EMBRACE THE CHAOS

LACK OF METADATA FOILS MANY SPATIAL DATA INFRASTRUCTURE INITIATIVES. KRIS GOODFELLOW AND BRIAN GOLDIN ARGUE THAT WE CAN LEARN FROM THE WEB HOW TO FIND AND IMPROVE MAP DATA

It's time to reimagine the spatial data infrastructure (SDI). While web mapping has exploded since 1993, and millions, if not billions of euros, have been invested in the creation of SDIs, the usage has languished, the data is undiscoverable and too much valuable content is inaccessible. As leaders of SDI initiatives, we want to swing open the windows and let in some fresh ideas.

Before moving forward, let's take a spin in the way-back machine. It's 1993. Remember 'You've got mail', the hiss of a 56-bit modem and the Mosaic browser? Around this time the US National Research Council proposed the idea of an SDI, according to an Esri white paper on the subject. The goal was to move GIS away from the world of print maps to a new, networked digital one. SDIs were envisioned as a way to promote seamless coverage, broad sharing and creative implementations. An SDI, it was postulated, would foster usage and increase the overall value of a GIS. This was a bold and important vision.

What followed was a top-down approach to building the solution. Technological frameworks, base layers and metadata standards were defined. Standards were thoroughly debated. This was all done to establish best practices, so the SDI would be accurate, usable, discoverable, and shareable. And this foundational work was a good start.

Changing times

But the world has changed, and for all of the hard work and debate, the usage of SDIs simply has not followed. An article by Nama Raj Budhathoki, Bertram Bruce and Zorica Nedovic-Budic entitled 'Reconceptualizing the role of the user of spatial data infrastructure' explored the subject in *GeoJournal* in 2008. It stated: 'There are indications that current SDIs which follow this top-down model, are under utilized.' The authors gave examples of SDIs in Europe, India and Nepal with limited uptake, and attribute this problem to the 'passive role of the users and inadequate attention to the users' work practices'.

Meanwhile, as of 2011, Google Maps had some 2,300 mash-ups and 1.1 million users of its the API. In 2013, OpenStreetMap has had almost 900,000 edits from some 50,000 users and is accessed by millions of people on FourSquare via another hot startup MapBox (see page 24). These alternative models of an SDI have been used for everything from finding a good pint to helping victims of Hurricane Sandy (http://mashable.com/2012/10/29/google-crisis-map-hurricane-sandy/) find shelter or victims of the flooding in Indonesia get help. (http://blog.google.org/2013/01/responding-to-severe-flooding-in.html).

In the US, the bedrock of this model is frequently free, publicly available data. That data is enhanced by the efforts of private enterprise. But it is the citizen reporter or GIS hobbyist who provides the finishing

Voyager allows users to instantly find map documents, datasets, web mapping services and other geospatial resources on the desktop or across the enterprise
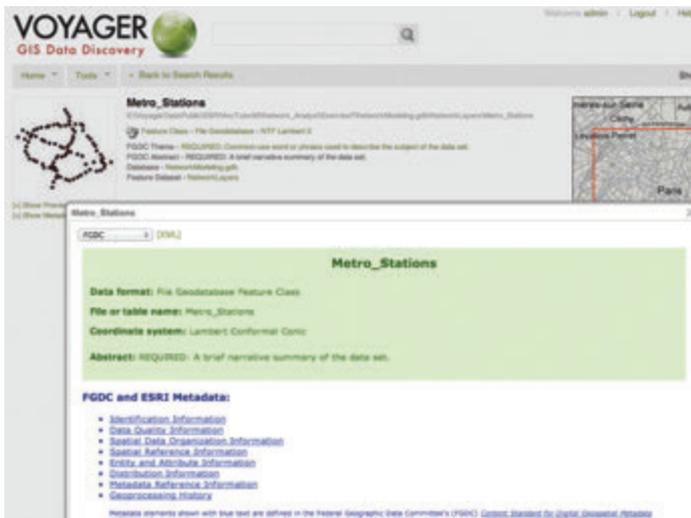
Documents can be found in Voyager using spatial or keyword searches, regardless of whether the file has metadata associated with it

On Voyager detail pages, users can see the relationship between a layer and the map documents that use it

touches. This informal network achieves many of the goals of an SDI – a platform for creating valuable map applications we could not yet imagine. Further, it fosters a community of editors who work to maintain the accuracy of the base map and do so for free.

