



Data Sources

1

Vít Pászto, Andreas Redecker, Karel Macků, Carsten Jürgens,
and Nicolai Moos

Abstract

This chapter is devoted to the overview of the data fundamentals as regards data models and sources accompanied by geomatics, remote sensing, and economy. Description of such data sources is complemented with the basics from respective disciplines to provide a thematic context to the reader. The chapter starts with a summary of the most commonly used data models, starting with tabular and attribute formats. It is then followed by the spatial data models, including vector and raster data core principles. Since the geospatial domain is heterogeneous in terms of different data formats, the list of interoperability data sources and services is provided. Emphasis is also given

to the international and selected national data sources, both non-spatial and spatial. This part is mainly covering the economic (socio-demographic) topics. At last, a remote sensing perspective on data sources is introduced, pointing out the most important Earth observation data. The whole chapter focuses on the major data models and sources, so it serves as a gateway to further exploration of existing data storages.

Keywords

Data models · Formats · Data sources · Data portals · Satellite archives

V. Pászto (✉)

Department of Informatics and Applied Mathematics,
Moravian Business College Olomouc, Olomouc, Czech
Republic

Department of Geoinformatics, Palacký University
Olomouc, Olomouc, Czech Republic
e-mail: vit.paszto@gmail.com

A. Redecker · C. Jürgens · N. Moos
Geography, Geomatics Group, Ruhr-University Bochum,
Bochum, Germany
e-mail: andreas.redecker@rub.de;
carsten.juergens@rub.de; nicolai.moos@rub.de

K. Macků
Department of Geoinformatics, Palacký University
Olomouc, Olomouc, Czech Republic
e-mail: karel.macku@upol.cz

1.1 Data Models

1.1.1 Basic Tabular and Attribute Data Formats (by Vít Pászto)

In this section, the most used data formats will be briefly introduced. Some of the data providers offer several options regarding data formats. Therefore, it is liable to mention the main characteristics of such formats.

1.1.1.1 TXT

This is the most common data format using plain text. The text could be supplemented by the special symbols for row endings, blank spaces, and tabulators. The suffix for this data format is .txt. Since the format is mainly plain text (with very

limited options for formatting), it is possible to open the .txt file in most of the software and even with the simple text editors (like Wordpad). Thus, the greatest advantage of this format is its interoperability.

1.1.1.2 CSV

Comma Separated Value (CSV) is a simple and standardised format for data storage. Individual records are separated by comma (in some case by a semicolon, blank space, or another tabulator) and the format is classified as a delimiter-separated format family. Most of the tabular software is capable of working with CSV. The format is interoperable, interchangeable and in most cases in the form of a plain text (storing both text and numbers).

1.1.1.3 XLS/XLSX

Files with .XLS/XLSX extension are formats of Microsoft Office package, namely with Excel, and is one of the most used and widespread format. The data is stored in tables, which are organized in spreadsheet and sheets. Label XLS/XLSX is basically “only” a suffix for an Open Office XML scheme (OOXML). The format is binary (i.e. needs specialised software/plugins to be opened), while XLSX represents zipped XML file, and was introduced by Microsoft in 2007. Data stored in XLS could be still opened in the newer Excel version. Thus backward compatibility is secured.

1.1.1.4 XML

This abbreviation stands for eXtensible Markup Language (XML) using structured markup language with constructs such as tags, elements, and attributes. Since its introduction in 1996, XML has become a basis for many other formats (e.g. XHTML, SVG, KML, Microsoft Office, OpenOffice and others). XML files are used mainly for data exchange due to its simplicity, openness, and platform independence. Moreover, the format is machine-readable, easily and quickly searchable and convertible to other formats. Most of the metadata is stored in XML files.

1.1.2 Spatial Data Models (by Andreas Redecker)

Performing scientific analysis implies the use of data and systems that can process this data to gain new insights into the characteristics and interdependencies of research objects. Taking advantage of the information on where an object is located and how it is delimited leads to the field of spatial analysis. It implies the use of a Geographic Information System (GIS) that can process a special type of data referred to as spatial data, geospatial data, geographic data or just geodata.

The understanding of the meaning of GIS varies from “just an application” to “a system of hardware, software and geodata”. The latter refers to the fact that besides a particular program, also the data used has to be suitable for spatial analysis. This can also apply to the requirements for the hardware, depending on the kind and size of the geodata utilised. Considering the available data and the aims of a spatial analysis, it can be necessary to use different software products to apply the appropriate methods to the geodata.

Depending on the source and purpose of the geodata there are two completely different models to represent objects from the real world: Raster data and vector data.

1.1.2.1 Raster Data

Raster geodata represents an area in the real world by an array of square cells with a certain edge length referred to as resolution in ground units (mostly meters). It is spatially referenced to the real-world space by the coordinate of the centre of the upper-left cell and – if necessary – rotation angles for orientation (Fig. 1.1).

