Mapping Place-Based Context

From “Digital Ethology: Human Behavior in Geospatial Context,”
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

From “Digital Ethology: Human Behavior in Geospatial Context,”
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

6

Geospatial Information Technology Systems for Digital Ethology

Thomas Brinkhoff

Abstract

Today, large amounts of digital data about human activities are generated and stored in databases. These data are often geospatial (i.e., locations on Earth are directly or indirectly referenced). To analyze the digital footprint of human activities in their environment, geospatial information is essential because spatial (and temporal) proximity to events may indicate meaningful relationships. The processing, analysis, and presentation of such information require a deliberate handling of geospatial data as well as the use of suitable software tools and frameworks. This chapter provides a short review of the geospatial information technology (IT) systems that can be used for digital ethology. It introduces the main concepts of geospatial information, presents several types of IT systems for handling geospatial data, and discusses their suitability for digital ethology. Special attention is given to the handling of very large geospatial datasets, to the use of geospatial analysis and aggregation methods, as well as to the application of comprehensible visualization techniques. Besides the usage of out-of-the-box functions, more complex geospatial analyses may need to use application programming interfaces for specific solutions.

Introduction

As introduced by Paus (this volume), the objective of digital ethology is to study human behavior—as well as its constraints and consequences vis-à-vis the built environment—in the natural environment by analyzing its digital footprint. In many cases, behavior and information about the environment relate to some geographical place(s) on Earth. This statement is obvious when one considers that human activities directly influence the built environment, such as when people cover land areas with buildings or a street artist covers the wall of a building with a mural. Many digital datasets also contain direct

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

or indirect spatial references, such as place names, postal codes, and coordinates. Thus, a suitable management and analysis of geospatial data can foster digital ethology. This mainly results from the first law of geography by Waldo Tobler (1970), who stated that "everything is related to everything else, but near things are more related than distant things." The claim "from individuals to communities and back" requires (among others) the handling of very large geospatial datasets, the use of suitable geospatial analysis and aggregation methods, as well as the provision of comprehensible visualization techniques.

In this chapter, the main concepts of geospatial information, georeferencing and geospatial data models are introduced. Discussion then follows on IT systems that are typically used for handling geospatial data, including their key properties as well as their assets and drawbacks for digital ethology.

Geospatial Information

The main characteristic of geospatial information and data is their reference to locations relative to Earth. For example, geospatial information describes the surface of Earth, refers to real-world objects like buildings and bridges, allows the planning of cities or other areas, defines abstract entities like municipal or postcode areas, or describes spatiotemporal events like traffic congestions and floods (Bartelme 2022). Objects with geospatial information are called geospatial features. The digital impacts of human behavior are often geospatial. In many cases, spatial and temporal extensions are important and can occur at different levels of granularity, with exact or fuzzy boundaries. For a digital representation of geospatial information, we need suitable data models. These models are encoded for storing, processing, and exchanging geospatial data.

Georeferencing

The spatial reference of a geospatial feature can be established in different ways (Longley et al. 2015):

1. *Names and codes*: A location is described by place names, address data, code numbers, or similar information. Common codes are postal codes as well as codes for administrative or statistical areas, such as the ISO 3166 (International Organization for Standardization) and NUTS (Nomenclature of Territorial Units for Statistics) used in the European Union. Code schemas subdivide areas and typically have a hierarchical structure. For instance, a NUTS-1 unit consists of one or several disjoint NUTS-2 units. A typical drawback encountered when place names are used from web pages, tweets, or similar sources as data is that the place names are often ambiguous or have vague boundaries (Markowetz et al. 2005).

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

2. *Symbolic reference*: Spatial reference is created through information that reflects a situation in a way that is comprehensible to humans and that refers to other objects. Driving instructions from navigation systems (e.g., "turn right at the next intersection") are an example of symbolic references. Further examples are often contained in social media messages (e.g., tweets, Facebook posts).
3. *Coordinate reference systems* (CRS) provide a framework that allows locations to be described as coordinates. This framework consists, among others, of a mathematical figure approximating the surface of Earth and a horizontal (geodetic) datum for assigning coordinates to points on this surface. For a geographic CRS, a position of the Earth's surface is defined by angular measures related to the equator and prime meridian of the ellipsoid (see Figure 6.1). For example, WGS84 coordinates are geographic coordinates. To display geoinformation on a flat surface (e.g., on paper or on a screen), a mathematical mapping of positions of the Earth's surface onto the plane is required and provides simpler and faster computations than geographic coordinates. Depending on the projected CRS chosen for this purpose, smaller or larger distortions may occur in terms of the location and the size of the area. National institutions and web applications often use projected coordinates.
4. *Linear referencing systems* provide another form of georeferencing that describes positions on a line feature (e.g., a road or pipeline) by a distance measure from a defined point of reference. These distances are typically stored by measures or m-coordinates.

