

# Mehr Raum für die Statistik

„Es ist nichts, was den geschulten Verstand mehr kultiviert und bildet, als Geographie“

Immanuel Kant

Im Navigationsgerät im Auto, bei der Anfahrsbeschreibung eines Webauftrittes oder im mobilen Telefon bei der Aufzeichnung der Laufrunde, Geodaten begleiten uns auf Schritt und Tritt. Bei der Ermittlung und Auswertung von Daten in einer immer stärker digitalisierten und vernetzten Gesellschaft kommt dem „Wo“ eine stetig zunehmende Bedeutung zu. Die Wirtschaft nutzt Geographische Informationssysteme (GIS) schon lange zur Optimierung von Standorten, des Vertriebes und der Logistik.

Neben der Entfernung zu Ressourcen und Kundinnen und Kunden spielen für die Ermittlung eines Standortes auch das Straßennetz oder weitere demographische Faktoren der umliegenden Wohnbezirke eine bedeutende Rolle. Die raumbezogenen (Geo-) Daten können durch die Kombination mit Merkmalen wie beispielsweise dem durchschnittlichen Haushaltseinkommen in der näheren Umgebung angereichert werden und damit zu einer ausschlaggebenden Entscheidungsgrundlage für oder gegen einen Standort oder das Zielgebiet einer geplanten Marketingkampagne werden. Auch der öffentliche Sektor kann die Erkenntnisgewinne durch den Verschnitt von Geo- und Fachdaten zweifellos für seine Aufgaben nutzen. Die Schadstoffbelastung einzelner Wohnbezirke oder die Gefährdungszonen von Naturkatastrophen lassen sich ebenso mittels GIS darstellen wie die Abbildung von Bevölkerungszahlen oder von aktuellen Baustellen im Straßennetz.

## Geokodierungsprozess

Die statistischen Ämter des Bundes und der Länder, welche durch die Änderung der §§ 10 und 13 Bundesstatistikgesetz (BStatG) im Jahr 2013 die rechtliche Möglichkeit erhalten haben, statistische Daten mit Geodaten zu verknüpfen, sind ein Teil der öffentlichen Dateninfrastruktur. Die amtliche Statistik bietet einen außerordentlich großen Quell an Informationen. Das Wirtschaftswachstum und die Bevölkerungszahlen werden ebenso erfasst wie die Kartoffelernte, die Baufertigstellungen oder die Einrichtungen der Kinder- und Jugendhilfe, um nur ein paar Beispiele zu nennen.

Die statistischen Daten liefern eine Entscheidungshilfe für eine Vielzahl von Interessengruppen. Für eine Familie mit kleinen Kindern kann es womöglich von Interesse sein, Kindertagesstätten auf einer Karte zu finden, die innerhalb einer bestimmten Fahrzeit um den eigenen Wohnort liegen und somit für die Familie in Frage kommen. Ein weiteres Beispiel, um bei den Kindertagesstätten zu bleiben, lässt sich mit der Berechnung der Bevölkerung unter 6 Jahren im Umkreis von zwei Kilometer konstruieren. Damit könnten Personen mit politischer Entscheidungsbefugnis auf

eventuelle Versorgungsengpässe in Teilen von Gemeinden aufmerksam werden, die bei einer rein tabellarischen oder gemeindeschaffen Betrachtung der Daten vielleicht nicht ins Auge fielen. In der amtlichen Statistik werden neben den statistischen Daten regelmäßig auch die Anschriften der Erhebungseinheiten als Hilfsmerkmale erhoben. Diese können mittels Geokodierung für die GIS lesbar gemacht werden. Für die Bestimmung der Parameter bei der Geokodierung gilt es, kartographische Grundlagen zu beachten.

Bei der Projektion der mehr oder weniger kugelförmigen Erde auf eine zweidimensionale Karte müssen Verzerrungen<sup>1)</sup> bei dem Abbild in Kauf genommen werden. Je nach Thematik der Fragestellung gibt es zahlreiche Möglichkeiten zu projizieren. Die bekannteste Projektion stellt vermutlich die Mercator-Karte dar, auf der die Fläche stark verzerrt ist. Die Insel Grönland wirkt hier beispielsweise in etwa so groß wie Afrika, obwohl Afrika in Wirklichkeit ungefähr 14 Mal größer ist. Das kommt daher, dass der Nordpol, eigentlich ein Punkt, bei dieser Projektion genauso lang gezogen wird wie der Äquator.