For each raster cell value can be stored that represents the characteristic of the represented object within the area of the cell. These values can be of different numeric types like integer or float to represent the desired properties of the object such as height, temperature, brightness etc. or to document codes for classes of land use as a result from a classification process. Different rasters with the same geometric properties can be

superimposed to constitute a layer stack like a common colour image that consists of three layers, each of them representing one colour of red, green and blue.

To save raster models digitally many different data formats are available as with imagery from any kind of digital camera. These can be complemented with geospatial parameters in additional files of the same name but with a different suffix.

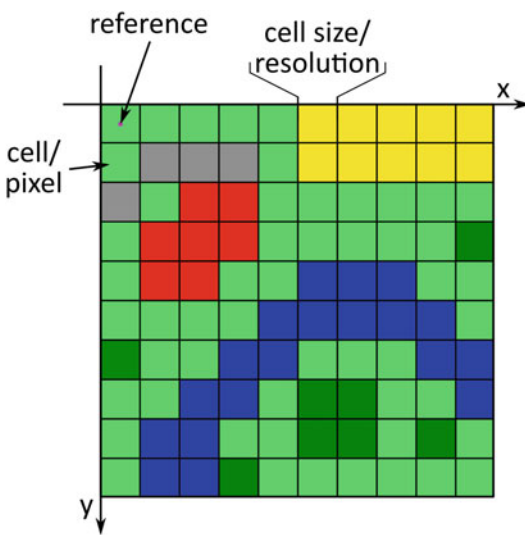


Fig. 1.1 Schematic example of a raster geometry in an image coordinate system. (Source: Author)

Special geodata related file types for raster models are holding the geospatially relevant information within the header of the file (Table 1.1).

All of these file types support compression techniques to reduce the amount of data that has to be stored on storage systems like SSDs or HDDs. Some of the codecs (coder/decoder) that are used to compress and uncompress the data are not able to completely recover the original condition of a raster and are referred to as lossy codecs. For imagery that only has to be viewed visually, these might suffice. But for most geospatial analyses of raster data, the use of lossless compression codecs is vital.

1.1.2.2 Vector Data

Vector Data – also referred to as feature data – represent individual objects (features) of the real world. These are modelled as geometries at a certain location holding attributes about their specific properties. A collection of similar features with a same set of properties forms a feature class.

Depending on the geometric dimension of the objects modelled, a feature class consists of points, lines or polygons.

- Points are defined by x-, y- and – if desired – z-coordinates describing the location of an object. They do not have an extent.

Table 1.1 Common raster-formats, extensions and codecs for geodata (Source: author)

	File extension	Spatial reference in an external file	Spatial reference stored internally	Available codecs for compression	quality
Tagged Image File Format (TIFF)	.tif	.tfw		LZW	lossless
Geo-TIFF	.tif		x	LZW	lossless
Graphics Interchange Format (GIF)	.gif	.gfw		LZW	lossless
Portable Network Graphics (PNG)	.png	.pgw		DEFLATE	lossless
Joint Photographic Experts Group (JPEG File Interchange Format) (JPEG/JFIF)	.jpg	.jgw		JPEG	lossy
JPG2000	.jp2		x	JPEG 2000 (wavelet)	lossless/lossy
Image (IMG)	.img		x	DR-RLE	lossless
Enhanced Compressed Wavelet (ECW)	.ecw		x	wavelet	lossy
MrSID	.sid		x	wavelet	lossless

- Lines are represented by a series of connected points (vertices). They describe a pathway with a direction and a certain length, but also no expanse.
- Polygons are described by a line with a joint start- and endpoint. They describe an area with a certain acreage and a perimeter accordingly.

Every feature class consists of the mentioned geometries and an attribute table connected to these. Each row (a record) in the attribute table – together with the corresponding geometry – represents one object or feature respectively. With so-called multipart features, several geometries can make up one object connected to one record in the attribute table (Fig. 1.2).

1.1.2.3 Tabular Data

In addition to geodata with a direct spatial relation (expressed by the coordinates of points or vertices), other data with only indirect spatial relation can easily be incorporated in GIS-analyses. Here indirect spatial relation refers to an attribute that can be linked to a feature class holding the same information in its attribute table. Indirect spatial relation can be realised by administrative codes, ids, addresses etc.

The simplest file-type to hold this kind of data is a text file with separated values. Here a special

character is used to delimit the columns within each line of the file.

If tabular data directly contains columns holding coordinates of a known spatial reference system, in many GIS, it can be directly transformed to vector-geodata (point features).

1.1.2.4 Topology

A special characteristic of some vector-geodata models is the ability to deal with topology. Meaning the GIS verifies the compliance with predefined geometric relations between features in certain feature classes. For example, there is a rule that there shall be no overlap of features nor any gaps between the representations of administrative areas.

Vector-geodata gets stored in many different ways. These are mainly dependent on the application they are used in. Nevertheless, there is at least one quite common but simple format, which is supported by almost every GIS system.