Typically, IT systems require two- or three-dimensional coordinates to represent, exchange, and analyze geospatial data. In addition, operations on linear coordinates are often supported. Missing spatial references are a common

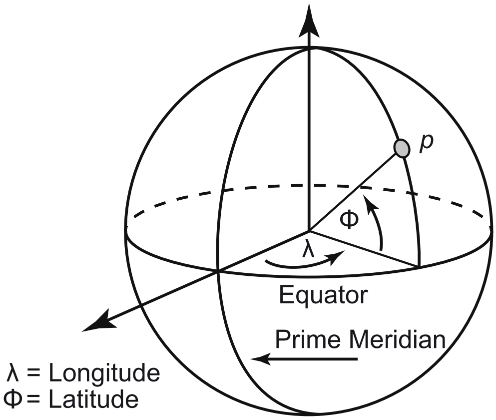


Figure 6.1 Illustration of geographic coordinates.

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

problem. Many photos exist, for example, that show some place on Earth but without a geotag to describe its position.

Modeling Geospatial Information

For an IT system to process real-world information, a suitable data model is required (Herring et al. 2022). To represent geospatial information, its essential properties must be considered. In addition to geometry and topology, nonspatial (thematic) and temporal properties may exist, yet only the first two properties are specific for geospatial data. The combination of space and time is of special importance because it can be used to describe the dynamics of a feature (e.g., the expansion of an urban area).

Similar geospatial features are typically grouped in layers (e.g., buildings, roads, rivers), as illustrated in Figure 6.2. Thematically related layers (e.g., all traffic layers) can form a grouped layer.

Geometry Models

Geometric properties of geospatial data are used to describe the location and extent of a place in space. As illustrated in Figure 6.3, two basic approaches are used: a vector and a raster model (Gröger and George 2022; Herring et al. 2022):

In a *vector model*, points are the base element that generates lines, surfaces, and (3D) solids. Coordinates describe the position of a point, and a sequence of two or more points creates a line. A surface is bounded by one or more closed lines, and it may have one or more holes. Figure 6.3 illustrates the vector model on the left side. A geospatial feature stores a vector geometry by a corresponding attribute.

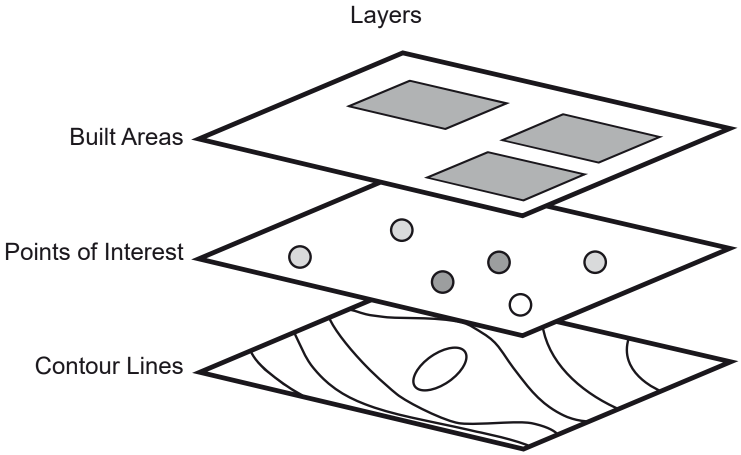


Figure 6.2 Layering of geospatial features.

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

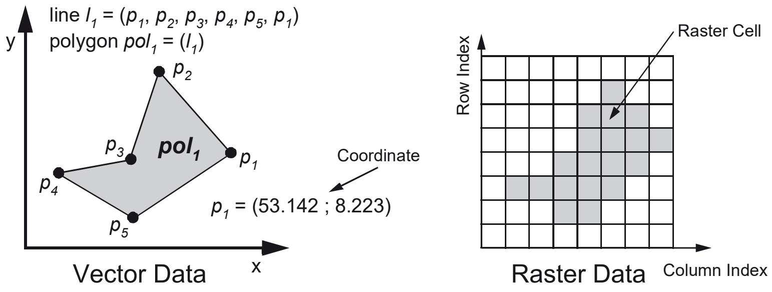


Figure 6.3 Vector model (left) and raster model (right).

In a raster model, coverages are functions from positions in space to values of some type. The most common implementation of a coverage is a raster. It decomposes the data space into similar raster cells (also called pixels), which are usually squares or rectangles and are identified by a column and row index. Each cell stores a single or composed data value. In the case of raster images, this value corresponds to color or brightness. In general, any type of value can be stored in the cells ("raster data"). The spatial reference must be established by georeferencing; for instance, by specifying the coordinates of the corner points of a raster. A georeferenced raster image is called raster map.

The properties of these two geometry models differ significantly. The vector model permits greater accuracy and better resolution scaling. In the vector model, a feature bundles an identifier, its geometry, and other properties. This connection can be used for further analyses. The raster model harmonizes well with important acquisition methods (e.g., aerial or satellite images) and output devices (screen).

For digital ethology, both models are useful. As described by Smith (this volume), vector-based administrative data and raster-based remote sensing data are important digital data sources. Balsa-Barreiro and Menendez (this volume) also list vector data (e.g., mobility patterns, census data, locations from personal wearables, point clouds from laser scanning) as well as raster data.

Topology

Topological properties describe the relative spatial relationships between geospatial features. Typical questions they address include: Which areas touch another area? Which lines intersect an area? Which lines are connected with another line?