Sobald das räumliche Bezugssystem festgelegt ist, können im nächsten Schritt für die Anschriften aus den Erhebungen mit Hilfe von speziell dafür angebotenen Anwendungen die Hoch- (X) und Rechtswerte (Y) ermittelt werden, das heißt die eigentliche Geokodierung durchgeführt werden. Diese Koordinaten werden den ursprünglichen Daten angehängt und ermöglichen dann das Einlesen dieser, nun zu Geodaten gewordenen Informationen in eine GIS-Umgebung.

In einem GIS werden diese geokodierten Daten zunächst als Punktwolke angezeigt. Für die endgültige Speicherung der Daten ist im BStatG eine Auflösung von mindestens 100 x 100 Metern vorgesehen, d.h. die Koordinaten dienen lediglich als Hilfsmerkmal zur Zuweisung zu Gitterzellen. Dafür wird das Gebiet (in unserem Fall das Land Niedersachsen) mit einem Raster aus 100 x 100 Meter großen Gitterzellen abgesteckt. Die Benennung der Gitterzellen ergibt sich aus den Koordinaten des süd-westlichen Eckpunktes der Rasterzelle mit den Bezeichnungen „N“ (North) für den Hochwert und „E“ (East) für den Rechtswert.

Dazu kommen noch Informationen zu dem genutzten Koordinatenreferenzsystem (CRS<sup>2)</sup> – hier wurde das von der EU für die Einhaltung der INSPIRE-Richtlinie<sup>3)</sup> vorge-

1) Grundsätzlich können drei Verzerrungen unterschieden werden, die auf einer Karte niemals vollständig beseitigt werden können: Winkeltreue, Flächentreue und Längentreue.

2) Coordinate Reference System.

3) INSPIRE = Infrastructure for spatial information in Europe. Ziel der INSPIRE-Richtlinie ist es, die EU-weite Dateninfrastruktur räumlich kompatibel und ohne Schnittstellenprobleme auswerten zu können. Beispielsweise soll eine bessere Vorbereitung auf und ein zielgerichteter Umgang mit grenzübergreifenden Naturkatastrophen über die administrativen Grenzen hinaus ermöglicht werden.

schriebene Raumbezugssystem „LAEA<sup>4)</sup>“ mit dem EPSG<sup>5)</sup>-Code „3035“ verwendet) und über das dargestellte Raster (RES<sup>6)</sup> – die Kantenlänge der hinterlegten Rasterzellen). Daraus lässt sich die eindeutige Gitterzellen-ID ableiten:

**CRS3035RES100mN12300E54300**

Diese Gitterzellen-ID kann nun als Hilfsmerkmal für die räumliche Lage des Fachdatums dauerhaft gespeichert werden und zukünftig für Geodaten-gestützte Auswertungen Verwendung finden. Bei der Verbreitung von Ergebnissen ist jedoch weiterhin die statistische Geheimhaltung maßgeblich, so dass für Auswertungen und Veröffentlichungen oftmals größere Gebietseinheiten, z. B. abgeleitete Gitter mit größerer Gitterweite von 1 x 1 Kilometer bzw. 5 x 5 Kilometern gebildet werden.

### Auswertungspotentiale

Durch die Nutzung gleich großer Gitterzellen sind Auswertungen möglich, die nicht an die administrativen Grenzen von Gemeinden, Postleitzahlgebieten, Wahlkreisen etc. gebunden sind. Gleichwohl ist die Auswertung auf diesen Ebenen weiterhin möglich und bleibt je nach Fragestellung auch durchaus zielführend.

Für die Aufbereitung der Daten können die Informationen in einem GIS flächenmäßig visualisiert werden. Dabei werden die Flächeneinheiten (Gemeinden, Gitterzellen, etc.) in einer Farbgebung je nach Ausprägungsgrad des beobachteten Merkmals dargestellt.

Abbildung A1 zeigt die Auswertung der Gitterzellen (5 Kilometer Kantenlänge) nach dem Anteil des Weizenanbaus an der landwirtschaftlichen Fläche insgesamt. Die Sättigung der Farbe steigt mit zunehmendem Anteil und lässt so auf den ersten Blick ein Muster erkennen. Der verhältnismäßig anspruchsvolle Weizen wird in Niedersachsen hauptsächlich in den (Sand-) Lössgebieten sowie den küstennahen Gebieten des Marschlandes (Kalkmarsch), die beide sehr fruchtbar sind, angebaut. Anhand dieser Karte wird auch deutlich, dass sich oftmals Auswertungen anhand von anteiligen Darstellungen anbieten.