Our challenge to you

So our first challenge to the SDI community is to expand your notion of who your users are and what your role is in this ecosystem. Perhaps you can achieve your goals by becoming an active player in a much broader community. Consider that the Arab Spring was driven by millions of tweets, not hundreds of newspaper articles. There are citizen reporters on every corner. In parts of Africa, there are more people with mobile phones than access to fresh water. There are thousands of developers building map applications who don't know a thing about GIS. These people may well be you greatest untapped resource.

Furthermore, consider the role of public-private initiatives in this open-source world. Should you be hosting a portal that can be used to layer data into ArcGIS, Google Maps or OpenStreetMap or all of the above? How do you expose an interesting analysis to GIS hobbyists? Will they ever find your portal? We would suggest that, yes, you need to be the owners of a base map with critical infrastructure and your expert GIS user should create enlightening map analyses. But equally, your GIS professionals should be active participants in these civic, mapping playgrounds.

Immediate gratification

This leads us to the second challenge for SDIs. In the beginning, SDIs followed the model of a library. It's not a bad one to be sure, but the internet has made us come to expect immediate gratification. The web of 1993 is not the web of 2013, and that calls for a re-examination of this model.

In our map-library model, a dutiful librarian fills out a metadata card and that will lead the GIS user to the best map, data or layer. Alas, GIS data is dynamic, digital and ever-changing. Your GIS user puts two layers together to create a new map, and then another and another, but did he or she update the metadata along the way to contribute back to the library? Probably not. Our map card catalogue is woefully out of date almost from the moment it is created, much to the frustration of our hardworking GIS librarian. The problem is so pervasive that the GIS user may stop coming to your library and your librarian may decide he or she has better things to do than fill out metadata cards that no one is using.

So embrace the chaos. If every webpage had to have meticulous and detailed metadata to be searchable and discoverable, what percentage of web pages would be shared? No doubt, it would be a tiny fraction of the excellent, available content. If we look instead to our parallel universe on the web, we will see search engine-optimised (SEO) websites by the boatload. The better the SEO, the more likely your page is to be found. The more it is found and used, the higher its value in search results. The more often a user contributes excellent, accurate content to the web, the more often it will be considered a trusted source and will also bubble to the top of search results. Quality can be assessed in ways other than the presence of metadata.

Our new SDI model

This is the model we would propose for our new SDI: leverage metadata where it is available but let the data describe itself. Metadata is one input, but the usage of the community and the contributions of key players may well be able tell us what is the best data. It's simple – good data gets used, bad data doesn't. Furthermore, we cannot assume that the presence of metadata equates to accuracy. These concepts are as disconnected as the metadata can be from the data itself.

We can provide a very concrete example of how this could work through our own efforts on this front. VoyagerGIS, our small software company, was born as a partnership between a development and programme manager from Esri and a lead developer in the open-source search community that built Lucene/Solr. Together they taught Lucene to speak spatial like a native. Our product, Voyager, works by walking

through the myriad array of GIS files on your C: drive, your server, your favourite web services and that secret data stash you keep on an external drive. It opens each one and extracts every bit of information it can about it, some of it from metadata, much of it not.

Creating an index of that data, Voyager can answer questions like how many times is that data source used and in what maps documents? Where are the broken links? Are you hosting duplicate data?

It allows users to search by keyword for, say, 'roads' and to use a map to only look in Italy for those roads. Most importantly, if the data changes, Voyager knows that it has, even if the metadata hasn't been updated.

Upon implementing a Voyager-based, public-facing, data-download portal, one SDI manager said he was finally "confident that users are getting the most up-to-date data. Users are not hitting some separate file that is outside our spatial data warehouse."

Our point is not to dismiss the value of metadata. Rather we give this example because it seems that finding the best data should be our ultimate goal, rather than finding or creating metadata. People are good at discerning the quality of the data they require. Does a student looking for data on the desert turtle for a school project require the same level of metadata as the researcher at the Nature Conservancy would? Probably not. And yet, that Nature Conservancy researcher who made an internal report may require still less detail than a researcher from the Congressional Research Service working for a senator heading to the floor. However, for all of them, not finding even a basic dataset leave them all at a dead end.

The final challenge

On the flip side, the final challenge facing SDIs and their portal front-ends today is information overload. A quick Google search for 'state GIS portals' yields some 3.4 million results. A keyword search on Geoplatform. gov, the US government initiative that 'aims to be the "trusted source" of information so interested citizens can find the good, relevant data when they need it', yields four original results and some 57,000 harvested results for the term 'test'. That is hardly

The Google Crisis Response team assembled a publicly available map and resource page to help track affected areas and provide updated emergency information for the millions affected by flooding

reflective of curated, top-quality data.