Topological properties can either be derived from geometric properties or explicitly represented by a data model. The former will typically be used to address questions related to digital ethology (e.g., which tweets correspond to an area of interest), since typically they do not need to be answered in a precise and consistent form (Kwan 2012). Routing is, however, an important exception. To compute the shortest path or to follow a road network for some distance requires a topological node-edge model. The nodes can represent the points in space and the edges the direct connections between two nodes with their essential properties like distance or travel time. In public health, for instance, this approach can be used to delineate hospital service areas (Wang 2020). In urban analysis, network analyses are used to determine the accessibility of particular areas such as parks for neighborhoods or specific population groups (Unal et al. 2016).

Maps

Maps visualize geospatial information and allow its contents to be communicated (Kraak and Ormeling 2021). For the presentation of geospatial data in maps on screen or in printed form, styling must be defined. Vector data can be visualized using graphical surrogates (symbols). Since points have no extension, they must be made presentable by special point symbols (icons). Specific symbols may be used for illustrating the semantics of line and polygon features.

Appropriate design rules must be defined for the thematic properties of geospatial objects or of raster cells (Buckley et al. 2022). Qualitative properties, which can be represented by (finite) enumerations (e.g., place category), can be visualized by a graduated color scheme or by symbols. Quantitative properties, which originate from a (in principle infinite) number range, can be represented by a color gradient. Alternatively, intervals can be formed, such places with less than 1,000 inhabitants or places with 1,000 to 4,999 inhabitants. Nominal properties such as names and codes as well as quantitative or qualitative properties that are difficult to represent by icons or colors can be added to a map by using labels. In addition to the definition of properties like font and text decoration, the application of label placement rules is important for comprehensible maps (Been et al. 2006). Appropriate design rules can be defined for different scale ranges with respect to a layer.

Generalization is an important concept for the design of maps (Brassel and Weibel 1988) and comprises

- the selection of important information (e.g., only cities with more than 100,000 inhabitants are displayed),
- the simplification, aggregation, and/or classification of data depending on the current scale (e.g., the presentation of individuals vs. the visualization of groups of a minimum size),
- the emphasis of important information (e.g., by using a special symbol or color), and
- the displacement of features so that they do not overlap with other objects (e.g., schematic plans of transport networks abstracted from exact position and emphasize topological connections).

Well-designed maps allow a broad audience to visualize the results of an analysis: from a wide array of experts to common citizens. They help lead the viewer to draw proper conclusions and identify the next steps of analysis.

Standardization and Data Formats

As discussed by Kum et al. (this volume), data access and cleaning are important steps for data analysis. To enable a smooth data exchange or "interoperability," the provided data must be accessible through standardized models and formats (Sondheim et al. 1999). In the field of geoinformation, two organizations play an important role for the standardization of data at the international level (Kresse et al. 2022). The Open Geospatial Consortium (OGC) has established a large number of specifications and other recommendations, many of which are reviewed and ultimately published by the ISO Technical Committee 211 Geographic information/Geomatics (ISO/TC 211) as standards of the 19100 series. Important standards include the following:

- ISO 19107 Spatial Schema is a conceptual data model that describes the spatial properties of geospatial features. ISO 19136 GML (Geography Markup Language) implements this model for interoperable data exchange using XML (Extensible Markup Language). GML is often integrated into an application-specific data model. For example, CityGML (Kolbe 2009), which is a well-known OGC specification for digital city models, follows this approach.
- ISO 19125 specifies a subset of ISO 19107, especially for use in spatial databases and geospatial web services. This simple feature model defines also WKT (Well-Known Text) and WKB (Well-Known Binary) as open encodings for data exchange.
- ISO 19115 Metadata is the accepted metadata model for geospatial data (Brodeur et al. 2019). In addition to obtaining basic information like content, representation, and geometric extent, metadata is useful for accessing quality (Dassonville et al. 2002) and for determining provenance (Beilschmidt et al. 2017).

The data formats GeoPackage and KML (Keyhole Markup Language) are two important OGC standards for data exchange. In addition to the ISO and OGC standards, other encodings are used for geospatial data. For vector data, shapefiles and the text-based GeoJSON (JavaScript Object Notation), formats are most important. For georeferenced raster data, GeoTIFF is often used for representation. The identification of coordinate reference systems is mostly done by EPSG codes.

Geospatial IT Systems

Location is a central aspect of human activities. Thus, digital ethology requires IT systems to analyze, process, and visualize geospatial data. Here, a suitable selection of systems is presented and discussed with respect to their applicability for digital ethology.

Many IT systems are available for processing geospatial data. In a broad sense, each may be referred to as a geographic information system (GIS). This term, however, refers to more specific types of systems. To distinguish them from other geospatial IT systems, the term is used only for traditional GIS in the following sense.