1.1.2.5 The Shape-Format

Initially, the Shape-format was introduced by the company ESRI as a simple data structure for the exchange of vector geodata (ESRI 1998). In the meantime, many other providers of GIS-applications have adopted it to provide a simple interface for the import and export of geodata or to provide a modest data structure for small projects. The Shape-format does not support topology. Each feature class can only hold features of one geometric type. Information on the spatial reference system for the coordinates used in a dataset is not obligatory but at least possible. Many providers of geodata utilise this format to provide data product-independently. A shape-feature class consists at least of three obligatory files with the same name but with different suffixes:

- *.shp: the main file, holding geometries
- *.dbf: attribute table in dBase-format
- *.shx: index file for the link between geometries and attributes

Additional information get stored in further optional files like:

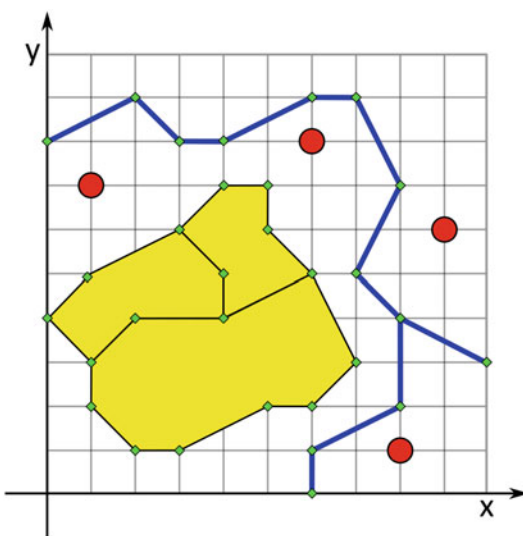


Fig. 1.2 Schematic example of the three different feature geometries (point, line, polygon). (Source: Author)

- *.sbn/*.sbx: spatial index (generated automatically)
- *.prj: spatial reference system
- *.shp.xml: metadata

1.1.2.6 Geodatabases

For the efficient use of geodata in (larger) projects, almost every GIS supports some kind of geodatabase. Geodatabases are database management systems (DBMS), which support the handling of spatial data. Some of them even offer functions for geospatial analysis directly within the DBMS. Most geodatabases can store raster geodata as well as vector data. Furthermore, they provide features to organise data like in folder structures and take care of spatial reference systems and topologies.

1.1.2.7 Spatial Reference Systems (SRS)

The spatial reference of geodata consists of coordinates that are related to the earth's surface by some kind of coordinate system. For this, a mathematical model of the earth's shape is required, which the coordinate system can be linked to. Usually, according to the earth's form, this model is a "flattened" (oblate) ellipsoid (of revolution), mostly defined by the parameters of its semi-major axis and its inverse flattening. Sometimes an additional gravitational model is applied to account for divergences between the ellipsoid and the geoid – the earth's real appearance (Snyder 1987).

The geodetic datum describes the linkage between the geoid and the idealised shape of the ellipsoid. It consists of the ellipsoids parameters and those for its orientation related to a known precisely measured point or a network of precisely measured locations on the earth's surface.

The internationally most common datum for geodata is the World Geodetic System 1984 (WGS84). In Europe, the European Terrestrial Reference System 1989 (ETRS89) defines the reference for coordinates of current geodata. It is based on the Geodetic Reference System 1980 (GRS80) that consists of a reference ellipsoid and a gravity field model like the WGS84.

To locate positions on the earth's surface by coordinates geographical or projected coordinate systems are applied to the modelled surface (Fig. 1.3).

1.1.2.8 Geographical Coordinate Systems

Geographical coordinates relate to a grid composed of vertical and horizontal circles around the earth – the so-called parallels and meridians – as a base for coordinates measured in degrees referred to as latitude and longitude.

Latitude describes a location's distance to the equator measured parallel to the earth's axis. The longitude measures its distance parallel to the equator related to the base meridian, mostly defined by the meridian that crosses the location of the Royal Observatory in Greenwich.

1.1.2.9 Projected Coordinate Systems

For easier reading of maps and plans and less complex computing of distances and areas, projected coordinate systems provide a flat rectangular grid (Cartesian coordinate system) as a reference for measurements in metric units.

For Europe, the Universal Transversal Mercator (UTM) Projection (ETRS89-TMzn, EPSG-Code 3038-3051) is the official reference system for conformal pan-European mapping with scales

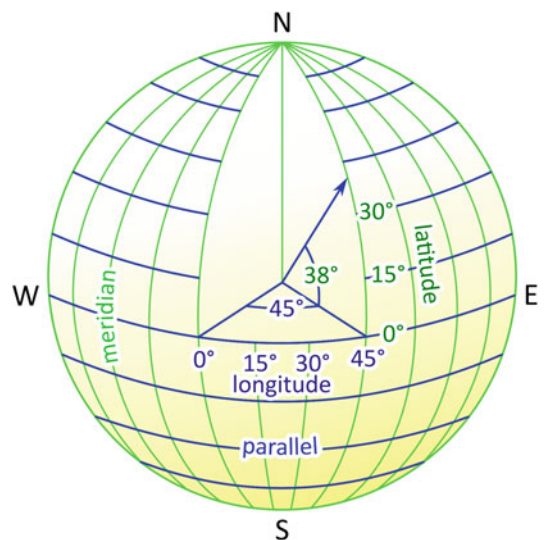


Fig. 1.3 Illustration of a geographical coordinate system. (Source: Author)

larger than 1:500,000. Less detailed maps are recommended to be drawn using Lambert conformal conic (ETRS89-LCC, EPSG-Code 3034) for conformal pan-European mapping at scales smaller or equal to 1:500,000 or using Lambert Azimuthal Equal Area Projection (ETRS89-LAEA, EPSG-Code 3035) for true area spatial representations in pan-European spatial analysis and reporting. All three of them are linked to the geoid – represented by the GRS80 – through the ETRS89 (European Commission 2014a).

