

WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD?

WHEN IS A STANDARD NOT A STANDARD?

WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD?
STANDARD? WHEN IS A STANDARD NOT A STANDARD? WHEN IS A STANDARD NOT A STANDARD?

KML IS SUPPOSED TO BE A STANDARD BUT GLITCHES IN IMPLEMENTATION MEAN INTEROPERABILITY IS STILL AN ISSUE. **DANIEL BALOG** LOOKS AT THE POTENTIALLY FATAL CONSEQUENCES AND WHAT CAN BE DONE TO FIX THE PROBLEM

There are multiple meanings of the term 'bad data' in a variety of contexts, but one broad definition in the geospatial context is 'any data with an undocumented or badly documented specification'. The ramifications of this kind of bad data can be severe, leading to incompatibility with mapping software, lost revenue and operational downtime.

In the geospatial world, we rely on a rich variety of data formats, ranging from the incredibly simple (like GeoJSON) to the very complex (like domain-specific data for military and aviation such as MIL-STD-2525 and AIXM 5.1 respectively). However, because there are many different data formats that are produced and coded in so many diverse ways, developers and end-users can find themselves in a challenging situation where they have access to meaningful data but are unable to analyse it properly, because it is produced in an invalid manner or has been insufficiently coded. This can render it useless outside the proprietary software where it is produced.

The developer's dilemma

Keyhole Markup Language (KML), an XML notation for geographic annotation and visualisation, is one of the most prevalent formats for geographic data. It was (as the name would suggest) built by Keyhole, which was acquired by Google in 2002. KML is also the most common format used by Google

Earth, which was originally called Keyhole Earth Viewer. KML files specify features such as place marks, images, 3D models and technical descriptions that can be displayed on maps in geospatial software such as Google Earth. In its most basic form, a KML file specifies a location's longitude and latitude, although the view can be made more specific with other data, such as tilt, heading and altitude, which can define a camera view, a timespan or a timestamp.

As it developed, Google Earth saw overwhelming success and popularity, with both businesses and hobbyists using it. However, this popularity led to the creation of bad data, which has had implications throughout the industry. KML as a standard has always been open, with a semi-ambiguous, human-readable description of what a KML file should contain. But parts of the KML specification were wholly specific to the implementation of Google Earth. This means that data providers have become reliant on these specific parts, which only conform to Google Earth. This ultimately led to people relying on data without even realising they weren't conforming with the spec.

For KML specifically, there are different mistakes a developer can make that may have knock-on effects. If you make a mistake at an XML-level, you can create an XML tag that's not properly closed. There's also the possibility of XML and KML schema-related issues. An

XSD (XML schema Definition) specifies how to formally describe the elements in an XML document – basically the rules of what you can do within one XML spec. One KML-specific XSD says that if you have a KML root-node, it has to contain either one document child node or one place mark child node. If non-specialists were developing data, they wouldn't know how to do this, so would dump multiple documents or folders into the root node, which would add different entries. Google Earth is able to work even if rules are broken, and because of Google's dominance, other geospatial software developers have to work around and fix poor KML.

An awkward situation

Many developers are primarily building on Google Earth, so are unaware of how it has shifted data standard goalposts. It's arguably not the fault of the people producing data on Google Earth because issues grow organically from the software itself. Google Earth's flexibility pushes users towards a very forgiving format where you can put anything in and whatever comes out is the official standard.

This creates an awkward situation for developers who are conscious of bad data – either stay true to the standards and miss out on the features that the bad data has built on top of the specification or feed into the bad dataset to conform

```

    <!-- image height & count < 2) {
    <!-- image height : data.$image.outerHeight();
    <!-- image width : data.$image.outerWidth();
    <!-- image height;
  }
  <!-- image width;
  <!-- image height;
}
  <!-- image height;
}

```

with common (yet incorrect) practices. KML specifications do exist, but following these word for word means, in most cases, that you would be unable to load and use many online KML files. The reason behind this is because many of the people who create the data only test it against Google Earth.

Interoperability headaches

Ultimately, end-users just want to be able to use their data in whichever analysis or visualisation programs are required to get whatever work they are working on done. However, the ramifications of bad data can be financially costly and result in hefty compatibility issues. Interoperability is the main reason that the bad data issue must be solved. This has been backed up by the

USERS DON'T WANT TO WAIT FOR A NATURAL DISASTER OR NATIONAL SECURITY INCIDENT TO DISCOVER THAT THEIR KML DATA IS INCORRECT

not-for-profit Mitre Corporation, which in KML Best Practices for Interoperability stated that 'data expressed in the KML format can support a variety of needs, including emergency response' and 'data published in the KML format needs to be interoperable in applications and systems, including Google Earth... limited budgets require that best practices be adopted in order to maximize cost savings'.

Users don't want to wait for a natural disaster or national security incident to discover that their KML data is incorrect and can't be used in mission critical piece of software.

Collaboration for clean data

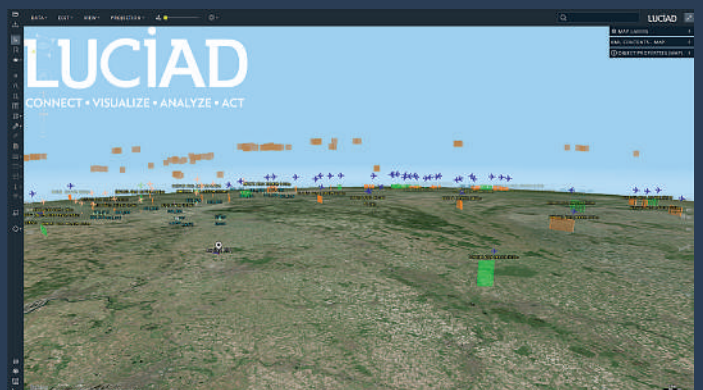
KML was adopted as an international standard by the Open Geospatial Consortium in 2008 to assure its status as an open standard. This means not only having documentation but also operational tests that developers can use to verify their own code. Code is tested and then receives a compliancy percentage, and as soon as the code reaches 100% compliance, the developer receives a badge that they can put on their website or product, assuring users and prospects of its data quality. If organisations participate in developing standards, then they automatically have better support for them. More importantly, participation from a wide array of industry players means that decision-making is not dominated by just one company, encouraging competition and therefore enhancing innovation.

To catch bad data before it gets out of hand, we need transparent cooperation to create proper testing and better standards. This will help to broaden what KML data can do (for example, adding 2D onto its existing 3D capabilities) and will improve the overall quality of the data format. KML can be vitally important in emergency situations, so must be ready to be used wherever necessary when disaster hits.

Daniel Balog is project lead at Hexagon Geospatial (www.hexagongeospatial.com)



St Peter's Church in Leuven, Belgium, visualised in 3D using KML data. Flight paths are tracked in the background



KML data can be used to track flight paths in real-time