Tools Rush In

Jonathan W. Lowe

Jonathan Lowe draws on his six years of experience writing Net Results to examine current trends in, and predict the future of, geospatial tools — data, hardware, and "plumbing."

Milestone birthdays or anniversaries, great losses, and extended vacations are some of the rare occasions prompting today's busy people to stop and reflect on their past journeys, where they've alighted, and what lies ahead. This month's edition of *Geospatial Solutions* marks our 15th anniversary of publication, prompting the magazine's regular contributors to sink down into our armchairs, gaze wistfully across the veranda, and slowly piece together the meaning of it all.

Even after writing about emerging geospatial technologies for the past six years, I still can't claim to be much of an industry tarot card reader. Nonetheless, I have reviewed my fair share of emerging technologies, so my contribution to our 15th anniversary's collective musing will be to review the past and present in hopes of glimpsing the future of geospatial tools.

What are our tools, how and why has our technology changed over time, and what might surface next?

Net Results columnist **Jonathan W. Lowe** covers the role of emerging technologies in the exchange of spatial information. Lowe is the owner of Local Knowledge Consulting (Berkeley, California), where he designs and implements spatial Web sites. He can be contacted at info@giswebsite.com.

I'll look for innovative advances in data, hardware, and "plumbing," or how we exchange the results of our work interoperably.

Geospatial tools don't just pop into existence; they're a solution to some problem, or sometimes a solution in search of a problem. Considering some of the unsolved problems still troubling our industry will suggest what tools might one day appear to solve them.

Data: Tool Fuel

Tools aren't much use without something to manipulate, and as spatial data change, so do our tools. So, how have our data been changing? Commercial and government datasets keep gaining detail and granularity, with an inevitable increase in storage size, maintenance challenges, and data-management options. Ten years ago, the typical geospatial project revolved around data conversion from paper to digital formats. Municipalities hired teams of GIS specialists to build spatial datasets of their streets, utilities, tax parcels, police beats, emergency response districts, and many other features, all from scratch. Today, that basemap work has been completed for most major metropolitan areas. Now cities are appending more detailed, specialized data to their bases, such as building footprints, oblique aerial imagery, traffic accidents, and daily criminal activity. Some are experimenting with real-time data capture, such as tracking emergency response vehicles and police cars. A smaller group of practitioners are also exploring strategies for archiving outdated data or managing it in an active spatiotemporal structure supporting historic views and change analysis.

What are our tools, how and why has our technology changed over time, and what might surface next?

Appending specialty and real-time data or storing spatiotemporal records makes datasets grow, and these ever-larger datasets require sophisticated access methods. If an organization can afford extensive and detailed spatial data, then it is also likely to support a pool of data users all needing simultaneous access to view or edit that dataset. Spatial databases are today's most common tools for serving multiple simultaneous users while remaining highly performant. Spatial datasets perform well by indexing data in advance of their users' queries. Performance is routinely the top requirement cited in user surveys, far above ease of use or cost. Consequently, database vendors compete to have the fastest engines, offering not just one index strategy, but several specialty indexes for spatial data, or the ability to index multiple dimensions and partition data by space or time. Different data types get attention from one year to the next. For instance, one of today's data management battlegrounds is the performance-to-volume ratio when querying seamless imagery collections. Peeking over the horizon is vendor competition to dominate the spatiotemporal data-management market.

Storing topology in the database is emerging as an alternative to some traditional GIS vendor strategies of computing it on the fly. With their initial data-conversion work over, data owners are intent on guarding the integrity of their holdings, such as by formalizing rules that prevent new corrupt data from entering their databases in the first place. Laser-Scan specializes in this capability, not only raising alarms if incoming data violate a rule about proper geometry or relationships to other datasets, but also fixing those problems in existing datasets. For instance, given a rule that building footprints do not overlap curb lines, Laser-Scan's software can traverse a spatial database to find any offending building and curb objects and automatically correct them by adjusting one or the other until they satisfy the rule. This rules-based data-management approach hinges on a detailed data model, leading some vendors to explore rules in the form of ontologies (see Net Results, June, 2005).

Well-established data vendors, such as NAVTEQ, Tele Atlas, and Ordnance Survey (OS), have been experimenting with their data models for years. Though their spatial databases may contain a single master schema, they can filter and transform that core data to pull many different products with different end purposes from one original source. OS, for instance, currently provides attributes only for physical features, but will soon also provide functional attributes. In other words, a building footprint with the form "church" may actually function today as a youth hostel due to an adaptive reuse city-planning scheme. Storing not only the form but also the function of certain objects will allow OS to provide two different products for the same collection of geometries. Expect to see a wider variety of data products from commercial data vendors in the future as they tailor their offerings to fit specific customer needs without having to maintain multiple data sources.