The UTM-System covers Zones of 6° width by superimposing the so-called prime meridian of the zone with the vertical line at $x = 500,000$ m of the coordinate system. This practice of so-called false easting avoids calculations with negative values west of the prime meridian within a zone. The counting of zones starts at the International Date Line with the first prime meridian at 177° west of Greenwich. Hence zone 32 covers the zone three degrees west and east about the meridian 9° east of Greenwich.

Y-coordinates refer to the zero-latitude thus representing a location's absolute distance to the equator in meters.

1.1.2.10 Application of Geodata Models and Formats

Besides the technical properties of geodata, they can also be distinguished by their content. Typical fields of applications for raster data are for example, imagery, height models, land use classes, population data, atmospheric parameters like temperature, precipitation etc.

1.1.2.11 Imagery

The results of imaging sensors like cameras or scanners are stored in raster datasets. In this context, the values of the raster cells or pixels, respectively, sometimes are referred to as digital numbers (DN). They represent the quantised intensity of electromagnetic energy that the sensor was exposed to. Depending on the amount and range of the energy recorded, they are positive integer numbers of different bit depths defining the number of gradations between the lowest

and the highest signal value. This defines the radiometric resolution expressed in bits of binary numbers. Standard bit depths are 8 bits representing 256 levels for consumer cameras and up to 16 bits representing 65,536 levels used with professional sensors.

Further aspects of digital imagery are explained in Sect. 1.5.

1.1.2.12 Digital Elevation Models

There are different kinds of models representing continuous surfaces. These digital elevation models (DEM) are differentiated as:

- DSM: Digital surface model, describing the height of the earth's surface, including all objects in the landscape.
- DTM: Digital terrain model, representing the terrain without vegetation or human-made objects.
- DHM: Digital height model also referred to as normalised DSM (nDSM) having the heights of all objects on the bare DTM (resulting from the calculation DSM-DTM).

The values of DEMs are usually of some floating-point data type to allow negative values as well as decimal numbers.

The raster-model is very common to represent this kind of geodata, but there is also a special vector data model for surfaces. Triangulated irregular networks (TIN) express surfaces by triangular areas resulting from a network formed by lines connecting mass-points of known heights.

1.1.2.13 Network Datasets

Another special vector based model for geodata is a network dataset. It is a collection of different vector feature classes and tables containing the all necessary information for performing network analyses: The network itself (holding attributes for the impedance of the edges), possible turns, barriers etc. Further information on this kind of geodata can be found in Part I, Sect. 3.3 in Chap. 3.

1.1.3 Geodata Interoperability (by Andreas Redecker)

For the exchange of geodata, it is vital to have data structures and methods that follow standardised rules. With these providers can advertise the properties of their data in a mutually intelligible form to potential users on the one hand. On the other hand, agreed formats and data structures allow the exchange of the data between different systems that internally might operate with individual, i.e. proprietary data models.

Many geodata is highly dynamic, and the exchange of that information can be very time-dependent. Therefore besides the exchange of files geodata more often are provided as services. That means that a user can directly use a provider's data by accessing it via a network. After receiving a standardised request, the provider's system will transfer the desired information to the user in a standardised format. This can be metadata about the data provided as well as the desired data itself.

Besides proprietary protocols standardised request and transfer methods are commonly used especially within public infrastructures. The central organisation that defines most of the standards to describe and transfer geodata is the Open Geospatial Consortium (OGC, <http://www.opengeospatial.org/>).

For geodata services, special formats support the delivery of spatially or thematically limited extracts of a provided dataset. Some of them even support the streaming of the data to be able to transfer large amounts, especially with raster data.

The most important standards that allow real-time access to (distributed) geodata over the internet are the OGC standards WCS, WFS and WMS.

1.1.3.1 WFS

A Web Feature Service allows interacting with geodata in a geodatabase on the level of single features (vector data). It supports request for:

- metadata about the service (in XML-format)
- description of datasets (in XML-format)
- delivery of feature data (geometry and attributes in GML-format)

- manipulation of the features (edit, create, delete, lock)

1.1.3.2 WCS

A Web Coverage Service provides access to raster data. Depending on its configuration level it offers services for:

- metadata about the service (in XML-format)
- description of certain datasets (in XML-format)
- delivery of raster data (raster formats)
- complex requests
- data processing
- data manipulation

1.1.3.3 WMS

The Web Mapping Service standard allows requesting geodata by stating the extent and choice of layers or requesting attribute information for single objects from a geodata service supporting this standard (for raster and vector data). Depending on the request it returns:

- metadata about the service (in XML-format)
- a raster image with a map (in a common raster format)
- attribute information (in XML-format)

Whereas WCS and WFS are designed to deliver data for further processing, WMS is intended to provide maps for display (Fig. 1.4).

