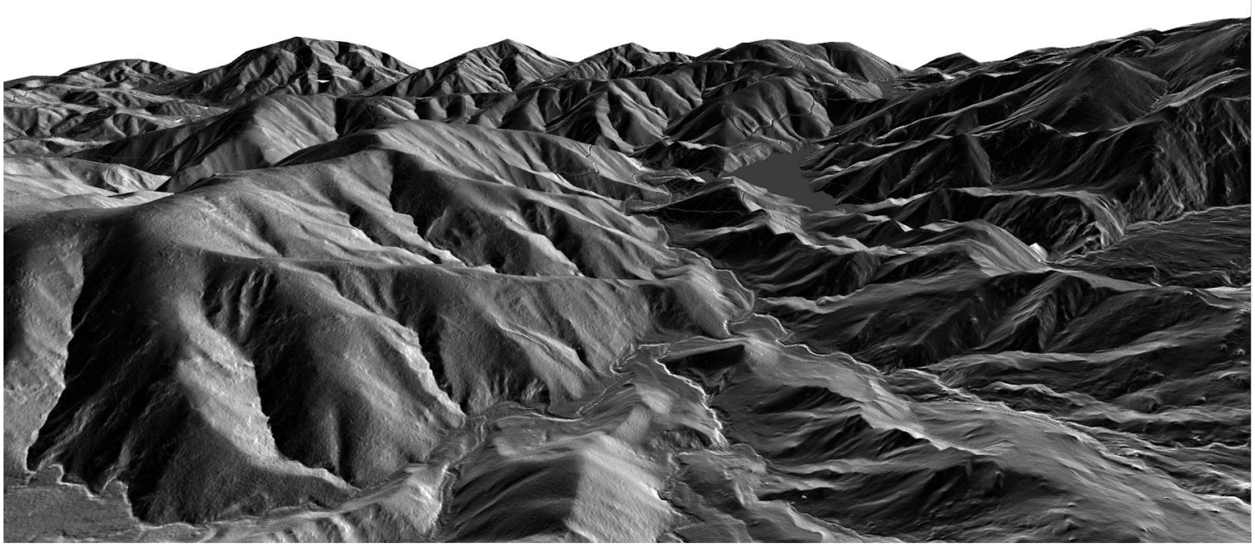
OpenTopography: Demonstrating Impact Through Audience Engagement

Juliana Casavan

Abstract

OpenTopography is a web-based cyberinfrastructure platform that enables public access to high-resolution topography data and processing tools. Over the past decade, there has been exponential growth in Earth-oriented topographic data through publicly funded programs. These data are not always easy to work with due to their massive size and complexity. OpenTopography makes these data easy to discover, access and process, thereby expanding their impact for research, education, and other applications. Over the last decade, the OpenTopography team has developed its community of users by strategically utilizing education and outreach, building from available resources, and listening to the needs of different stakeholders.



Credit: LINZ Data Service.

Origins

Mapping the earth's surface above and below sea level is an activity with far-reaching importance and many applications - from scientific research to planning and modeling for natural hazards, to engineering and construction questions for the public and private sectors. As technologies have advanced to more rapidly capture this data, the challenges of managing and processing these large datasets have become clear. OpenTopography was developed as a means of having a repository of research-oriented, high-resolution topographic data, both as a service to those who survey and collect it, and those whose work relies on using or analyzing it.

Though OpenTopography officially launched with funding from the National Science Foundation (NSF) in 2009, it first began well before that, evolving from the Geosciences Network (GEON) project led by PI Chaitan Baru at the San Diego Supercomputer Center (SDSC) at UC San Diego. GEON[1] was an NSF-funded project designed to facilitate collaborations between computer scientists and geoscientists to tackle earth science data challenges. At that time, Viswanath Nandigam, a researcher in PI Baru's Advanced Cyberinfrastructure Development Lab was tasked with developing a custom multi partitioned IBM DB2 database specifically for managing massive volumes of lidar datasets on the IBM DataStar cluster. Christopher Crosby was a graduate student interning with Dr. Ramon Arrowsmith at Arizona State University, Dr. Arrowsmith was one of the geoscience co-PIs on the GEON project. Chris Crosby arrived at ASU to begin working as a graduate student with Ramon, and had experience working with lidar (Light Detection and Ranging - remote sensing method that uses light to measure distances) data of the Northern San Andreas Fault from an internship at the U.S. Geological Survey. Lidar was a new technology at the time with obvious potential but also considerable technical challenges.

Ramon and Chris recognized the challenges of lidar and also the potential opportunities to tackle these challenges available through the GEON Project and the collaborations with colleagues at SDSC. After discussions and presentation of some simple prototypes built at ASU, Viswanath Nandigam, a researcher in PI Baru's Advanced Cyberinfrastructure Development Lab was tasked with developing a custom multi partitioned IBM DB2 database specifically for managing massive volumes of lidar datasets on the IBM DataStar cluster, SDSC's newest supercomputer at the time.