Geographic Information Systems

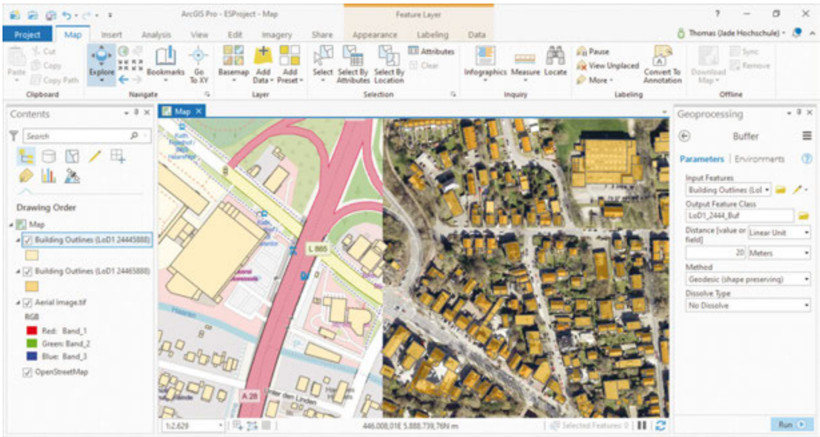
A GIS is a computer-based system designed to collect, manage, analyze, and present geospatial information (Bartelme 2022). The acquisition of geoinformation comprises not only the data input, by using the GIS, but also the import of geospatial data encoded by different data formats. Update functionality is provided as well. Management includes the appropriate description, structuring, storage, and retrieval of geospatial data. GIS supports both the vector and raster models. Data are typically organized by corresponding layers. In addition to the proper datasets, metadata describing the geospatial information (e.g., area coverage, date of origin, data format, acquisition type) should also be provided.

The analysis functionality of a GIS serves to gain information and knowledge from existing geospatial data. It allows measurements, coordinates transformations and geometry analysis functions such as buffering, overlay and nearest-neighbor search, topology analysis functions (e.g., routing), interpolations, approximations, and simulations. Results can be new geospatial datasets, alphanumerical data, statistical evaluations, reports, and other forms of data. A parameterized execution of combined processing steps allows an automation. Users can define workflows in GIS by visual programming languages like the ModelBuilder in ArcGIS.

A GIS provides a graphical user interface. The main component is a map that displays one or several layers according to user-defined styling rules. Temporal developments can be depicted by animated maps.

Well-known GISs include ArcGIS Pro from ESRI, GeoMedia from Hexagon, and the open-source system QGIS. In Figure 6.4, the user interface of ArcGIS Pro is shown as an example. The map is visualized in the center. The layer list on the left contains four layers: two layers are vector layers and represent outlines of buildings. An aerial image is depicted as raster layer in the right part of the map. Open Street Map (OSM) is used as background layer in the left part. The ribbon contains basic GIS tools. The dialog on the right asks for the input parameters of a buffer computation.

From “Digital Ethology: Human Behavior in Geospatial Context,”
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813



Input Features: Building Outlines (Lo|

Output Feature Class: LoD1_2444_Buf

Distance [value or field]: Linear Unit

20 Meters

Method: Geodesic (shape preserving)

Dissolve Type: No Dissolve

Figure 6.4 The user interface of ArcGIS Pro. Attribution for the building outlines and aerial image: Extract from geodata of the Landesamt für Geoinformation und Landesvermessung Niedersachsen ©2023, used in accordance with the data license (https://www.govdata.de/dl-de/by-2-0). OSM used per the Open Data Commons Open Database License by the OpenStreetMap Foundation (https://www.openstreetmap.org/copyright/en).

In digital ethology, a GIS can be used to select, prepare, and combine different types of source data, to analyze and visualize geospatial data, and for data export. For this purpose, a large set of analyses are possible. The map overlay is the traditional analysis operation for combining geospatial information, and has been used since the 1960s to detect areas of urban expansion (Steinitz 2016). To extract geospatial data from individuals, neighborhood and proximity functions are essential. A typical example are studies that investigate whether minority and low-income populations are disproportionately exposed to industrial pollution (Sheppard et al. 1999). In another example, Yin and Shaw (2015) examined the relationships between physical movements and social closeness evolution. Aggregation and cluster algorithms are important to form groups and detect spatial patterns. In a study by Leong and Sung (2015), clustering was used for crime analysis. Charreire et al. (2012) identified built environmental patterns by using a GIS cluster analysis and investigated relationships between walking and cycling facilities and body mass index. Westerholt (2019) estimated hot spots from geospatial social media datasets and found, in contrast to other city districts, that the Asian quarter in San Francisco was a hot spot of messages during Chinese New Year celebrations.

It is important to note, however, that a GIS has limitations for use in digital ethology. For instance, GIS is not designed to handle very large sets of single features. A certain grade of aggregation should thus be done beforehand. Analyses that are very time consuming and may need a distributed processing of data should not be done within a GIS. Furthermore, a user must be aware that a GIS is a proprietary software.

Virtual Globes

Virtual globes are programs that enable the representation of Earth based on a three-dimensional model. In addition to satellite and aerial images, a virtual globe contains further geospatial data (e.g., streets, railroad lines, place names) and georeferenced objects (e.g., photos, Wikipedia articles, 3D building models). Google Earth is a well-known example of a virtual globe.

Virtual globes also allow the creation of mashups by integrating user data. In Google Earth, the KML data format can be used for defining point, line, and area geometries as well as corresponding visualization rules. It is also possible to integrate raster maps and map services as well as user-defined 3D building models.