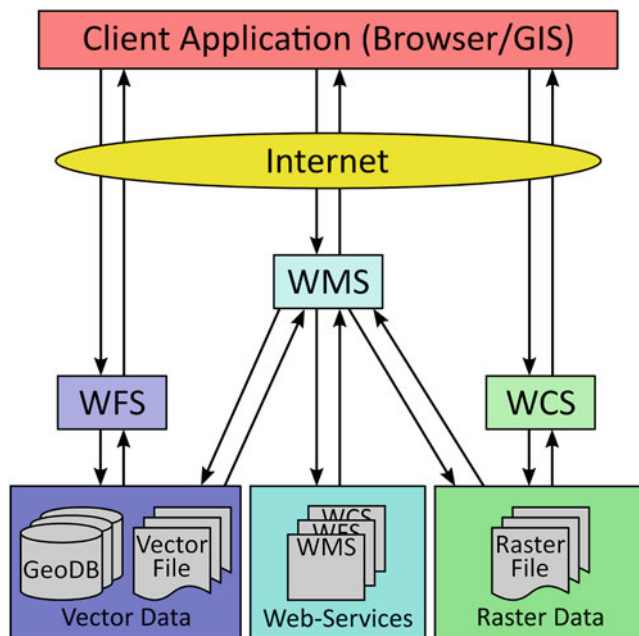
1.1.3.4 GML

The XML-based Geography Markup Language was defined by the OGC as a universal format for the storage and transfer of geodata. Besides feature (vector) data, it can also be used to represent coverages (raster) and sensor data.

1.1.3.5 WKT/WKB

The markup language Well Known Text is used to describe vector-geodata in a human-readable, easy transferable way. It is supported by many applications that comply with OGC standards. Its

Fig. 1.4 Overview scheme of OGC-web-services.
(Source: Author)



binary counterpart Well Known Binary is used to handle geospatial data within databases.

1.1.3.6 KML/KMZ

The Keyhole Markup Language is an XML-based format for the transfer of 2D and 3D geodata within internet-based applications like maps and earth browsers. KMZ files contain zip-compressed KML content. Initially developed for the use in Google Earth it became an OGC standard later on.

1.1.3.7 GPX

For the exchange of records from GPS-receivers, the GPS Exchange Format was developed by the company TopoGraphix. It represents waypoints, routes and tracks as coordinates with attributes in an open XML scheme. It can be handled by many applications.

1.1.4 Metadata (by Andreas Redecker)

Information about the characteristics of geodata and geodata services is important for the

reliability of most analyses. General descriptions about the objects held in the geodata as well as information about the spatial reference, resolution, attributes, geometric accuracy, origin, copyright and many other aspects make up the so-called metadata. Usually, it is held in a special .xml-file delivered with the data itself. International standards for the description of geographical information are defined by the ISO (International Organization for Standardization):

- ISO 19115:2003 Geographic information – Metadata
It defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. (ISO 2018a)
- ISO/TS 19139:2007 Geographic information -- Metadata -- XML schema implementation
It defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115. (ISO 2018b).

Standardised metadata are the key to the Infrastructure for Spatial Information in the European

Community (INSPIRE) that is aimed to easily share and use spatial data within the EU (European Commission 2014b).

1.2 International Data Sources (by Vít Pászto, Karel Macků, Andreas Redecker, and Nicolai Moos)

1.2.1 Eurostat

Eurostat represents the main official statistical body of the European Union with its headquarters in Luxembourg. The main task of Eurostat is to provide high-quality statistics about and for Europe (Eurostat 2018a). Thanks to these statistics, we can compare individual countries and/or various regions in a comprehensive way based on factual information. Most of the data that Eurostat collects comes from national statistical offices, which are obliged to report selected statistical indicators to Eurostat. In this sense, Eurostat serves as a common European statistical office for all member countries. For more information about Eurostat mission, goals and history, please, go to the official website – <https://ec.europa.eu/eurostat>.

1.2.1.1 Eurostat Spatial Data

The main body collecting spatial data and information within Eurostat is called Geographic Information System of the COMmission (GISCO). This unit is responsible for maintaining the geographical databases, creating and publishing maps and map applications. Besides the data management, GISCO also cooperates with other Eurostat units and publishes research texts on various topics (e.g. Rural-urban typology, Urban Europe etc.). GISCO also leads their own activities, such as GEOSTAT initiative and Merging statistics and geospatial information in the European statistical system. More details on GISCO activities and data is available at – <https://ec.europa.eu/eurostat/web/gisco>.

