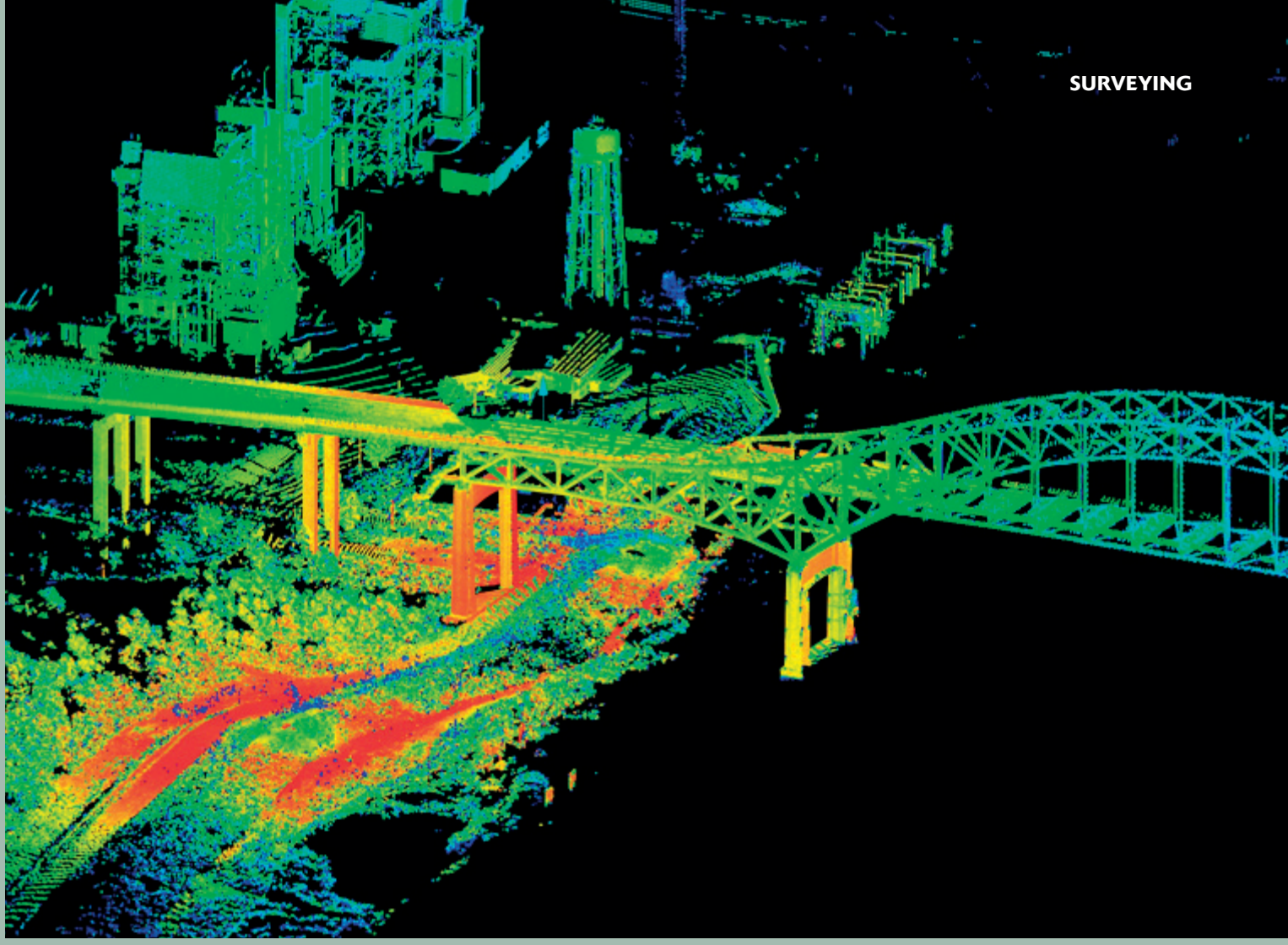


THE CLOUD BRINGERS

THE AMOUNT OF LIDAR DATA HAS GROWN RAPIDLY, RESULTING IN MANY NEW APPLICATIONS – AND CHALLENGES. JARED DOMINGUEZ AND ROBERT PARKER DISCUSS THE STRENGTHS AND WEAKNESS OF THE FILE FORMATS AVAILABLE IN WHICH TO STORE IT

LiDAR data initially comes from the scanner as a cloud of points and although it will later be used to make other products like contours and DEMs, it is still important to keep the point cloud. Often, these point clouds are written to standard ASCII files or the 'LAS' format. Alternatively, some systems use a proprietary format. What are the advantages – and disadvantages – of all these formats?