For digital ethology, virtual globes are helpful for checking hypotheses, validating the quality of given geospatial datasets, for finding explanations for outliers or unexpected results, and for verifying the results of analyses. Figure 6.5 shows the results of an algorithm that computes center points for localities (Brinkhoff 2020). Corresponding locality boundaries and Google Earth's imagery enable a user to check the results of the algorithm.

Spatial Database Systems

Database systems allow the permanent and secure storage of large amounts of information in databases and support efficient retrieval of data. The structure of data in a database follows the specifications of a database model to ensure uniform and consistent storage and update. In particular, the management software supports simultaneous multiuser operations. The relational database model is most commonly used. It allows the retrieval of data by SQL (Structured Query Language). Sometimes the relational model is extended by

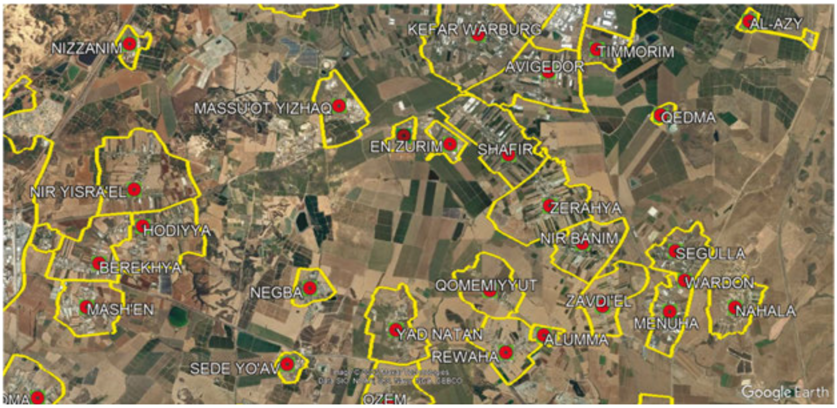


Figure 6.5 The use of Google Earth to compare the computed centers of Israeli localities with the locality boundaries and the globe's imagery. Attribution: Image © 2023 Maxar Technologies Data SIO, NOAA, U.S. Navy, NGA, GEBCO.

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

object-oriented functionality ("object-relational model"). For supporting large-scale application, NoSQL databases are gaining importance.

Spatial database systems allow the integrated storage and spatial retrieval of geospatial data (Brinkhoff 2022; Rigaux et al. 2011). The structure and semantics of geospatial data and the functionality of the spatial database system follow internationally accepted standards to ensure interoperability (ISO 19125 and ISO/IEC 13249-3 SQL/MM Spatial). A spatial database system can be used to manage the data of a GIS. It can also be run independently of a GIS to provide data to geospatial services or applications. Object-relational spatial database systems with extensive functionality are Oracle Spatial and PostgreSQL with the PostGIS extension. First NoSQL databases like MongoDB and Neo4j also support geospatial data (Guo and Onstein 2020). Their functionality is, however, rather limited compared with the aforementioned systems.

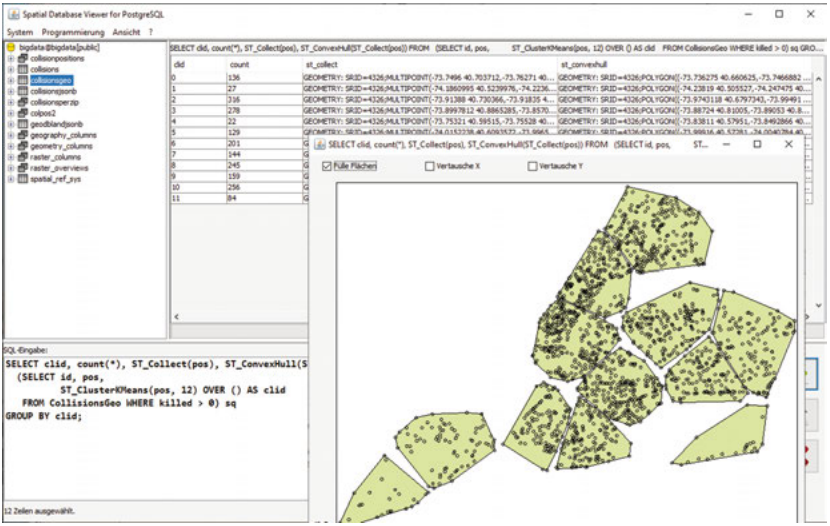
A primary task of spatial database systems is to support spatial queries. An example is the point query, which determines all geospatial features whose geometry contains a given query point. Other examples are window and region queries that compute all features intersected by a given query rectangle and polygon, respectively. A distance query finds all features whose geometry is located within a given distance to a query geometry, and a *k*-nearest-neighbor query (*k*-NNQ) determines the *k* nearest features for a query geometry. A spatial join allows combining two datasets and provides all pairs of features that fulfill a topological or distance condition. For an efficient processing of spatial queries, spatial database systems use spatial indexes like linear quadtrees (Samet 2006) or R-trees (Guttman 1984).

High-level spatial database systems support topological data models, 3D data, raster data, and spatial data mining techniques. The analysis of spatio-temporal data (e.g., for moving objects) is currently not supported by common spatial database systems.