Talking about datasets, GISCO provides reference geodatasets (geographically covering EU) in five main themes:

- Administrative/Statistical Units – this section contains geodata about administrative hierarchical units NUTS (2003–2016), Urban audit data (2001–2014), countries-level units (2006–2016), census geodata (2011) and communes (LAU2 units, 2006–2013). Most of the datasets are provided in Esri Geodatabase format, Shapefile, and some other data models.
- Population Distribution/Demography – the section includes three main projects, namely GEOSTAT population grids (2006 and 2011), Urban Clusters (2006 and 2011), and DEGURBA (Degree of Urbanisation 2001 and 2014). Except for Urban clusters, which is provided as a raster (TIFF, geoTIFF), all datasets are provided in vector format (Esri Geodatabase, Shapefile).
- Transport Networks – this section contains two major datasets – airports (2006 and 2013) and ports (2009 and 2013). Both geodata sources are provided in Esri Geodatabase and Shapefile.
- Land cover – as indicated by the name of this section, it includes data on Land Use/Cover frame Statistical Survey (LUCAS) with the reference year 2009. Again, geodata is prepared in Esri Geodatabase and Shapefile. Besides, there are links to related datasets of Corine Land Cover (CLC) and Urban Morphological Zones (UMZ) provided by the European Environmental Agency (EEA, now being translated to Copernicus programme).
- Elevation – in this part, the main focus is on geodata referring about digital elevation model (DEM) and its derived products (e.g. slope, aspect, coloured relief). This section contains Digital Elevation Model data (in two different coordinates systems), data on Aspect, Slope, Coloured relief, Hillshade, and Hydrography. All the datasets are available in raster format (GeoTIFF).

1.2.1.2 Eurostat Statistical Data

As a counterpart to the spatial part of Eurostat data, there is a statistical part containing great

number of tabular data with the possibility to link them together with spatial data. On the homepage of Eurostat, the first option to search for a data is a tab “Data”, which redirects the user straight to available databases (<https://ec.europa.eu/eurostat/data/database>). There exist several options on how to search for a data using Data navigation tree:

- Database by theme
- Tables by theme
- Tables by EU policy
- Cross-cutting topics
- New items (sorted by code)
- Recently updated items (sorted by code)

Besides Data navigation tree, the user can perform search “database by the theme” also via context menu; this option brings additional links to respective EU policy indicators. In the context menu on the Eurostat website, it is possible to browse the database by alphabet order (Statistics A-Z).

Also, there are special data products and services available at the Eurostat webpage –

Population Census 2011, Experimental Statistics, Bulk Download, Web Services, Microdata, Metadata and Data validation service. Although this dataset provides valuable information on specific topics or using specific (technical) approaches, only the main database will be further explored.

Searching and Downloading Data from Eurostat Main Database

Using any means of data search, it will bring the user to the list of main tables or databases, in which the specific topics are listed. It is worth to note that the Eurostat database contains hundreds of tables in various topics. Therefore, it is not possible to list them all in this book. In all cases, the tables are logically organised, and it is very intuitive to download data. In Fig. 1.5, there is an example of an expanded Data navigation tree with individual tables on the Economy and finance theme (main GDP aggregates in particular).

Basic information about the selected indicator is available by clicking blue “i” icon, yellow

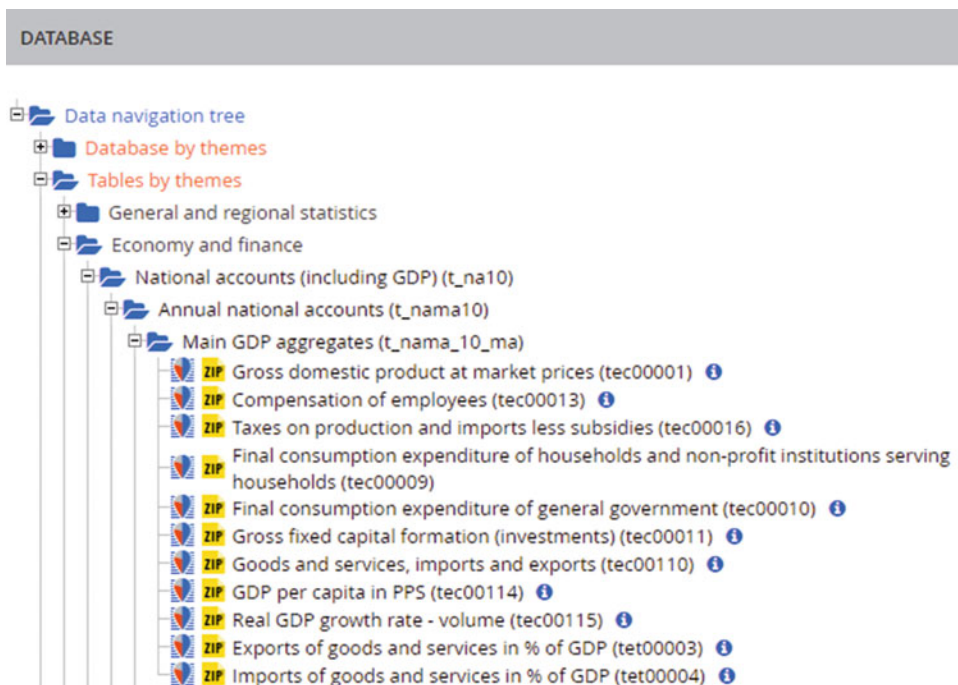


Fig. 1.5 Expanded Data decision tree on the Eurostat data website. (Source: (c) European Union 1995–2018)