There is a high cost to not finding the best data. The data in your SDI may well be the best data available. But how are you sharing it? The possible, available, excellent GIS data that could be contributed to a portal is vast and deep. But if it cannot be found, it is as silent as that tree falling in the forest.

Data that is hard to find will be overlooked. If it is overlooked then some lesser choice may be the only alternative or time will be wasted recreating it. The result may well be poor decisions, wasted time and missed opportunities.

We want to avoid this, I know. But to do so, we have to make the data accessible again, to a much larger audience through open search initiatives not portals.

Friends, hopefully you have taken a deep breath of fresh air, and not out of sheer exasperation with our proposals. Embrace the chaos and use your vast resources and talent to contribute to the evolution of non-traditional SDIs in a much broader community. It is an impossible challenge to attempt to control it, for in this broader universe, your resources are always going to be too small. As you go boldly forward, be driven by this new vision: six billion people finding and using your data to change the world.

THE POSSIBLE, AVAILABLE, EXCELLENT GIS DATA THAT COULD BE CONTRIBUTED TO A PORTAL IS VAST AND DEEP. BUT IF IT CANNOT BE FOUND, IT IS AS SILENT AS THAT TREE FALLING IN THE FOREST

Kris Goodfellow is vice president of product management and Brian Goldin is a co-founder and CEO at VoyagerGIS (www.voyagergis.com)
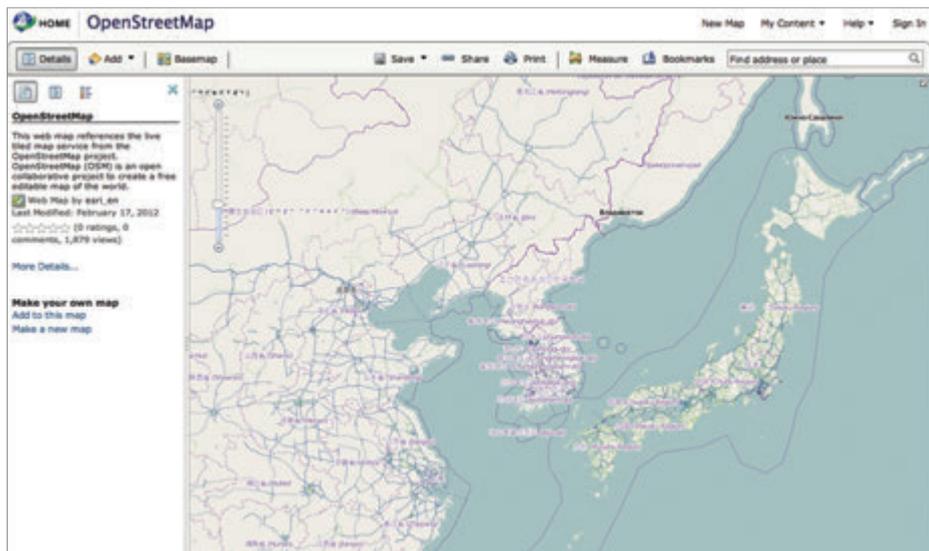
Users of ArcGIS Online can see a version of Open Street Map and add their own or other publicly available layers to the map
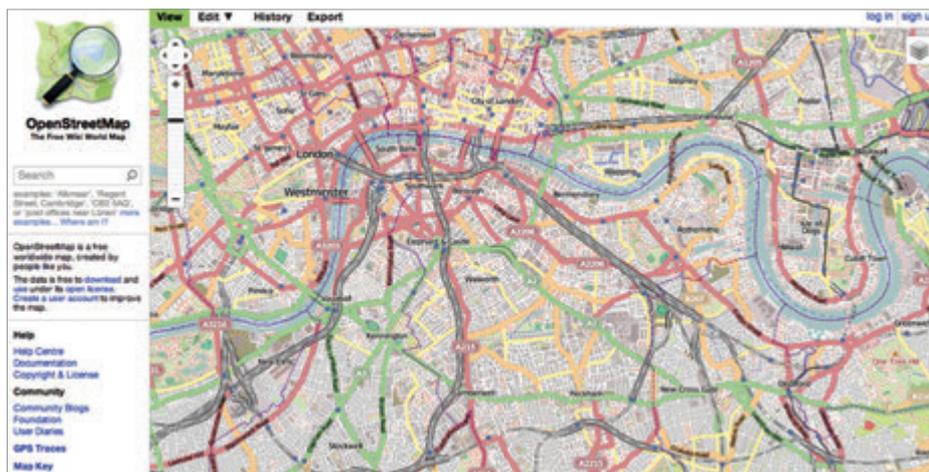
Open Street Map is an open source project started in 2004 with the vision of becoming a freely available map of the world