Internet Mapping

At the broadest level, Internet mapping refers to the capture, storage, production, and representation of spatial data via the Internet. The spatial component of such data introduces unique hurdles to standard data management and visualization vis-à-vis Internet technologies, and with increased pervasiveness of location-based services, Internet mapping is becoming a core way to assess patterns and observe relations across datasets. Such broad definition is necessary on the one hand because of the range of technologies involved in delivering an online map, and on the other hand because of the range of spatial practices one enacts in making maps online. An inclusive understanding of Internet mapping combines attention to its constituent technologies with these social and practical considerations. This entry synthesizes across these considerations, first providing historical context for Internet mapping, then discussing the current state of the field along with speculation on important future developments. Lastly, the entry covers pressing research challenges and concerns that emerge when exploring the field.

Where have we been?

As far back as the 1980s, academic researchers, technology developers, the military, and public administrators have recognized the unique challenges that geography introduces to Internet data transmission and representation, and began to produce the technical infrastructure and support new practices for doing so. Access and use of data across disparate locations was a challenge, but also the unique ways spatial data needs to be managed, analyzed, and visualized. Geographic Information Systems (GIS) were developed to address the latter challenge, and early Internet technologies adjusted to synthesize across the challenges. Distributed GIS, meaning multiple linked systems across disparate locations, served these purposes, and streamlined the process of creating and manipulating spatial data from GPS-enabled mobile devices. This approach to GIS comprised a strict definition of *Internet* mapping, as GIS licenses and data were often accessed through remote FTP (file transfer protocol) servers and not simply the world wide web; however, this era was still characterized by disagreement about analytical and conceptual overlaps amongst “Internet GIS”, “mobile GIS”, and “web mapping”.

Advances in technological infrastructure paralleled both the growth of the geographic industry symbolized by the private companies Esri and Intergraph, and increased access and interest in the world wide web. Websites such as Mapquest and the USA’s National Weather Service launched important early examples of Internet mapping that received extensive public interest. These quite often relied on Esri’s proprietary Internet mapping software ArcIMS, which fine-tuned the standard presentation, business logic, and data management tiers for specifically spatial data. Still, early Internet mapping platforms suffered from the relatively static nature of web pages at the time. For every user command – say, to pan or to zoom a map – the entire web page would need to reload to display the image representation of spatial data requested. This introduced two sources of inefficiency: reloading the page, and requesting new data only upon the user’s request.

By the early and mid-2000s, several possible development trajectories for Internet mapping narrowed in response to important socio-technical shifts. First, in 2000 the USA government discontinued the Selective Availability program on its Global Positioning System (GPS), which had limited the accuracy of civilian GPS units through signal scrambling. Discontinuing Selective Availability encouraged widespread adoption of GPS technology and enabled laypeople to produce spatial data, later to be incorporated into smartphone technology. Second, the asynchronous data transfer capabilities of most browsers greatly expanded, meaning new data could be downloaded and displayed without needing to reload the web page. Out of the range of technologies enabling this shift, AJAX (Asynchronous Javascript and XML) became the most popular. AJAX enabled web applications like Google Maps to download data adjacent to the user's current map view, speeding the spatial navigation process significantly. Termed "slippy maps", this interactivity allowed the map frame in a browser to change with the user's mouse – dragging the mouse would dynamically pan then map in the map frame, rather than needing to reload the entire page. The map thus "slips" with the mouse. Third, AJAX proliferated simultaneous to the displacement of Web Mapping Services (WMS) and Web Feature Services (WFS) to relatively small stitched-together images of data called "map tiles". Because the former required large graphical downloads exceeding the user's viewable range, the latter made map navigation more efficient in data downloads. This efficiency would later become crucial for early smart phone usage. The fourth socio-technical shift was the discovery by large data-collecting institutions such as Facebook, Twitter, and Flickr that they could profitably share limited access to their datasets and services through Application Programming Interfaces (APIs). By opening some usage of their platforms and data to web developers, companies holding large spatial datasets attracted more extensive use of their platforms and enabled creative re-appropriations – ultimately enabling new digital forms of profit generation.

