Why Open Data is not as Simple as it Seems

Open data is undoubtedly a great movement in the world of GIS. OpenStreetMap and Copernicus are just two examples that open many doors and facilitate new applications. There are, however, also various challenges associated with open data. From technical obstacles to commercial objections and political concerns, open data is a much more complicated subject than some of us think. Why is open data still a topic of lengthy discussions?

The concept of open data is not restricted to geo-information, nor is it something new. Open data has its origins in the scientific world in the 1940s, long before the rise of the internet, when Robert King Merton explained the importance of sharing scientific results. He suggested making the results of scientific research freely available to all, in order to stimulate knowledge growth and innovation. That concept still applies today and has proven to be very successful in, for example, the domain of open source software. However, open data is much more complicated than simply collaborating on work and sharing results to help humanity move forward. There is a lot more at stake.

Common Good

One of the reasons open data has become the topic of many discussions is because the motivations and concerns for making data open to the public can be very different in nature. There are protagonists of open data that have similar motivations as Robert King Merton once described. Inspired by the concept of open source software (which is a clear win-win mechanism for software development), they consider open data as a common good that stimulates collaboration and innovation. OpenStreetMap is a beautiful example of open data that is co-created in a similar way as open source software and knowledge bases such as Wikipedia.

Transparency

There are also advocates of transparency and openness, who believe that data should be available for everyone to view and use. Their motivations are mainly political and apply for governmental agencies and (semi-)public organisations. For example, one could state that data owned by the public sector has been funded by the public and should thus be freely accessible to the public. Another political motivation is to promote openness and transparency in an attempt to avoid or solve distrust in public-sector organisations. The majority of these organisations are still in the process of opening up their data treasure chest, with now still only a limited number of datasets being available.

Letting Go

As motivations for joining (or trying to avoid) the open data movement differ greatly in terms of their origin and goals to achieve, it is no wonder that open data is subject to a lot of debate. And as if this is not complex enough, there are also various other challenges to overcome, ranging from ethical to practical. For example, there is a large grey zone on what data should actually be shared and what should remain private for whatever reason. Many organisations struggle with this question. In a time when openness has become the default, on what grounds can you decide not to share a dataset? Put simply, open data is not just about throwing your data set out in the open. Open data is about letting go. Many organisations fear the concept of open data – for all or only a few of their datasets – simply because they are afraid someone will misinterpret or misuse the data. What if someone creates a map based on your dataset, distributes this map, and someone else makes a fatal decision based on an error in this derivative product? Sooner or later, this will raise the question: who is accountable? Some organisations thus decide to play it safe and keep some datasets behind locked doors.

Return on Investment

A commercial objective commonly heard is that data represents value and costs money to produce, so it must be paid for. This point of view is understandable for data products made by private companies – especially those that build their business model on data and information. For public organisations, such a perspective typically raises the discussion: shouldn't this data be available at no cost, since citizens have indirectly already paid for it? Fortunately, more and more public-sector data is being made available in various countries these days. From an economical perspective there are also advantages of open data. Making public data accessible provides opportunities for commercial exploitation, which can create an indirect return on investment to the government.

Accessibility

Finding the right open data can be quite a hassle, even when correct metadata is present. This brings me to, in my opinion, the biggest challenge of open data: its discoverability and accessibility. A data portal that offers all available data does not exist (yet), so users need to look for data in various online data repositories – each with their own policies, (often complex) user interfaces and limitations. In my view, there is still a lot of work to be done on this side of the data chain. There are a couple of online portals that have made a serious effort to disseminate large amounts of data. Examples for international use include Copernicus Open Access Hub, Natural Earth Data, USGS Earth Explorer and Esri's ArcGIS Hub. Many of the online portals have an intimidating user interface that probably discourages untrained and non-geo professionals to even have a go at it. Though some of them serve the professional GIS and remote sensing communities well, they will never reach a wide audience, simply because they are too complicated to understand and use. This greatly limits the potential of open data.

Knowledge Intensive

No matter how discoverable and accessible, some datasets remain 'knowledge intensive'. This means that only a limited number of users have sufficient technical background to understand how to process, analyse and use them. Think of map layers showing certain specialised agricultural indicators, or remote sensing imagery such as Synthetic Aperture Radar (SAR). Those remain the domain of experts, at least for now.

A Great Development

Open data is a great development, but much more complicated than it seems. Access to open data is only possible by solving the sum of technological, economic, political and communication challenges. Although we are certainly making progress, there are still a lot of things to do, many problems to solve and interesting discussions to have before we can really unleash the power of all the data that is out there and potentially open.

This article was published in GIS Professional October 2017