Stop sharing data via hard-drives, it's time to modernize your workflows.

Sharing Geospatial Data Effectively

BY KEVIN MILLER, CHIEF TECHNOLOGY OFFICER

Whether apparent or not, almost every data set and application today have a geographic component built in. It could be as simple as an X,Y location of your computer mouse within a web-application, the monitoring of assets and their locations, construction progress over time including how components fit together, or perhaps a more advanced self-driving car that needs to make sure it isn't running into that parent pushing their baby across the street. All of these examples include some sort of geospatial information, a standard that identifies the data having an implicit or explicit association with a location relative to Earth.

Over the past several years, the amount of spatial data available has exploded. Before the internet, thirty years ago, no one really had any detailed sense of where things were located except surveyors and geographers. It was extremely costly, and in many cases impossible, to track your assets based on location. Geographic dataset coverage was a challenge which was solved when internet services started becoming available in remote areas and not just in major centers. Within the last 10-15 years, we have seen systems start talking to each other, allowing users to bring increased information together, including location. Device coverage was also an issue which we are now solving as systems are getting smaller and less costly, allowing more and more "things" to record and provide all sorts of data. The common term for this development and architecture is known as the Internet of Things (IoT). Along with this, there are now even some companies working on building microchips that can essentially capture 3D points (like LiDAR) – small enough to fit within a mobile phone! There remains the question of how should a company manage and share all of this information from all of these devices to the rest of their team/stakeholders?

Our team at SOLV3D, deals with extremely large and complex geospatial datasets. Most of the time, our clients' data lives on computer servers, or hard drives somewhere and is utilized by one person or a small team of people (usually within the same organization), preventing this typically expensive data from being shared and used effectively. We are looking to change this and provide a method for sharing the information effectively over the internet.

Regardless of the sharing method of choice, many of the organizations we work with aren't in the day-to-day business of developing software or managing big IT software projects. This means they simply do not understand the architecture behind hosting and sharing data of this magnitude.

After speaking with dozens of individuals at companies throughout North America and Europe, it seems there some consistent thoughts in regards to the needs of users:

- They typically do not need to access all of the project data sets on a daily basis. Rather, they just need to access sub-sets or single tile/polygon of an area.
- They do not want to wait in order to see the latest data in an area. They want to be able to see it right away.
- They want to compare versions of data for the same area to see how the area has changed over time.

To meet these needs, the data needs to be:

- On the internet, accessible from a web browser and no special software needed.
- Organized, or merged, in a way that a user can use a map and easily find out their data coverage.
- Organized in a chronological fashion, and accessible with associated metadata that lets a user know what they are looking at, when the data was recorded, and allows the user to easily compare datasets of various vintages.
- Easily accessible, to allow analysis on the data such as measurements or comparisons.
- Downloadable for use in other off-line programs.
- Provided in the format of a known common file format and properly georeferenced.
- Accessible by as many users as needed, and not restricted to those with technical expertise.

While the hardware and skill sets of data collectors have evolved rapidly over the past number of years, to the point where they can collect data with millimetre accuracy and properly geo-reference it even if it is several kilometres under the ground, they often revert back to archaic practices when they attempt to deliver and disseminate this data. Typically, their data ends up on a USB hard-drive and mailed across the country to their client! Let's not get into the problems that can arise when a hard-drive is mailed across the country, but trust me, you're definitely taking a risk doing so.

We also see most data collectors employing different workflows for each hardware type they employ. This results in datasets that are all over the place in terms of consistency, especially when it comes to combining this data with other datasets. All of that hard-won data accuracy is for nothing if the data is lost in transit, or if the data is not consistent from dataset to dataset.

By sharing data over the internet, data collectors can dramatically reduce the risk of losing their data when they send it to their clients. This method also usually means that their clients can access their data immediately, and not several days later.

Sharing geospatial data over the Internet does come with its set of challenges however:

- Too much data can slow things down. Data typically needs to be formatted, indexed or split up to allow for quick access.
- Data needs to be in a common format. To be accessible by many people, the data needs to be organized and presented in a format that is easily consumable – i.e. not everyone has a high-end computer running some CAD software.
  - This all boils down to the user experience on how to visualize and interact with the data.
- The data needs to be accessible. Systems should be in place to allow people to collaborate/comment/work with other team members without too much hassle
- The data need to be secure, protected and only accessible to the people who have permissions to access it.
- The data needs to be available for users when it is needed. You can't have any downtime, or have the data slow to access.

As an organization, you may choose to take on these challenges yourself, however, this typically involves forming an entire professional development team. Instead, I recommend looking at companies or products that specialize in dealing with and managing these types and magnitude of data. By doing so, as they expand their product offerings to solve use-cases for their users, they can also take advantage of these features.

Companies that specialize in creating professional software, have the team and expertise to help you effectively manage your data. They are the ones that can help you deal with different types of data, along with making it easier for you and your clients to collaborate.