Most histories of Internet mapping tend to emphasize these technical innovations to the relative neglect of the social and political shifts that underwrote its development. As suggested above, the release of APIs signaled more than technical shifts – instead, it reflected a political economy in which private companies sought to identify new avenues of profit. The growth of the spatial web 2.0 (called the "geoweb") coincided with the state delegating more of its data production responsibilities to "the crowd" and private companies, while concomitantly rolling out new regulations and frameworks for generating private-sector profits from these new activities. Since the early 2000s social media have increasingly integrated a spatial component, producing new (big) datasets for Internet mapping, yet these data reflect new digital divides as adoption, access, usage, and skills diverge widely across the globe. In other words, Internet mapping is a social and political phenomenon in addition to a shift in the technology milieu.

Where are we now?

Both conceptual approaches to understanding the Internet – the technical and the social – inform recent interest in mapping Big Data. As more data are produced, circulated, and represented on the Internet, researchers have balanced attention to data velocity, veracity, volume, and variety ("the four Vs") with attention to Big Data's effects on knowledge production. Doing so helps account for the disparity in *who* maps what *kinds* of data, drawing in sharp relief the inequalities of Internet mapping. This research program borrows ideas and principles from critical, participatory, and feminist GIS to argue that such disparities create an unequal playing field for representing data through the Internet, ultimately privileging those in high-income countries and cities. On the technological side, web developers have responded to Big Data by writing new Javascript libraries such as D3 to help spatialize, analyze, and visualize large datasets on the Internet. Other recent Javascript libraries have been utilized for Big Data visualization, but also for standard mapping techniques; examples include Leaflet and OpenLayers.

Notably, while data transfer speeds have increased, the improved computational capacity of personal computers has meant that many of these technical innovations have shifted Internet mapping to more client-side processing. That is, much – if not most – Internet mapping now relies more heavily on Javascript, HTML5, and CSS, rather than the historically more common server-side applications using PHP or ASP.NET. Client-side Internet mapping places a higher computational burden on browsers and users' devices, but may improve interactivity and navigation efficiency. Innovations in server-side processing, though, include server-side Javascript, and Python web development frameworks, both of which having had strong contributions to delivering closed data to Internet mapping platforms. The exception, of course, is for institutions and individuals who serve their own spatial data to the Internet; doing this usually involves, beyond the standard web server, a GIS server to handle spatial data requests, as well as a GIS database server to access the data themselves. These may be filtered through a sort of configuration file, like Esri's "map document", for rendering specifications.

Currently, Internet mapping serves important functions for researchers, private enterprise, public administration, and the general public. Academic researchers are presented with new sources of data, as more people have become enrolled as sensors in the geoweb. APIs provide another source of spatial data such as social media, demographic and social data, environmental data, and primary data sharing. Current technological trends mean that cartographers are challenged by new conventions and best cartographic practices for effectively communicating complex data and visualizations on the Internet; these challenges have implications both for practitioners as well as for teaching students in higher education. Private enterprise has benefitted from the accelerated download transfer speeds around spatial data, as well as the strengthening trend toward openness and spatial data sharing. For example, the Seattle-based private company Socrata develops open data platforms, including some data with a spatial component, for governments at various scales, moving most data to private and distributed servers and web services. This, in turn, unifies for an entire government entity what could otherwise be scattered datasets, streamlining data access and data discovery for public administrators and citizens. Data not opened to the public are still often shared among government employees through Internet protocols. The general public now both produces and consumes spatial data through Internet mapping platforms, as the accessibility, interoperability, and integration of disparate datasets has increased.

Geoweb platforms such as Google Earth, Strata, and Yelp, for instance, serve as platforms for producing spatial data, but also for producing Internet maps. Their integration into everyday lives means that Big Data provides a new, partial, and variegated view on social processes. As well, it means the public can participates more strongly in Internet mapping practices. That is to say, with a small amount of technical expertise, members of the public can now become Internet mappers, and they now number much higher than was the case in the early 2000s. With greater technical skill, the public now has free access to spatial analysis operations such as buffers and network analysis through open-source Python and Javascript libraries.

