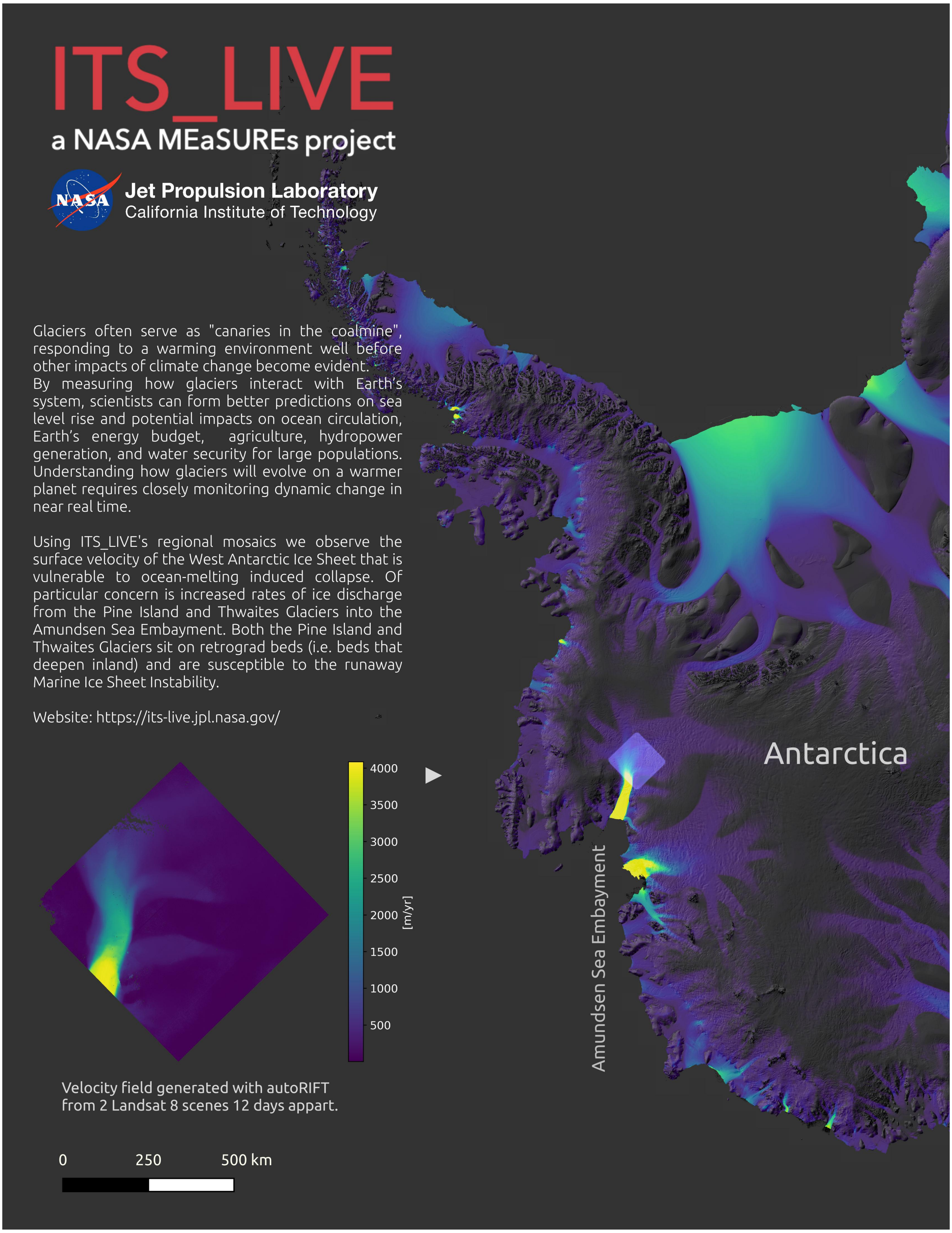

Figure 4: Info-graphic showing the surface velocity of the West Antarctic Ice Sheet that is vulnerable to ocean-melting induced collapse[12,13,14]