For digital ethology and other fields, spatial database systems are extremely helpful in organizing large and exceptionally large sets of geospatial data, enabling cooperation between different users and software systems, and for applying aggregation and analysis operations on geospatial datasets. In Figure 6.6, the SQL query selects from 1.7 million accidents in New York City those collisions that killed persons and clusters them by the k-means algorithm into 12 spatial clusters. Their convex hulls are depicted.

For a visualization of data, GIS or other tools are required. Although some spatial database systems provide a large set of geospatial functions, their capabilities are limited in comparison to GIS. Despite SQL standardization, the handling of data by spatial database systems involves the use of proprietary solutions. In very large datasets (e.g., geotagged information from social networks), big data analytics can be performed by special frameworks and NoSQL databases (Bordogna et al. 2017; Hoel 2022).

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813



```
SELECT clid, count(*), ST_Collect(pos), ST_ConvexHull(ST_Collect(pos)) FROM  (SELECT id, pos,       ST_ClusterKMeans(pos, 12) OVER () AS clid   FROM CollisionsGeo WHERE killed > 0) sq GRO...
```

clid	count	st_collect	st_convexhull
0	136	GEOMETRY: SRID=4326;MULTIPOINT(-73.7496 40.703712,-73.76271 40....	GEOMETRY: SRID=4326;POLYGON((-73.736275 40.660625,-73.7466882 ...
1	27	GEOMETRY: SRID=4326;MULTIPOINT(-74.1860995 40.5239976,-74.2236...	GEOMETRY: SRID=4326;POLYGON((-74.23819 40.505527,-74.247475 40...
2	316	GEOMETRY: SRID=4326;MULTIPOINT(-73.91388 40.730366,-73.91835 4...	GEOMETRY: SRID=4326;POLYGON((-73.9743118 40.6797343,-73.99491 ...
3	278	GEOMETRY: SRID=4326;MULTIPOINT(-73.8997812 40.8865285,-73.8570...	GEOMETRY: SRID=4326;POLYGON((-73.88724 40.81005,-73.89053 40.8...
4	22	GEOMETRY: SRID=4326;MULTIPOINT(-73.75321 40.59515,-73.75528 40....	GEOMETRY: SRID=4326;POLYGON((-73.83811 40.57951,-73.8492866 40....
5	129	[illegible]	[illegible]
6	201	G	
7	144	G	
8	245	G	
9	159	G	
10	256	G	
11	84	G	

```
SELECT clid, count(*), ST_Collect(pos), ST_ConvexHull(S
  (SELECT id, pos,
          ST_ClusterKMeans(pos, 12) OVER () AS clid
   FROM CollisionsGeo WHERE killed > 0) sq
GROUP BY clid;
```

12 Zeilen ausgewählt.

Figure 6.6 Performing a spatial clustering of collisions in New York City by using the PostGIS extension of PostgreSQL. Attribution: NYC Open Data (https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95).

Geospatial Services

Geospatial web services provide an important service by making geospatial data and maps available to a broad audience. Geospatial data or processing functionalities are available on the Internet (or Intranet) through common web protocols. Spatial database systems usually serve as the data source.

Both closed and open geospatial services are available. Closed services provide geospatial data exclusively for specific applications or libraries. Their protocol is not open and cannot be used by other systems. An example is the aerial and satellite imagery retrieved by Google Earth. The map services used by Google Maps, Microsoft Bing Maps, or Apple Maps also fall into this category.

Open geospatial services are usually based on general geospatial standards. In many cases, they are made available by public institutions, such as surveying agencies or statistical offices (Kresse and Danko 2022):

- The Web Map Service (WMS) (ISO 19128) is a portrayal service that computes user-specified map sections and provides them using common raster and vector map formats.
- The Web Map Tile Service (WMTS) produces raster tiles in predefined bounds and resolutions. This restriction can significantly increase server performance.

- The Web Feature Service (WFS) (ISO 19142) provides vector-based geospatial features that fulfill a given spatial and nonspatial query condition.
- The Web Coverage Service (WCS) provides raster-based coverages according to ISO 19123. A WCS implementation typically supports clipping, scaling, interpolation, and CRS transformation.
- The Web Processing Service (WPS) is a general framework for server-side geospatial computations.

Geospatial services can be integrated into GIS and web mapping applications. For digital ethology, portrayal services can be used to support a visual analysis of other geospatial datasets. The data and processing service are more relevant for data access. The main advantage of open geospatial services is their high grade of standardization and interoperability. Proprietary web services also exist (e.g., the feature service by ArcGIS Server). A WPS may be used to provide specific functionality to a broad range of users.

Geospatial Sensor Data

Human activities influence the physical environment. As described by Smith (this volume), data collected from sensors are a valuable source for deriving measures of the physical and built environment. In the case of *in situ* sensors, the location of the sensor and the observed area are (almost) the same. For remote sensing, these two locations differ. In both cases, however, the location of the observed area is of high importance.