Where are we headed?

The landscape of Internet mapping changes quickly, and any statement about current trends will likely be out-of-date before reaching the printing press. However, three socio-technical advances seem to hold the most promise for longevity, widespread adoption, and future development.

In contrast with raster and image-based data download and visualization, *vector tiles* stitch multiple small selections of vector-based data for the user. This works similarly to raster tiles, in that the data delivered to the user can be scaled to many resolutions, and is an efficient way of transferring spatial data to Internet maps. However, since the data remain in vector format, they are generally smaller and thus faster to download and display. Tiling the dataset further removes the need to download and display entire vector datasets at full resolution, greatly improving efficiency over standard vector data models. Typically, vector tiles are extracted from the server using information embedded within a URL. This URL extraction scheme was common with raster tiles, as well. The format follows a

zoom/x/y

approach. Here, *zoom* is a number usually between 1 and 20, where 1 displays the global scale and 20 can be as small as a few square feet. Thus, to display a neighborhood a number between 16 and 14 might suffice, and to show an entire urban area, 10-13. *X* and *y* represent the latitude and longitude values, encoded as decimal degrees (rather than degrees-minutes-seconds). This value pair, together, are the location of the center of the map. Together with the zoom value, the vector tile server determines the extent of vector data to deliver to the user.

As of 2017, MapBox Studio is a highly used vector tiling platforms for Internet mapping; OpenStreetMap also serves vector tiles of its data using its vector data server called the Mapnik engine. After having extracted these data, several available services may stylize and render the vector tiles before displaying for the user. MapBox offers this rendering service as part of its software, using the proprietary Mapbox GL Style Specification (the descendant of another standard, CartoCSS), and Mapzen and Stamen Designs rendering specifications using more open formats.

Second, web developers, social scientists, and data scientists are finding new ways to encode geographic data for more efficient transfer, storage, and display purposes, and the *TopoJSON* format appears to be increasing in value and use. TopoJSON is a data structure built on the JavaScript Object Notation (JSON), using key-value pairs for high human readability. JSON emerged to substitute for XML in encoding and transmitting data, and is easily parsed in JavaScript. The GeoJSON extension of JSON aims to create standards for encoding explicitly *spatial* data in the JSON format, but GeoJSON suffers from large file sizes. Its readability and extensibility pale in comparison with its long download and display times. The size of these objects varies, of course, depending on the dataset granularity/resolution and attributes included, but for comparison, an official 2015 California Census tracts shapefile at the 1:500,000 scale as a shapefile is 6.9 megabytes, and encoded as GeoJSON is about 8.9 megabytes.

The TopoJSON specification was designed to address this issue. Instead of individual vertices and lines for each object boundary, TopoJSON stores arcs shared by features. Thus, coincident boundaries are not duplicated: the boundary between Canada and the United States would be stored once as a shared arc rather than as two separate object boundaries. Additionally, with this approach users maintain topological consistency and smaller file sizes. The same file as above is 75% smaller as TopoJSON, at about 2.25 megabytes. Because TopoJSON is a relatively new standard, converting between shapefiles, GeoJSON, and TopoJSON remains an encumbrance to adoption and usage. As with other Internet mapping technologies, users must link to external libraries to display the files properly.

Third, the growing trend toward distributing spatial datasets for unrestricted public use comprises the *open data* movement. Open data platforms are increasingly common across many scales of government, including cities, regions, and countries. The motivations and manifestations differ widely across contexts, but are largely driven by efforts to promote government transparency, efficiency of data sharing, public entrepreneurialism, and broaden knowledge discovery. The proliferation of open data platforms in the twenty-first century has opened a flood of spatial datasets to Internet mappers, enabling broader options for spatial analysis and cartographic products. Importantly, in many ways this movement constitutes a subtle shift from earlier thinking and application in spatial data infrastructures, in that both promote richer collections of shared spatial datasets across a geographic area. These debates largely oriented around how to increase public participation and public access, and diverse solutions were proposed: those included the notion that publicly-funded data should be made free to the public, or that a cost-recovery program should be in place to recuperate the cost of producing the data. For the open data movement, however, the discourses seem to focus on the figure of civic data activists – "hacktivists" – and the kinds of transparency promoted by public-facing datasets.