ASCII

ASCII (American Standard Code for Information Interchange) is a published, standard code that uses binary numbers to represent common characters. Some common file extensions for these files are .txt, .csv, and .xyz. LiDAR data saved in this way is human-readable and can be opened in standard text editors.

Each point is defined as a set of values in a line of the file and each value is separated by a delimiter. For example, a row of data could be as simple as x, y, and z coordinates separated by a comma, with a new line for each set of coordinates. Sometimes, the files include one or more header rows at the top to clarify the types of values included in the file.

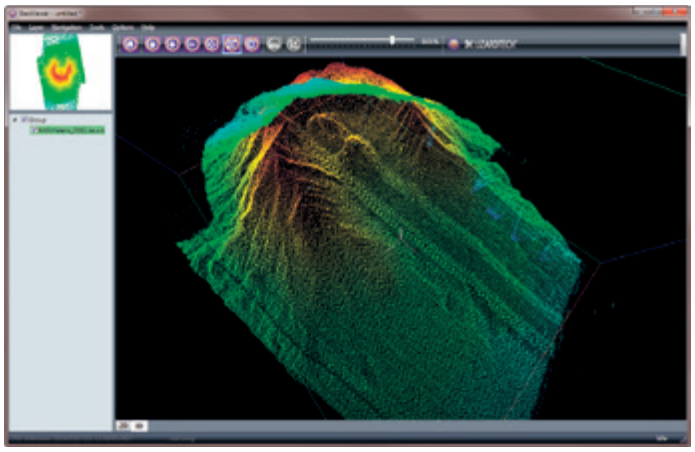
Each individual letter or number takes up eight bytes. This quickly adds up, even if the dataset is not particularly large. In comparison, without using the ASCII code, it is possible to represent each x, y, and z value in its entirety with just 16 bytes. Additionally, read and write performance can be extremely slow because most viewing applications attempt to load the entire contents of an ASCII file into memory. For these reasons, users are more likely to use a 'binary' file format such as LAS for their work.

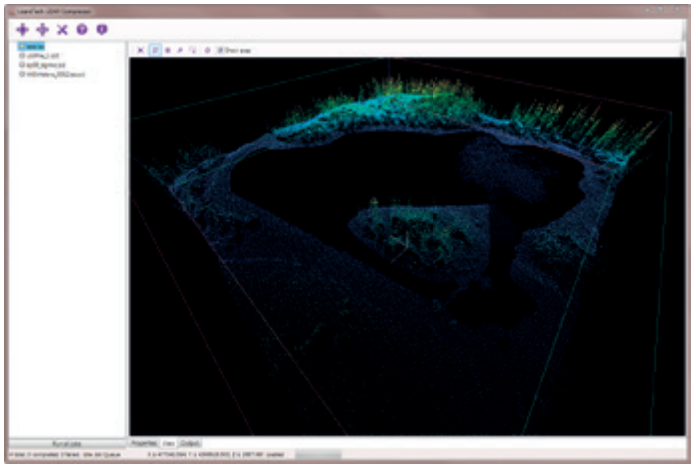
LAS

LAS is the most commonly used format. It is an open standard and can be read in virtually any application that supports LiDAR. It was created as a binary file format for the interchange of three-dimensional point cloud data and so generally contains x, y, and z coordinates. However, it can also contain other data, such as the number of returns, GPS time and colour values.

There are a few different versions of LAS available, from LAS 1.0 to LAS 1.4, and LAS 2.0 is in the works. As it is considered a de-facto standard, overall support for this file format should continue into the foreseeable future, providing users with a reliable way to share LiDAR data.

LAS provides several advantages over ASCII, including smaller file sizes and faster load times. However, there are some challenges with





LizardTech's LiDAR Compressor software is the only way to generate MG4 files

LAS, including the inability to realistically save extremely large point cloud collections in one file. Traditionally, users have separated large datasets across multiple LAS files, called 'LAS grids'. While this can minimise the memory footprint of LiDAR data, data management becomes more complicated.

LAZ

LAZ is an open source lossless compression for LiDAR developed by Martin Isenberg at RapidLasso. As LAZ is completely open, it can be implemented on any operating system and in any programming language for free.