There is a tipping point where a large dataset becomes too extensive and detailed for its vendors to keep current and still turn a profit. Arguably, Tele Atlas North America has already reached that tipping point, but it remains profitable by extending its enterprise to include a wide variety of distributed local providers who contribute but will never receive a Tele Atlas paycheck. For instance, when a local construction company maps a section of roadway to prepare for a dig, they may share their resulting highly accurate and current data with Tele Atlas, possibly in exchange for nearby data. Other sources don't even know they're helping; Tele Atlas scrapes construction announcements posted to the Web, for instance, to identify new street construction or changes to traffic rules.

These are effective approaches to the never-ending task of keeping spatial data up-to-date, but will data vendors go even further to outsource the general public? The United Kingdom's Rural Payments Agency (RPA, www.rpa.gov.uk) is responsible for maintaining a multimillion-polygon dataset of what they call “land parcels” to help track subsidies to farmers. On a regular basis, RPA sends every registered farmer a paper map of his or her land parcels and requests corrections. The farmers sketch where they’ve planted a new hedge or are growing a different crop and return the revised map to RPA. The medium of data exchange (paper) is an old one, but the concept is quite innovative, or perhaps born of necessity. Because there are far too many land-parcel polygons for RPA to maintain with its own small field staff, the agency instead solicits updates from the people who know the data best and care most — the farmers themselves.

Back in the digital world, public online mapping sites such as MapQuest, Yahoo! Maps, and Google Maps allow their users to place their own ad hoc points but do not then store those points as enhancements to the core dataset — and why should they? Who would care to see them? However, these sites (and some emerging commercial offerings) demonstrate that it’s technically possible for users to input geometry to a central map source via a Web browser interface. Might local governments someday use similar technology to enable citizens to update the spatial data for their own neighborhoods? Corporations cut costs by replacing their service personnel with computer interfaces and distributing responsibility to individual customers. Thanks to the Internet, we the people have become our own bank tellers, stockbrokers, and travel agents; might we also eventually serve as our cities’ collective geospatial data providers?

Though not exclusively geospatial, a knowledge-base development company called Cycorp has been building a structured collection of facts that it says will form the basis of artificial intelligence on the Web. The company already provides a subset of this collection (in Web Ontology Language format) free for public download by the name of OpenCyc (www.opencyc.org). This year, Cycorp will open its doors to the general public in the form of a Web site where anyone can submit questions to the Cyc engine. OpenCyc’s engine uses the knowledge base to parse the public’s questions and, when encountering a new word or idea, “learns” from the question. Cycorp will be able to answer some questions immediately, but others may take time as the system learns. After more than 10 years of development, Cycorp’s tools have enough background knowledge to know what they don't know. To expand the knowledge base, Cycorp will tap the collective input of the world’s Internet users. Though reminiscent of a science fiction novel’s plot, leveraging the Internet as a societal brain may not be as far-fetched as it sounds. Before dismissing the concept altogether, consider first taking a casual stroll through OpenCyc’s knowledge base, if only to see the results of more than a decade of meticulous work by a determined startup company.

Thanks to the Internet, we the people have become our own bank tellers, stockbrokers, and travel agents; might we also eventually serve as our cities’ collective geospatial data providers?

Hardware: Smaller, Cheaper, Hybridized

The public is not just a herd of microinformation cows waiting to be milked by private industry. We also have more access than ever to the tools of the geospatial professional’s trade, and we can pull in spatial data for our own use. For instance, for programming a real-time spatiotemporal application, Motorola GPS-enabled cell phones with the Java programming language built in are publicly available. Sensor data stream in freely from some government weather stations or can be purchased in preconverted spatial format from commercial meteorology data vendors such as Meteorlogix. Tiny sensors called motes lack the power supply to carry onboard GPS chips, but they are quite cheap and enable the masses to deploy their own sensor networks (see Net Results, May, 2004).

These fancy and publicly available spatial hardware toys are a lot of fun, but my selection for the rising star of the hardware family is one of the most humble. Radio-frequency identification (RFID) tags cost almost nothing, and therefore can be deployed in great numbers. RFID tags do not have their own power source; instead, they are antennae that bounce back a unique signal when bombarded with energy from another source. That other source, the RFID “reader,” then knows which RFID tag is in its vicinity. Bookstores use crude RFID tags to protect themselves from shoplifters — a tag is in each book, and the paired posts at the door of the store send the radio signals. Walk out without buying your book, and the tag will set off an alarm. Spatially, bookstores are testing for only one condition: the book is inside or outside the store. Similar approaches are being tested in the United States and United Kingdom by attaching RFID tags to cows’ ears to track their movement from one field to another. The systems don’t tell exactly where in a particular field a tagged cow might be, but they do tell when each cow was in a given field, and which other cows were there at the same time. For purposes of disease containment, field-level tracking meets the analysis requirements and, more importantly, is a cost-effective solution thanks to the low cost of RFID hardware.

