Geotag

Ate Poorthuis

Humanities, Arts, and Social Sciences

Singapore University of Technology and Design

Published as: Poorthuis, A. (2020). Geotag. In: A. Kobayashi (Ed.) *International Encyclopedia of Human Geography (Second Edition)*. pp. 137-140. <https://doi.org/10.1016/B978-0-08-102295-5.10551-7>

Abstract

The geotag is a geographic identifier that can be added to any digital object (e.g. a photo or a message). Enabled by a host of new technologies that emerged around the turn of the century, today much digital content has such a geotag attached. These large, geotagged datasets, ranging from simple websites and Wikipedia to social media platforms and transportation records, affect the world around us, as well as how we conduct geographic research about that world. Although a relatively humble concept, the geotag has enabled an entire wave of new types of geographic research that are often labelled as “digital geography.”

Keywords

big data, data science, digital geography, geocomputation, georeferencing, geotag, geoweb, internet geography, social media data, volunteered geographic information

Introduction

A geotag is a geographic identifier that can be added to a digital object such as a photo, message or website. This geographic identifier can be a relatively precise set of position coordinates but also a character string that refers to an address, zip code or other (vernacular) geographic place indication. The geotag provides geographic context to data and can also enable people to find information specific to a certain geographic location. Today, digital content is often geotagged in some way or another. This geotagged digital content has enabled two significant changes. It has changed how Geography, as a discipline, conducts research but it has also affected many socio-spatial processes in a more direct way. In other words, the geotag has changed both the way we can study the world and has also affected the world itself.

History

In the early 1990s, satellite imagery and the Tagged Image File Format (TIFF), a digital file format for storing images, became increasingly commonplace. In order to make such imagery more manageable, producers and users of mapping software and satellite imagery collectively proposed a standard, the GeoTIFF, to attach geographic information to satellite imagery that was often stored as raster TIFF images. This geographic information came in the form of a *tag*, an additional label attached to the image that contains a geographic location. Through reading these tags, software can immediately understand which specific part of the world a satellite or aerial image is referring to and thus place each pixel of the digital image in the ‘right’ location.

While the process of geotagging digital images was initially the remit of professional mapping software, this changed rapidly in the following years. Around the turn of the century, digital cameras and smartphones with built-in cameras became a mass-consumer product. Similar to the TIFF standard, consumer images, often stored in the JPEG file format, adopted the Exchangeable Image File Format (EXIF) which allows metadata to be attached to the image through a tag-based system. Following the older GeoTIFF standard, this can also include a location tag or label, for example as derived from the built-in GPS receivers in these consumer products.

The second half of the 2000s subsequently saw the proliferation of a new type of internet services, which together are often referred to as Web 2.0. These services allowed the internet to transform from a medium of *consuming* information to a medium of *producing* information and, specifically, made it easier for internet users to create their own web content, such as blogs, social media posts and photo galleries. For example, the social photo sharing platform Flickr (started in 2004) allows internet users to upload their own photos to public photo galleries that can be browsed and viewed by others. Leveraging the EXIF information attached to many of these photos, this resulted in one of the earliest consumer-produced repositories of geo-tagged information, allowing Flickr users to display their photos on a map and others to search for photos based on a specific location.

Other Web 2.0 services, such as blogs, Wikipedia, and social media websites such as Facebook and Twitter also allow users to geotag their content. For instance, people can “check in” to a location on social media, alerting others to where they are. They can also attach a specific location tag to their post, showing where the message was written. All of these pieces of digital information thus have a geographical location, or geotag, attached to them.

In addition to personal uses, there are also numerous commercial uses for the geotag. For instance, online mapping platforms, such as Google Maps, can overlay additional information on top of the map. In addition to conventional navigation, people can use these mapping platforms to find nearby businesses, landmarks, and other points of interest – each with their own geotag that pinpoints its exact location on the map, making it searchable to a broad public.

Although the geotag is a relatively simple piece of metadata, it has had a profound effect on the production and consumption of online and offline information. Consequently, it has affected how geographic research can be conducted, and even has changed the nature of some social and geographic processes.

Changing geography of information production

As the geotag expanded beyond the digitalization of satellite imagery analysis to the broader reaches of the internet, it changed how and who can produce geographic information. The geotag is the foundation of an array of associated technologies that enable laypeople to create geographic content. Examples range from tools that allow individuals to produce a map with pinned locations of their favorite restaurants to more complex citizen science platforms where residents can contribute data points to track local air quality. Geographers have closely studied and documented this phenomenon since the 2000s under different terms, including Volunteered Geographic Information, Neogeography, Geoweb and User-Generated Geographic Content.