The open data movement presents conceptual and practical challenges to researchers and policymakers, which are currently the focus of much debate. First, as conceptualizations of the digital divide increasingly integrate discrepancies in digital skills and software licensing capacity, it becomes important to consider who the open data movement actually helps. Data availability has certainly always limited the kinds of analysis researchers and the civically-engaged could conduct, but to effectively leverage many datasets requires a skillset not held by the general public. This inequality pertains equally to the intricacies of spatial analysis, as to sound web-cartographic principles, as to technical GIS and web mapping skills. Moreover, software licenses and hardware each present formidable hindrance to widespread mapping and analysis of open data resources. Second, benefits to transparency should be critically evaluated, and to date there is a marked lack of empirical case studies to lend evidence to claims of transparency. The institutions creating and maintaining open data platforms retain the prerogative to release datasets or not, and the politics around these decisions remain opaque. Third, a small number of private, for-profit businesses develop and maintain the majority of large-scale open data platforms. This raises questions around how the profit motivator shapes the kinds of data made publicly-available, and the ways these third parties should be held accountable. Fourth, policymakers should carefully consider the kinds of datasets that *could* improve transparency, and evaluate the public's access and ability to analyze and map them. This is a conceptual question with strong material impacts. For example, one consideration might be whether election spending and campaign funding datasets should be made available, or whether a dataset should be made available for map visualization rather than simply a shapefile download. Further issues around privacy protection are raised for policymakers in the open data movement.

Why and how should we be cautious?

Just as critical GIS debates in the 1990s and 200s foregrounded the processes and institutions that shape our knowledge, so too should researchers and practitioners be wary of the way Internet mapping increasingly frames how we perceive the world to be. In critical GIS debates the private GIS company Esri constituted a major focus of antagonism, as researchers interrogated the company's – and GIS's – complicity in militarism, corporate incursions into the public sector, and a naïve realist approach toward science; now, for Internet mapping, these questions may be reframed toward private companies like MapBox, Socrata (the open data company), and others. There are pressing questions around the ways in which these private companies, and the software they offer to the public or to clients, delimit the forms of interrogation and knowledge production that we may execute in Internet mapping. Such an enquiry might be couched more broadly within a concern for the ways of knowing that are closed off or marginalized from Internet mapping. Comparing data coverage between OpenStreetMap and Google Maps for places in the Global South – say, for example, the Gaza Strip or Dhaka, Bangladesh – illustrates one potential form in which this concern may manifest. Early research suggests that many technological inequalities observed earlier persist, yet we require new conceptual material to account for the distinct technological environment.

Another way such longstanding debates re-emerge in Internet mapping is through the persistence of the Mercator projection. The Web-optimized version of the Mercator continues to be the de facto standard projection used in web mapping. The Mercator projection preserves local angles, and is thus well-suited for the navigation and network routing purposes to which it is often put in web applications. However, cartographers have long noted that the colonial-era projection, when viewed at small scales, exaggerates area closer to the poles and conversely minimizes the sizes of the Global South countries near the equator. This effectively makes European and American countries more visually dominant, and consequently represents them as more powerful, diverse, and important. The Web Mercator reinforces those power relations in Internet mapping. Technical capacities are expanding rather quickly, though, and websites such as Jason Davies's "Map Projection Transitions" demonstrate innovative ways diverse projections might be integrated into Internet mapping platforms to circumvent problematic representations of the earth. Historically, some popular platforms have also experimented with other projections, as well. The field currently needs a broader repertoire of creative solutions to the complex social and mathematical problems associated with projections.