The simple user interface they developed - "the GEON LIDAR WORKFLOW (GLW)" - made it possible to process and analyze lidar data through a web browser using high-performance computing (HPC) and web based portal software. This turned out to be a novel approach to processing large volumes of geospatial data, with broader utility. After several presentations at

[1] GEON began with funding from NSF's Information Technology Research (ITR) program: https://acid.sdsc.edu/projects/geon.

meetings, and the addition of a couple of important early datasets to the system, they noticed that the GLW was being used by communities with large datasets that needed to be shared.

The big ah-ha came when the NSF EarthScope[2] project approached GEON to host their GeoEarthScope lidar dataset, ~5,000 sq km covering active faults in the western US collected as a community resource for research. GeoEarthScope was one of several early initiatives to collect this kind of data, and it was clear that these large community datasets needed a system for discovery, access, processing, and analysis to achieve their full impact. When it was time to think about the next round of development, the notion of a simple platform that would make it easy to share these large complex datasets with researchers and educators, stood out as extremely valuable, both for data preservation and access, and so OpenTopography was “officially” born.

Sustainability Status & Strategies

OpenTopography has enjoyed consistent growth and development since its launch. The platform functions as a two-sided market; its value relies on the continual addition of new data as well as on the demonstrated value of that data to those who can use it. Both sides of the equation appear to be working quite well. As of May 2020, over 120,000 unique users have run jobs via the OpenTopography portal (accessed OpenTopography compute resources to process or analyze data); and over 8,000 new users created accounts in the past 12 months. In addition, they have continued to add to the number and range of data hosted on the site, to date, they offer access to over 14 trillion data points, covering ~65% of the US.[3] These data come from NSF-supported projects, other federal agencies (including the USGS’s 3D Elevation Program), and international partnerships (e.g., Land Information New Zealand).

As an NSF-funded project created by and for researchers, demonstrating value to science is a priority. OpenTopography has done this by measuring the research that scholars have accomplished; there are 505 peer reviewed publications that utilized data and/or tools from OpenTopography, leading to 8844 citations in the literature, and an H-index of 46 .[4]

OpenTopography operates with a modest financial “footprint,” and has received several rounds of funding since 2009, primarily from the National Science Foundation. In addition to the 2 FTE this funding covers, the project has benefitted from stable partnerships with San Diego Supercomputer Center, Arizona State University, and UNAVCO, which contribute staff time, technical resources, and institutional support.

Today, the OpenTopography portal has 35,000 registered users with an average of over 650 new users registering for accounts each month. Not only does the team place a premium on driving usage of the platform, but they have also developed and maintained a dashboard of

[2] NSF EarthScope: https://www.nsf.gov/pubs/2017/nsf17511/nsf17511.html.
[3] OpenTopography analytics: https://portal.opentopography.org/analytics.
[4] Definition and origins of h-index: https://guides.lib.umich.edu/c.php?g=282982&p=1887449.

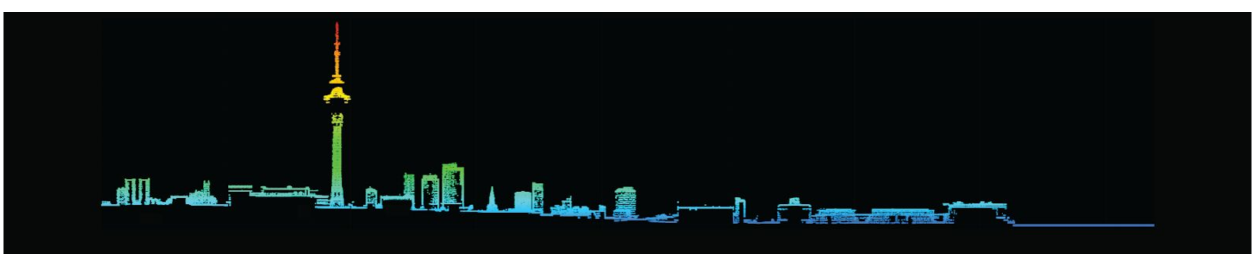
metrics and key indicators that helps them - and their stakeholders - easily measure the progress they are making and their impact.

There are several strategies the team has employed that have helped them to achieve this. The first is the use and emphasis on outreach and education, the second using modularity and available tools instead of building from scratch, and the third, listening to users and meeting demand.

Sustainability Strategy #1: Value of outreach and education

As OpenTopography was being developed, the team offered presentations and taught short courses at Southern California Earthquake Center to highlight the availability of lidar data covering active faults. This is a community that the leadership team was already familiar with and a part of, and this connection helped drive the first wave of users to the platform.