Shifting the potential to produce geographic information and create maps and related products to a larger number of people has sometimes been characterized as a democratization of geographic knowledge production. To some extent, this might indeed be the case. An initiative like OpenStreetMap, whose mission it is to create a free, editable map of the world, has hundreds of thousands of users across the globe. It is arguably one of the largest repositories of open geographic knowledge and has involved more people in producing map content than ever before. In this way, there are clear linkages between the effect that the geotag has had on knowledge production and what citizen science and participatory approaches to Geographic Information Systems are trying to accomplish.

However, in other cases these new ways of information production can reproduce rather than break down social differences. For example, OpenStreetMap and other mapping platforms alike struggle with gender bias, with male users forming a large majority in these communities. This can and does affect what gets represented within geotagged content and what, sometimes literally, stays unmapped. More globally, access to the internet itself already has a specific, uneven, geography but the production of content for and on the internet is even more unevenly distributed. Content from and about the Global North is overrepresented on many internet platforms that range from conventional websites and Wikipedia to social media and specific-purpose sites such as OpenStreetMap and the social coding platform Github.

As the production of geographic information changed with the advent of the geotag so did the geography of that information. Because of the increasing prevalence of digital geographic information, physical or offline places are now augmented by digital data and services. In other words, physical places are represented through digital information but are also partly made of this digital information. In this way, the physical and the digital co-constitute many of the everyday places we live in. In some cases, this connection is a matter of simple enhancement where a digital layer of data sits on top of the existing, physical space (e.g. restaurant reviews). In other cases, places simply would not be the same without this new digital layer of geographic data and code. This is reflected in how many of our day-to-day decisions related to travel, eating and entertainment are now augmented and changed by various websites, apps and digital services.

Geotags also make visible larger concerns around the social and ethical implications of new digital technologies. In light of these technologies, privacy – and an individual's rights and expectations of privacy – has been intensely debated. Although many digital technologies have similar implications, the geotag makes these implications tangible in important ways. For example, geotags hidden within the metadata of a digital image are often invisible to the user and thus might inadvertently be shared publicly when the image itself is shared, giving away private information on a user's home or other sensitive locations. With geotags attached to a wide variety of other digital data, socio-spatial behaviour of individuals over long periods of time can potentially be reconstructed. Importantly, anonymized data can often be de-anonymized with relative ease when that data contains some kind of geotag.

Geographic research with the geotag

Before the existence of large geotagged data sets, most data collection methods available to geographic researchers were time and resource-intensive. For example, transportation research often relies on the use of travel diaries that are meticulously filled out for a day or a week by willing research participants. Across many of Geography's subdisciplines, geotagged digital data has the potential to alleviate some of these resource constraints by making rich and detailed data accessible to researchers without the need for additional data collection. Although this type of data has only been available for the last 15 years, and we are still gaining an understanding of its limitations, its potential is clear, especially in contexts where it was previously not feasible to collect data through more conventional means.

For transportation research, mobile phone records; taxi, ride and bike-share trajectories; public transportation smart card records; and social media check-ins all contain geotagged metadata that can be used to analyse people's movements at the city, country and even global scale. Similarly, the rich content of social media data, and its relevance across a number of domains, makes it a useful data source for political, economic and urban geography, particularly when paired with the geotag, which allows researchers to see how (the content of) conversations on social media differ across space and time. Such data can be analysed across spatial scales, from studying the global attention to different consumer brands to analysing local variations in the use of amenities and public spaces in urban neighborhoods.

In disaster response, OpenStreetMap and other mapping platforms have been used as a valuable data source as they often contain near-global and real-time coverage on people's use and perception of place. Finally, visual platforms such as the photo-based social platforms Instagram and Flickr allow for the study of (urban) life, not only in public spaces but also inside of buildings – a site that is traditionally more challenging to research.

As geotagged data has become more common in geographic research, this has also brought forward the need to engage with new computational methods and technologies – often referred to as data science. Because of the size of many geotagged datasets, geographers find themselves engaged with new database technologies, distributed computing approaches, cloud computing and other technologies often referred to together as cyberinfrastructure, that are designed to make it easier to read, process and analyse datasets that can extend beyond the capabilities of a single personal computer.

Apart from the sheer size of geotagged datasets, the type of data that can now be used in geographic research has changed as well. For example, web sites and social media platforms alike provide a large quantity of unstructured text content that is challenging to analyse with qualitative research methods developed for small-N contexts. Instead, computational techniques such as Natural Language Processing (NLP) are leveraged to process and gain insight from such datasets.

Similarly, in addition to more conventional satellite imagery, photo-sharing platforms like Instagram and Flickr provide such a large amount of visual data that more conventional visual research methodologies, which rely on the researcher as the primary interpreter of these images, cannot be feasibly utilized. Instead, in these contexts computer vision techniques provide an alternative way to analyse and detect objects within these images. Computer vision is currently (much) more limited than

the interpretation of human researchers but its main advantage (scale and speed) is crucial for processing larger geotagged datasets. This is reflective of a more general trend in analysing geotagged datasets. Rather than solely relying on the design and interpretation of the researcher, both supervised and un-supervised machine learning techniques are employed more frequently to aid in processing and analysing large datasets.