Because RFID tags are passive and don't know where they are except in relationship to an RFID reader, they've gotten less geospatial industry attention than (larger, power-hungry, expensive) GPS devices. However, humble-but-cheap RFID and sophisticated-but-fussy GPS may become increasingly regular bedfellows. To date, GPS manufacturers have not found workable solutions to the problem of indoor tracking, where RFID can work quite well. For instance, finding packages stored in large warehouses demands a spatial solution not easily implemented with a GPS. But by embedding RFID tags on the points of a grid in the concrete of a warehouse floor, and by attaching an RFID reader to the bottom of a forklift, it's possible to track the forklift's movements inside the warehouse. Combined with a GPS, that same forklift can be tracked when it's outside as well, providing a complete facilities-tracking solution with a hybrid technology approach.

Plumbing: Web Services and Ontologies

Examples demonstrating smooth interoperability of multiple technologies are not yet commonplace, although much of our industry's discussions revolve around this topic. Why is interoperability so difficult to achieve? The problem stems in part from human nature. We humans work by specializing and collaborating, usually in that order. We retreat to the efficiency of our little clans to solve a specific problem, then later share our work with the larger tribe, hoping also to benefit from the work of other clans. It's then that we discover "disconnects." Though specialized development does yield faster results than design by committee, it does so by delaying coordination of mental models among the larger group of potential users. For instance, law enforcement agencies of neighboring states may want to capture criminals fleeing across shared state borders. Why not just share criminal database records to solve the problem? After all, both agencies have the same conceptual directive — to catch criminals. When they compare their information systems, however, the underlying database structures or even the overall system architecture may both be quite different, and neither agency's systems may be capable of changing to match the other's. One state's conceptual approach may be to organize crimes by time and severity, whereas another may organize by the offenders themselves. One may use a file-based approach, while another uses a database.

As an industry, we have been struggling with this shift from clan to tribe, from specialization to collaboration, for at least the past five years, if not longer. One approach to the problem has been to trumpet the need for standards. Unfortunately, many legacy systems are too expensive to overhaul to a common standard, and there isn't necessarily a universal standard in the first place. The best solution yet to emerge has been Web services, which leaves the underlying implementation, database schema, architecture, and so on, to each specialized clan's preference, but provides a recipe for extracting the clan's information for the larger tribe's purposes using a standard protocol.

As a result, geospatial tools are appearing that are neither front ends nor back ends, neither user interfaces nor databases. Rather, they are the Web-services plumbing between the data and its presentation. The free open-source GeoServer (http://geoserver.sourceforge.net) product is one example, and can provide data between almost any open database connectivity–compliant database and Open Geospatial Consortium (OGC, www.opengeospatial.org) Web services–compliant mapping tool. On one hand, the Web-services approach frees database and presentation tool vendors from the odious task of maintaining separate codestreams for each database or client tool with which they exchange data. On the other hand, Web services commoditize spatial data viewing tools and databases to some degree, sending the clans off in search of some new, more complicated problem in which to specialize in order to keep their market edge.

Worth noting in the geospatial Web services discussion is the fact that OGC's guidelines for spatial Web services don't exactly match the World Wide Web Consortium's guidelines for Web services. OGC has adopted the language and justification common to the larger information technology (IT)-wide Web services initiative, but has only partially followed the movement. Researchers of both Web services and the Semantic Web are publishing approaches that bring Web mapping services and Web feature services into closer compliance with the IT industry standard. Whether or when they will become mainstream within the geospatial industry is anybody's guess.

Tools from the Woodwork

Because "Net Results" typically investigates emerging technology, the work I most often review is fairly raw in comparison to offerings from long-established professional vendors (though sometimes their new efforts qualify as emerging technology as well). As a result of these ongoing explorations of new geospatial technologies, I've come to believe that the tools of our trade are the physical manifestations of our industry's goals and dreams: The dreams come first, the tools second. Where spatial dreams still struggle most rigorously to become reality, one often finds the most innovative emerging technologies. There continues to be significant geospatial development activity in the open-source community, as evidenced by the growing number of available products and publications such as O'Reilly's soon-to-be-released "Mapping Hacks" by Schuyler Erle, Rich Gibson, and Jo Walsh, or "Web Mapping Illustrated" by Tyler Mitchell, to name just two. These efforts and publications carry with them a philosophy that both data and tools would better benefit society by being directly accessible to all. Whether or not you agree — or find this approach feasible for your livelihood — it is undeniably producing some very useful tools and, maybe more importantly, dreams that will influence and change our industry's future. And even if they have no influence whatsoever, they're a hoot to play with and read about!

Many readers of *Geospatial Solutions* have well-defined jobs that demand proven, reliable industry solutions. Mirroring our lives, our work seldom affords us the spare time to reflect on where geospatial technology's dreamers may lead us in the misty future. I hope this anniversary survey of emerging data, tools, and dreams assures you that our industry remains a living, changing, and exciting place to work, and that even the most buttoned-down implementation might still benefit from an occasional new tool. ◍