Among the tactics they have found most useful are attending events and conferences, which surfaces opportunities that they hadn't thought of before, like Data as a Service. During a discussion with a representative from a state agency that had recently created a program to capture Lidar data of their state. Chris learned that they would greatly benefit from a platform that stored, managed, and made their data accessible. With this new insight, Chris started attending other state geospatial meetings to better understand the problem. His hunch on this was not wrong, after talking with many geospatial representatives at conferences they had a better understanding of the problem state agencies would have building their own platform. To date, OpenTopography is distributing data for the states of Indiana and Utah, and Land Information New Zealand.



Profile through Auckland, NZ lidar point cloud data. Sky Tower at center left. Credit: OpenTopography.

The OpenTopography platform was built with the capability to capture detailed metrics from the outset which enables them to record user behavior on the platform. This has been an invaluable way to understand what the users are doing and who the users are so they can reach out to those communities to better understand their needs and prioritize services. They also use this information to build partnerships and gain insights for future development when connecting with leadership in these communities.

Sustainability Strategy #2: Building Efficiently: Not reinventing the wheel

Though it can be rewarding and valuable to build solutions from scratch, sometimes the best and most efficient solution is to reuse existing technologies that best solve the problem. The OpenTopography team has taken steps specifically to be flexible and have the ability to move quickly.

One example is the monitoring system that they have added to the platform that was originally built at SDSC, called Opal Toolkit as part of an NIH funded project “National Biomedical Computation Resource. By utilizing this software package on their platform, not only do they gather detailed analytics but they can also be instantly notified of failures or problems and have the capability to respond quickly. This may seem simple, but the real value to the user is in customer support, which strengthens trust in a service.

The team believes that excellent instrumentation of their own system is critical to providing responsive customer support. “Good metrics and instrumentation of the system helps us be responsive, which I think users genuinely appreciate. Getting an email from a real person that says ‘hey, we know that job failed, we fixed it, sorry for the inconvenience. Try it again now,’ goes a long way towards building loyalty,” says Crosby.

Another great example of OpenTopography’s efficiency is in using open source options when they are available. Part of their sustainability practice has always been using tools and resources that are openly available instead of building from scratch. This has enabled them to utilize the expertise of a broader community and tap into advancements that others have made. With short grant cycles and a lean team made up of individuals that have only a small percentage of their time dedicated to the project, it makes it a simple decision to use open source.

Sustainability Strategy #3: Letting Demand Lead: Looking Beyond NSF for Funding

OpenTopography, after 10 years of support from NSF over three funding cycles, is - like many grant-funded initiatives - still largely dependent upon core funding from NSF. Currently, this funding provides two full-time employees spread across seven team members, Crosby acknowledges that the grant-making space is increasingly competitive, and diversifying their funding strategy is important.

In addition to NSF grants, they have been able to capture institutional support from SDSC, Arizona State University, and UNAVCO. This includes resource support from SDSC with access to new technologies being developed that need to be tested. The most considerable non-cash support comes from the people themselves. As is the case with many Science Gateways, this is not their primary job or responsibility, and they rely on the passion of the entire team to make it successful.

Despite relatively consistent funding from NSF and institutional support, the team has taken steps to consider other options for funding. One area they are exploring involves data management for state and national agencies that have GIS and lidar data they would like to have stored and managed by OpenTopography, as an estimated cost for a state to build and maintain its own resource would be about $140 million.

When faced with the “offer” of ingesting a large amount of data, the OpenTopography team faces a challenge - while it might enhance the platform to offer more data, there are significant costs required to do so. In order to support all of these data requests, identifying new funding through other agencies or via revenue has quickly become a part of the conversation for the team. As scientists and researchers, they see the value in the data. However, when the data suppliers don’t have the financial means, the question becomes, will the OpenTopography user community benefit by making this data accessible via the platform, and is it appropriate for OpenTopography to host these data using funding from the NSF? Increasingly, OpenTopogrpahy is asking non-NSF data suppliers to provide resources to cover the costs of data hosting.

In a podcast by StateScoop, Phil Worrall, former Executive Director of Indiana Geographic Information Council (IGIC), stated: "Honestly (OpenTopography) made the difference between success and failure of the use of the data because it made it so readily available, and really broke down the barriers to having the normal mom and pop engineering firms and surveyors get access to the data in a form they can use.” [5]

Currently, the entire state of Indiana is available on OpenTopography, and they see engineers accessing the data to use in designing things like irrigation ditches. The question now is, what is the cost-benefit of being able to easily interact with these datasets for engineers, surveyors, and others that are in the business of profit. If for-profit sectors show a willingness to pay, this could become a part of the sustainability model for the project.

[5] StateScoop podcast: https://soundcloud.com/statescoop/gis-addressed-episode-6-3d-elevation-data.

Closing Thoughts

While OpenTopography is a traditional grant-funded effort, the team has taken steps towards testing the viability of their gateway as a sustainable product. The team has done this by building a robust and devoted user base, and by remaining attuned to the market. The team has been able to keep its focus on propelling the platform forward and making decisions based on the needs of their users balanced with what is feasible for the project. While working with minimal funding and a lean team, OpenTopography has had a long life so far with their eyes set on growth.