Although the potential of these new datasets is clear, there are also a number of key concerns related to geotagged data. First, while administrative and purposefully-collected datasets often have stringent measures to ensure data quality, these may not exist for geotagged content. A mapping data source such as OpenStreetMap can be more, or less, accurate than more conventional data sources depending on the context. Furthermore, although geotags can be derived from GPS sensors and thus contain exact latitude/longitude coordinates, these coordinates might not always be precise or accurate, especially in urban environments.

Second, as pointed out in the previous section, not everybody is represented equally in geotagged data. The exact nature of this bias is often unknown and because geotagged datasets are generally very large, bias may get snowed under by the relative size (and the supposed statistical power that comes with size). Making inferences based on the analysis of such datasets, specifically if they are to influence policy and practice, has to be done with great care and caution – and only if the bias present within such data can be qualified or quantified.

Third, research with geotagged data brings forward a number of significant ethical questions. For example, conventional social research happens frequently under the auspices of Institutional Review Boards or similar ethical oversight. Participants in such research are only doing so based on informed consent. However, people generally contribute geotagged content to digital platforms for a completely different purpose (e.g. sharing photos with friends) and might not fully realize their content is public, let alone subsequently used in research. Can such data be used in geographic research just because it is public? And if consent would need to be given first, how can the logistical challenges of asking consent from millions of, potentially anonymous, users be addressed? Another ethical concern is the relative ease with which many of these geotagged datasets can be accessed or downloaded. This can potentially change what kind of research geographers are conducting because some types of questions can be more easily answered with this new trove of data, at the expense of important questions that still require more conventional datasets and research methodologies.

Beyond the geotag

Despite the relatively simple concept – a geographic identifier that can be attached to any digital object – and its beginnings in geotagging satellite imagery, the geotag has significantly altered who, how and in what quantity geographic data is being produced. This new data and its associated technologies have not only changed the everyday spaces and places around us but also how we conduct geographic research. This has led to an increasing use of geotagged big data in geography and an increased synergy with computer science and data science field. As a result, we see a renewed interest in quantitative geography that has been labeled as geographic or spatial data science.

Interpreted more broadly, this change could also signify a digital turn within the discipline as whole. The geotag not only enables the adoption of new (quantitative) methods and datasets in our research,

it also warrants new types of geographic inquiry into how the digital (co)produces, alters and reproduces spaces and places and has made the digital itself a site for geographic research.

Further Reading

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53.

Ash, J., Kitchin, R., & Leszczynski, A. (2018). Digital turn, digital geographies? *Progress in Human Geography*, 42(1), 25–43.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130–139.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3), 571–590.

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

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.

Haklay, M., Singleton, A., & Parker, C. (2008). Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, 2(6), 2011–2039.

Leszczynski, A., & Wilson, M. W. (2013). Guest editorial: theorizing the geoweb. *GeoJournal*, 78(6), 915–919.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Shi, L. (2015). Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Annals of the Association of American Geographers*, 105(3), 512–530.

Poorthuis, A., Zook, M. A., Taylor Shelton, M. G., & Stephens, M. (2016). Using Geotagged Digital Social Data in Geographic Research. In N. Clifford (Ed.), *Key Methods in Geography*. SAGE.

Ritter, N., & Ruth, M. (1997). The GeoTiff data interchange standard for raster geographic images. *International Journal of Remote Sensing*, 18(7), 1637–1647.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2011). GeoWeb and crisis management: issues and perspectives of volunteered geographic information. *GeoJournal*, 78(1), 21–40.

Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, 78(6), 981–996.

Zook, M. A., & Graham, M. (2007). Mapping DigiPlace: geocoded Internet data and the representation of place. *Environment and Planning B: Planning and Design*, 34(3), 466–482.

Glossary

GeoTIFF – a standard that describes how a geographic identifier can be embedded in an image stored as in the Tagged Image File Format (TIFF). The geographic identifier is stored as a tag in the metadata of the file. The standard was first proposed by Ritter and Ruth in 1997 to enable the georeferencing of satellite imagery.

Geoweb – the amalgam of digital geographic content, software and platforms that is enabled by the geotag and its use in (Web 2.0) online platforms

User-generated Geographic Content – a term, similar to Volunteered Geographic Information (VGI), used to indicate geographic information produced by individual users of online services. In contrast to VGI, it does not emphasize the supposed voluntary nature of this content production.

Volunteered Geographic Information – geographic information produced by individuals, often on Web 2.0 websites such as Flickr and OpenStreetMap, ostensibly on a voluntary basis.

Natural Language Processing – a machine learning approach to analyze language, often employed to analyze large amounts of text data and, for example, aid the analysis of sentiment or topics without requiring human interpretation.