Big Data proponents continue to approach Internet data collection as if these data represent society writ large, neglecting the tenacious digital divide and the inequalities of social representation. This fact has ramifications for how we think about Internet mapping. Among the many sources of Big Data there remain glaring gaps: those who forego social media, those not living in places with environmental/social sensors, "blank spots" on the digital Internet map, those who actively hide their digital activities, and more. As before, observing the global coverage of Google StreetView illustrates the spatial variegations of data and mapping coverage. A burgeoning research program – from which some citations above come – orients around the question of how this digital divide presents itself in new ways in contemporary society. Internet mappers would do well to keep these inequalities in mind and to understand the partial nature of Big Data.

These three paragraphs have on some level invoked an ethical framework guiding Internet mapping research and use, but Internet mapping also presents new challenges to the ethical principles enshrined explicitly in legal codes. Issues regarding privacy are among these ethical considerations, as Internet mapping introduces new ways of triangulating multiple data sources, thus potentially breaching standard privacy protection expectations. New practices such as geotagging, digital presences and profile creation lower the barriers to privacy breaches. These new challenges and considerations include control over the use and accessibility of personal and personally-identifiable information. *Consent* factors into these conversations in that while some may agree to publicly publish their information, the non-determinant nature of Internet mapping development means that future uses for the data may arise that the user could not foresee. The user may not have consciously consented to all possible future uses of the data, and perhaps consent should therefore be re-secured for new purposes and applications. A salient example of this is the Girls Around Me app, which for a brief period (2011-2012), displayed female-identified social media users nearby the user's location. The app gathered public information from the location-based check-in service FourSquare and combined it with public Facebook profiles; however, critics raised questions about the degree to which those social media users were aware of this use of their data. Similar concerns were raised about the Please Rob Me application, which streamed the same data sources to display when someone is not at their place of residence. Further research is needed to clarify the ways these principles, concepts, and rules apply to the Internet mapping sphere.

More insidious implications emerge in the context of vulnerable populations, such as in humanitarian work, emergency management, or around ethnic or sexual minorities. The American Association for the Advancement of Science is currently conducting research into the ways social media and crowdsourced spatial data can be used safely and effectively for humanitarian aid delivery and humanitarian rights violations. This work is situated within a small but growing body of work looking at how to protect users of spatial technologies, and the broad implications span from Facebook's real-name policy to conflict-zone digital humanitarianism, from Google Maps's use of unpaid digital labor to surveillant smart cities.

While technological innovation has enabled Internet mapping, it is an inherently social phenomenon. Pressing research needs orient around the ways people use Internet mapping, and the ways it shapes what we know about the world and our place in it. Novel approaches to spatial data are allowing practices and explorations previously beyond the purview of single-user desktop GIS.

Further Readings

Battersby, S., Finn, M., Usery, L., & Yamamoto, K. (2014). Implications of Web Mercator and Its Use in Online Mapping. Cartographica, 49 (2), 85–101.

Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. Geoforum, 42 (1), 6–15.

Graham, M., Hogan, B., Straumann, R., & Medhat, A. (2014). Uneven Geographies of User-generated Information: Patterns of Increasing Informational Poverty. Annals of the Association of American Geographers, 104 (4), 746–764.

Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shifts. Big Data & Society, 1, 1–12.

Muehlenhaus, I. 2013. Web Cartography: Map Design for Interactive and Mobile Devices. Boca Raton, FL: CRC Press.

Peng, Z.-R., & Tsou, M.-H., (2003). Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks. Hoboken, NJ: John Wiley & Sons.

Tait, M. (2005). Implementing Geoportals: Applications of Distributed GIS. Computers, Environment and Urban Systems, 29, 33–47.

Warf, B. (2012). *Global Geographies of the Internet*. New York: Springer.

See Also

CyberGIS

Digital Earth

Digital Divides, Geography of

Geoweb

Internet GIS

Location-based Services

Volunteered Geographic Information

Web GIS

Ryan Burns

Department of Geography

University of Calgary