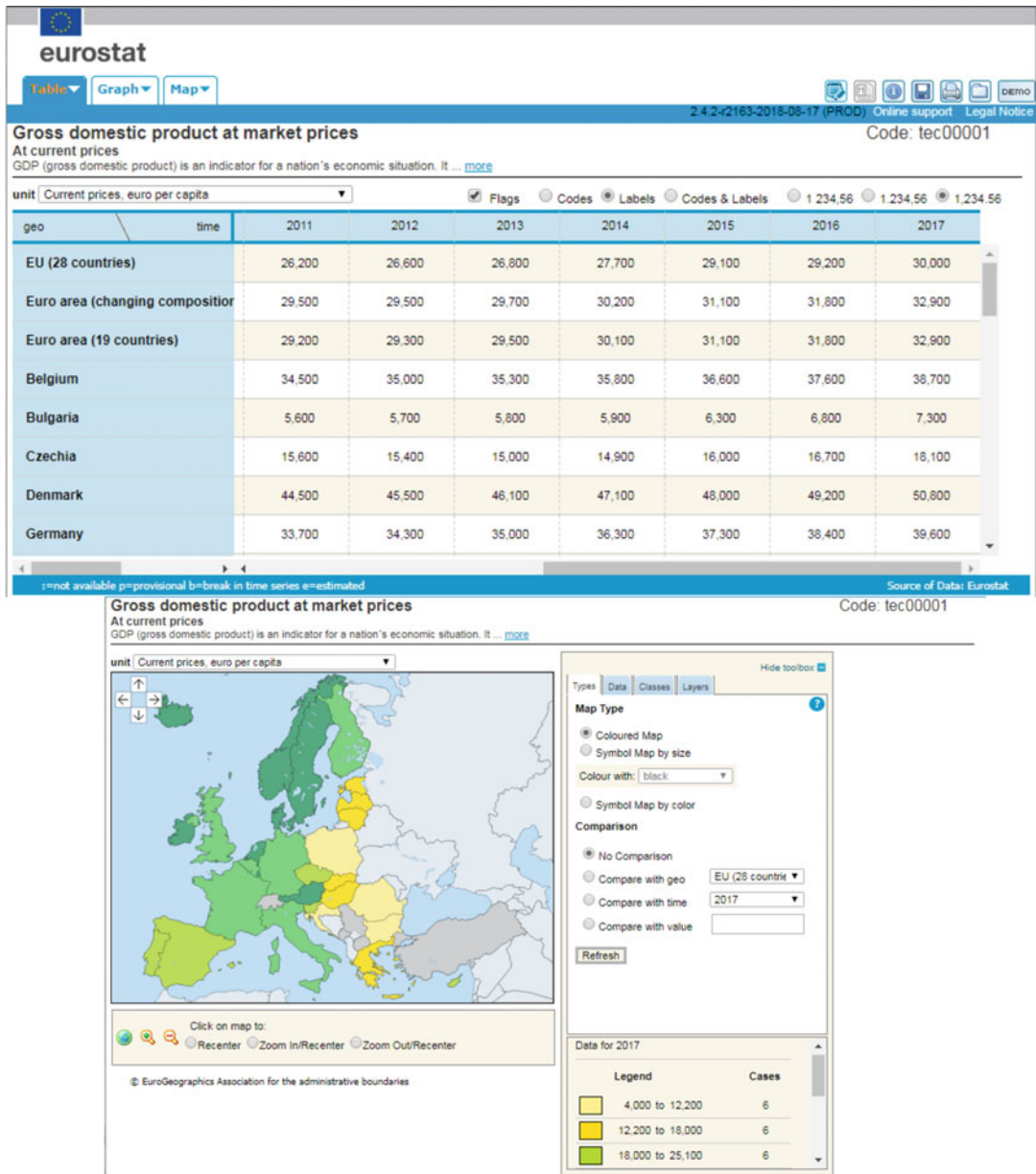


Fig. 1.6 Interactive interface for data selection, customisation and download on the Eurostat data website. (Source: (c) European Union, 1995–2018)

“zip” icons stand for downloading a data in a TSV format, and the first “marker” icon takes the user to the interactive interface with graphs, table, and map (if available). In this interactive environment (Fig. 1.6), it is possible to customise selection in tables, change the visual representation (graph or map) and download data selected

data (by clicking “floppy disk” icon). Once submitted to download, a several data formats will be available to choose – XLS with or without footnotes (with and without short descriptions), HTML (with and without short descriptions), XML, PDF (with and without short descriptions),

and TSV as a possibility to download complete table.

Besides the datasets listed above, it is important to note that Eurostat produces research publications, manuals and guidelines, working papers, yearbooks, brochures and leaflets, methodologies, books, digital publication, Statistics Explained, and other materials. To see the whole scope of these valuable sources of information, check <https://ec.europa.eu/eurostat/publications/all-publications>.