Bei der Betrachtung der Verteilung von absoluten Werten, wie beispielsweise der Anzahl der sozialversicherungspflichtig Beschäftigten im Gastgewerbe, werden in der Regel die bevölkerungsstärksten Gebiete, wie Großstädte, hervorstechen, auch wenn die Bedeutung dieses Wirtschaftszweiges für eher dünn besiedelte und periphere Regionen, wie die Lüneburger Heide oder die küstennahen Landstriche, proportional deutlich größer ist.

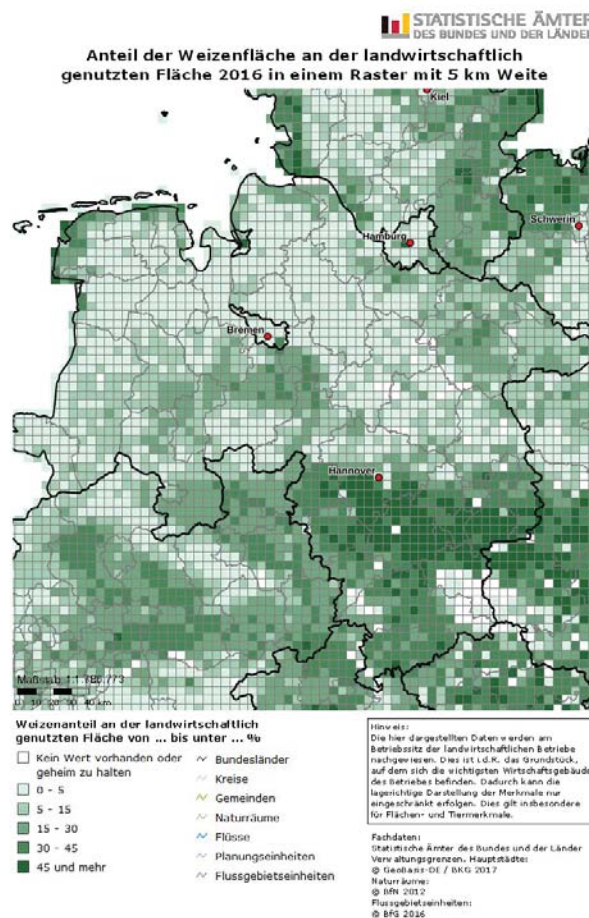
4) Lambert Azimuthal Equal-Area Projection.  
5) EPSG (European Petroleum Survey Group Geodesy) hat eindeutige Schlüsselnummern für Raumbezugssysteme vergeben.  
6) Resolution.

### Ausblick

Die Digitalisierung der Arbeitsprozesse und der Kommunikation sowie das Voranschreiten des Internets der Dinge erzeugen immer größere Datenmengen. Für die amtliche Statistik ergeben sich daraus neue Möglichkeiten der Datenerfassung und -auswertung. Ein aktuelles Projekt testet z. B. die satellitenbildgestützte Erfassung landwirtschaftlicher Flächen. Eine Realisierung dieser Methodik könnte nicht nur den Statistikämtern die Arbeit erleichtern, sondern böte ggf. auch für die Berichtspflichtigen Entlastungspotential, wenn dadurch weniger Erhebungsbogen ausgefüllt werden müssten.

Ein weiteres Feld, welches nicht nur in der Statistik diskutiert wird, ist die Nutzung und Auswertung von Big Data. Hier bieten sich ganz neue Datenquellen und Erhebungsfelder an. Die Verortung von Mobilfunkdaten kann eine gute Einschätzung über die tatsächlichen, kleinräumigen Aufenthalte der Bevölkerung im Tagesverlauf oder bei akuten Krisensituationen ermöglichen. Die amtliche Statistik ist aufgrund ihrer Verpflichtung zur Geheimhaltung sehr geeignet diese sensiblen Daten auszuwerten und der Öffentlichkeit anonymisiert, transparent und wertungsfrei zur Verfügung zu stellen.

### A1 | Anteil des Weizenanbaus an der landwirtschaftlich genutzten Fläche



(Quelle: <https://www.atlas-agrarstatistik.nrw.de/>)