The industry has taken a keen interest in LAZ and many popular applications have added read and write support for the LAZ format. Although any software application can potentially write to LAZ, many users use LASzip, also developed by RapidLasso, to create LAZ files. A large number of respected organisations, public and private, offer LAZ downloads of their products, including the USGS, NOAA and the US Army.

LAZ appeals to many in the industry for a variety of reasons. You get access to powerful, useful technology with no licensing fees as well as the ability to modify and improve the software to meet your needs. However, it takes a certain level of technological ability to take full advantage of what the technology has to offer, and users relying on the community for support or updates could have to deal with unexpected costs.

Overall, the technology has gained vast acceptance and has quickly become a popular standard for LiDAR compression.

MrSID Generation 4 (MG4)

MrSID Generation 4 (MG4) is a binary file format that attempts to resolve some of the problems of storing and accessing large LiDAR collections. MG4 is able to represent point clouds losslessly with significantly smaller file sizes. It also creates an optimised index of the entire point cloud, allowing for extremely fast extraction and visualisation. Since file size is not an issue, many users create entire-area MG4 mosaics using LAS grids as sources. This enables them to easily manage and extract from a mosaic of their entire collection, rather than find and gather individual files to make up an area of interest.

Since MG4 is a proprietary format, there are some restrictions. For example, there is only one way to write an MG4 file – LizardTech's LiDAR Compressor software. This means that other software applications can view and access MG4 data but cannot create and edit it.

Optimized LAS (zLAS)

Esri released Optimized LAS (zLAS) in January 2014. Like MG4 and LAZ, it compresses LiDAR data losslessly, may be used directly without decompressing the file, and creates a spatial index and statistics that result in an 'optimised' file. Compression to zLAS is offered for free through Esri's LAS Optimizer software. Esri recently made it possible for developers to add support for zLAS to applications. However, there are still few companies that support reading or writing to zLAS files.

Despite the newness of this file format and its similarity to LAZ and MG4, zLAS is definitely a format to watch in the future. Esri's prominence means its efforts with LiDAR will surely have a big impact. Some in the open source community are worried that the proprietary format might overshadow the open source accomplishments of LAS and LAZ but Esri has indicated a willingness to figure out a path forward with the community.

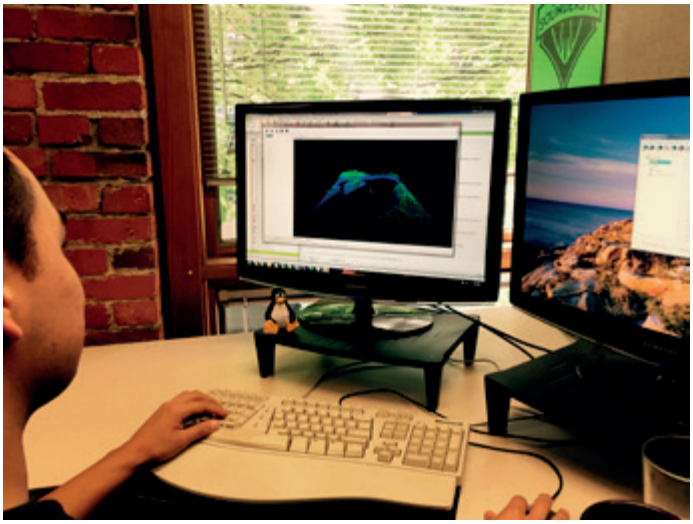
Conclusion

Of all the file types above, most people consider LAS to be the standard file format for LiDAR because it is open and has been accepted everywhere for many years now. But MG4, LAZ, and zLAS have shown that there are more efficient ways to store and access LiDAR data. MG4 and LAZ have been around the longest and so have been adopted by more users in the industry. Although zLAS is new, it should see quick adoption as well.

More testing is needed to effectively compare these three formats, but in general, they accomplish the same goal of smaller file sizes with fast, efficient access to the data. Users are encouraged to try each format for themselves and examine the benefits.

MG4, LAZ, AND ZLAS HAVE SHOWN THAT THERE ARE MORE EFFICIENT WAYS TO STORE AND ACCESS LIDAR DATA

Jared Dominguez is a technical writer and Robert Parker is a sales engineer at LizardTech (www.lizardtech.com)



A LizardTech employee compresses LiDAR data with LiDAR Compressor