For geospatial sensor data, the OGC has developed an architecture for a geosensor web (Botts et al. 2013). It comprises several data models and geospatial services. First, the Sensor Model Language (SensorML) provides a metadata model to describe sensors and includes information about sensor identification, observable properties (phenomena), and the location of measurement. Second, the Observations and Measurements (O&M) data model (ISO 19156) defines an encoding of observations, including phenomenon and measurements. Finally, the Sensor Observation Service (SOS) allows querying sensor descriptions and observations by spatial, temporal, and further filter conditions.

The SensorThings API (application programming interface) represents another current OGC approach to process and provide geospatial sensor data. It takes concepts from the Internet of Things (IoT) into account (Işıkdağ 2020) and specifies a data model. In this model, a data stream groups observations that refer to the same phenomenon and are measured by the same sensor. The location can be given by the location of the sensor or by a feature of interest that describes the location being observed by the sensor. This may be the same as the sensor's location but it may also differ. The SensorThings API provides a web-based interface for requests and operations and is based on the REST paradigm (Representational State Transfer) and JSON (JavaScript Object Notation). It supports spatial, temporal, and alphanumerical query conditions as well as insert, update, and delete operations. The SensorThings API extends MQTT (Message Queuing Telemetry Transport), which is an important IoT protocol, and enables the transmission of measurements between devices, even if the bandwidth is low or delays occur.

Services for geospatial sensor data allow for the simple integration of large sets of sensor data into other applications. An important field of application are smart cities (Al Nuaimi et al. 2015; Meier and Portmann 2016). Smart cities utilize multiple technologies to improve health, transportation, energy, education, and other services important for their residents. Sensor measurements are a central ingredient for controlling these services as well as the focus of many studies aimed, for instance, at understanding how cities influence social behavior (see Balsa-Barreiro and Menendez, this volume). Privacy issues are of paramount importance and must be observed.

Geospatial sensor data are often not available for all locations in a study area. If the measured phenomenon is continuous, samples can be extrapolated using geostatistical methods provided by GIS, in particular by kriging[1] (Lorkowski 2021).

Geospatial Application Programming Interfaces

Because data processing in the context of digital ethology often requires complex and time-consuming processing steps and algorithms, one solution is to include these into statistical software, data mining software, big data frameworks, or similar packages. Still, the capabilities for processing and visualizing geospatial data vary. If a problem requires more geospatial operations, the integration into a GIS, discussed above, might offer a solution.

Another solution is to program a stand-alone software that uses a geospatial programming library. Such APIs are provided by GIS vendors (e.g., ArcGIS Maps SDK) or exist as independent solutions. For the Java programming platform, the open-source library JTS (Java Topology Suite) is often used as implementation of the simple feature model defined by ISO 19125. Ports into other programming languages are available.

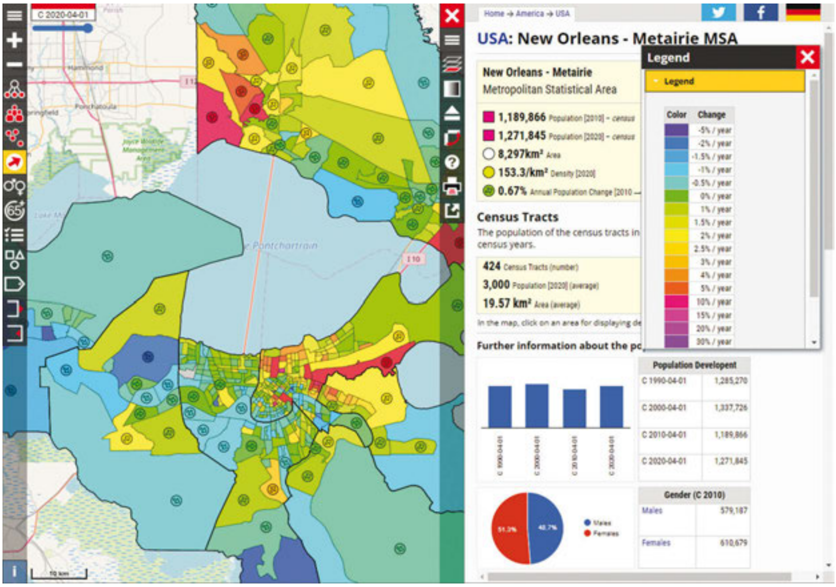
A more comprehensive solution is the open-source GIS toolkit GeoTools. This Java API allows the representation of geospatial features and coverages. They can be uniformly queried from databases and web services. Further processing capabilities, such as coordinate transformations, raster-vector conversions, and rendering, are provided.

---

[1] From a limited set of sampled data points, kriging estimates the value of a variable over a continuous spatial field: (a) the spatial covariance structure of the sampled points is determined by fitting a variogram; (b) weights derived from this covariance structure are used to interpolate values for unsampled points or blocks across the spatial field.

APIs can also be used to present geospatial data and maps via web applications. In the case of web mapping, the map is embedded into a web page. Navigation and information functionality are provided. Geospatial data are usually obtained via services and spatial database systems. The creation of web mapping applications is supported by various JavaScript-based geospatial APIs. For example, most GIS vendors offer specific software packages that can be used to convert a GIS project into a web application. Corresponding APIs also exist for proprietary geospatial services. Another approach is to use independent software libraries. A prominent representative is the free open-source software OpenLayers, which allows the straight integration of various data formats (including GeoJSON, GML, KML, OSM) and open geospatial services (including WMS, WMTS, OSM Tile Service). A popular alternative is the JavaScript library Leaflet, which offers less functionality but is easier to apply.