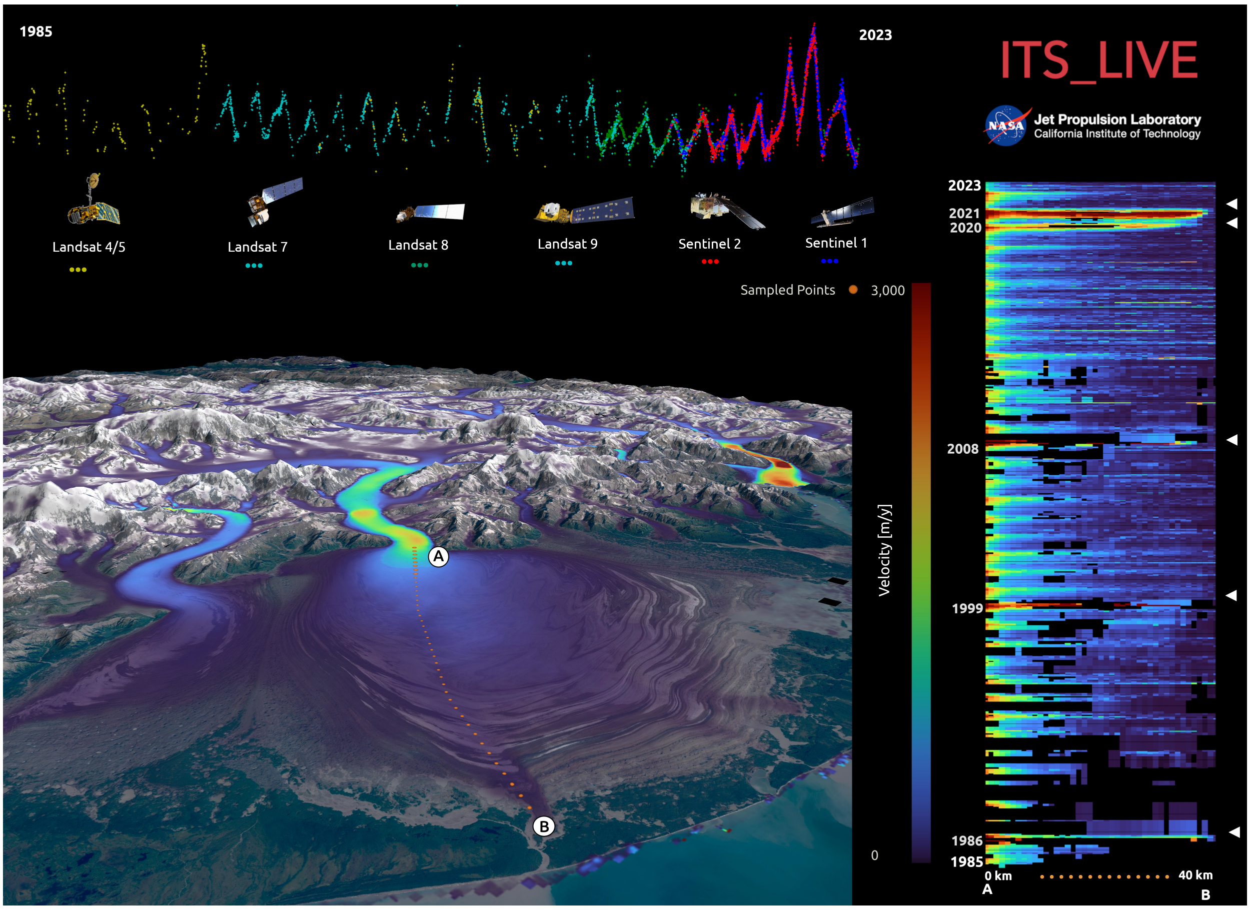

Figure 5: What looks like a frequency spectrogram on the right in this figure is a velocity profile along sampled points in Malaspina glacier in Alaska (above) , with the first point at the lower end of the ice (0 km at the left edge). In this plot the ice is flowing from the right to the left (A to B). One can see the seasonal cycle in speed (slowest in Fall) in the upper glacier, and the surges of the Malaspina in indicated years, including 2020 and 2021, when the entire glacier along this centerline reached speeds > 2500 m/yr, or 6.8 m/day.

[9] P. Rudiger, J.-L. Stevens, J. A. Bednar, B. Nijholt, Andrew, B. Chris, A. Randelhoff, J. Mease, V. Tenner, maxalbert, M. Kaiser, ea42gh, J. Samuels, stonebig, L. B. Florian, A. Tolmie, D. Stephan, S. Lowe, J. Bampton, henriqueribeiro, I. Lustig, J. Signell, J. Bois, L. Talirz, L. Barth, M. Liquet, R. Rachum, Y. Langer, arabidopsis, kbowen, holoviz/holoviews: Version 1.13.3 (2020).

[10] T. A. Scambos, M. J. Dutkiewicz, J. C. Wilson, R. A. Bindschadler, Application of image cross-correlation to the measurement of glacier velocity using satellite image data, 1992.

[11] J. R. Fahnestock, D. E. Dow, Mappin: A web native browse tool for the NASA JPL ITS_LIVE project's ice velocity dataset, in: 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (IEEE UEMCON 2023), New York, USA, 2023.

[12] I. Joughin, R. B. Alley, Stability of the west antarctic ice sheet in a warming world, Nat. Geosci. 4 (8) (2011) 506–513.

[13] C. A. Greene, A. S. Gardner, N.-J. Schlegel, A. D. Fraser, Antarctic calving loss rivals ice-shelf thinning, Nature 609 (7929) (2022) 948–953.

[14] K. A. Naughten, P. R. Holland, J. D. Rydt, Unavoidable future increase in west antarctic ice-shelf melting over the twenty-first century, Nature Climate Change (Oct. 2023). doi:10.1038/s41558-023-01818-x.
URL https://doi.org/10.1038/s41558-023-01818-x