Figure 6.7 depicts a web page that visualizes the population development of census tracts for the New Orleans–Metairie Metropolitan Statistical Area between the 1990, 2000, 2010, and 2020 census. The original census data were aggregated to small census blocks and larger census tracts by the U.S. Census Bureau for reasons of manageability and privacy. The census tract geometries are also provided by the same agency. To produce the depicted map, several further steps, outlined below, are necessary that use some of the presented geospatial tools.



USA: New Orleans - Metairie MSA

New Orleans - Metairie
Metropolitan Statistical Area

1,189,866 Population [2010] - census
1,271,845 Population [2020] - census
8,297km² Area
153.3/km² Density [2020]
0.67% Annual Population Change [2010 –

Census Tracts

The population of the census tracts in
census years.

424 Census Tracts (number)
3,000 Population [2020] (average)
19.57 km² Area (average)

In the map, click on an area for displaying de

Further information about the po

Color	Change
🟪	-5% / year
🟦	-2% / year
🟦	-1.5% / year
🟦	-1% / year
🟦	-0.5% / year
🟩	0% / year
🟩	1% / year
🟨	1.5% / year
🟨	2% / year
🟨	2.5% / year
🟧	3% / year
🟧	4% / year
🟧	5% / year
🟥	10% / year
🟥	15% / year
🟪	20% / year
🟪	30% / year

Population Developent	
C 1990-04-01	1,285,270
C 2000-04-01	1,337,726
C 2010-04-01	1,189,866
C 2020-04-01	1,271,845

Gender (C 2010)	
Males	579,187
Females	610,679

Figure 6.7 Population development of census tracts in the New Orleans–Metairie Metropolitan Statistical Area. Attribution: Thomas Brinkhoff, City Population, https://www.citypopulation.de/en/usa/metroneworleans/.

From "Digital Ethology: Human Behavior in Geospatial Context,"
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

Census tracts of different years differ in their boundaries and are not immediately comparable. For computing adapted population figures, previous census blocks need to be assigned to 2020 census tracts. This task can be solved by overlaying the census block polygons with census tracts polygons. Another (simpler and more robust) approach is to determine a representative point for a census block and use it for a unique assignment to a census tract. The U.S. Census Bureau provides a centroid for a census block. Since centroids may lay outside of the original geometry, they are not really suitable for this task. Instead, it is better to determine (exactly or approximatively) the "visual center" of a polygon. In Brinkhoff (2020), the method of Garcia-Castellanos and Lombardo (2007) is favored because it can be programmed using an API like JTS or a script within a GIS.

The original census tract polygons are too bulky for a web application and need to be generalized for this purpose. An individual generalization of polygons would, however, produce gaps and slivers between the polygons. Thus, a topological data model has to be constructed before performing the generalization. Comprising parish polygons can be neatly computed by merging related census tract polygons.

The map is rendered by using the OpenLayer API, which requests the polygons from a geospatial service and retrieves them from a spatial database system. The original CRS is WGS84. The integration of other geospatial services works best, however, with the Pseudo-Mercator projection. Thus, the web mapping API transforms the coordinates. The background OSM is requested as raster tiles. Other background maps can be integrated by a user by specifying a WMS or a WMTS service. For the visualization, a suitable color gradient is defined. Arrow icons depict in the map the population increase or decrease. They are placed on the visual center of the corresponding polygon. Only icons that fit into the corresponding polygon are shown.

Conclusions

This chapter has highlighted geospatial IT systems that can be used for digital ethology. For such analyses, vector as well as raster data are often required. To achieve a high grade of interoperability, geospatial standards for data models and data formats should be observed. This requirement concerns not only the access to input data but also the provisioning of research results and can be fulfilled by using standardized geospatial web services.

GIS is the basic tool for geospatial analyses. It supports the acquisition, management, analysis, and visualization of geospatial data. For digital ethology, the combination of various databases as well as neighborhood and proximity functions are essential. The latter group of functions is often accompanied by network analyses. Aggregation and cluster algorithms provided by GIS are important for forming groups as well as for detecting spatial patterns and hot spots. Results can be new geospatial datasets, alphanumerical data, statistical evaluations, reports, and other types of data. In order to check hypotheses or validate data quality as well as to find explanations for outliers or unexpected results and to verify the results of analyses, we need a suitable visualization of geospatial data by maps in GIS or in virtual globes.

For large geospatial datasets, the use of spatial database systems is advisable as they support spatial queries and provide a basic (or in some cases a rather extensive) set of geospatial analysis functions. Spatial database systems also serve as a data source of geospatial web services. Sensor data are an important source for data about the environment and human activities. Geospatial standards and frameworks facilitate the access and the analysis of such sensor data.

For repeated data access or complex analyses, an automated execution of combined processing steps is necessary. Visual programming languages in GIS support the definition of such workflows. For the specification of more complex geospatial analyses as well as for geospatial web applications, several application programming interfaces exist.