1.2.2 OECD

Organisation for Economic Co-operation and Development (OECD) is international body gathering countries from 36 countries across the globe (most of the EU countries, USA, Canada, Mexico, Chile, Turkey and others). The main goal of OECD is “to promote policies that will improve the economic and social well-being of people around the world” (OECD 2018). OECD creates a platform for international cooperation, sharing experiences and problem solving of social, economic, and environmental topics. Besides policy issues, OECD analyses and compare data with a focus on future development predictions.

To support policy and decision making, OECD runs special data portal “OECD Data” (<https://data.oecd.org>), where one can search for a data using hypertext, or browse data by county or topic. All the datasets are supported with relevant methodological guidelines and explanations. The statistical data are also available in a standalone application, in which user can search, filter, customise, visualise and download statistical data covering various themes:

- General Statistics
- Agriculture and Fisheries
- Demography and Population
- Development
- Economic Projections
- Education and Training
- Environment
- Finance

- Globalisation
- Health
- Industry and Services
- Information and Communication Technology
- International Trade and Balance of Payments
- Labour
- National Accounts
- Monthly Economic Indicators
- Prices and Purchasing Power Parities
- Public Sector, Taxation and Market Regulation
- Productivity
- Regions and Cities
- Science, Technology and Patents
- Social Protection and Well-being
- Transport

Similarly to the Eurostat database, the datasets are organised into expanding tree-system on an interactive website. After data selection, a table with data appears supported by explanations of the indicators and some other metadata (Fig. 1.7). It is possible to visualise the data as a chart (scatter plot, bar or line chart), user can customise data selection, layout and even table options (e.g. decimal places, empty rows etc.), manage and save queries, and most importantly download data. Download options depend on a selected indicator, but in general; it is possible to choose from XLS, CSV, XML, PC-axis, and others (e.g. complementary Word files). For some indicators, a bulk download as a RAR file is also available.

1.2.3 UN

As the United Nations (UN) is a well-known institution, just a brief note about its mission is to be mentioned. The UN is the international and intergovernmental organisation established in 1945, and its goal is constituted in the Charter of the United Nations, most importantly to protect human rights, freedoms, and a wide range of basic societal principles (e.g. healthcare, social equality and many others).

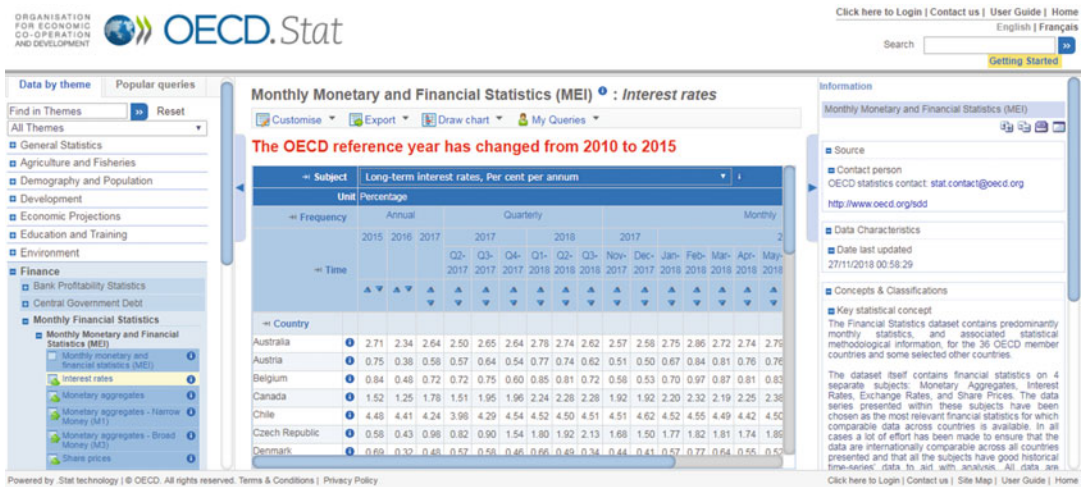


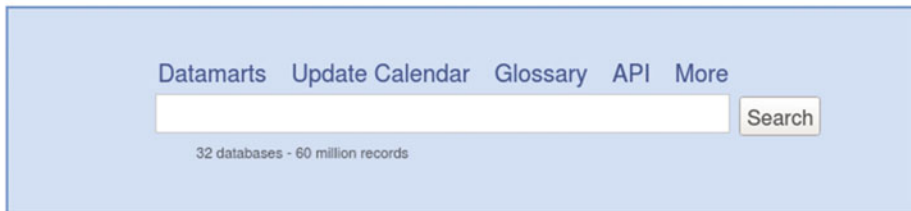
Fig. 1.7 Interface of OECD database after a selection. (Screenshot from OECD.Stat webpage)

The main body within the UN responsible for statistical data dissemination is its Statistical Division (UNSD). Statistical Division coordinates activities of international, national and other statistical organisations. Its primary focus is on data collection, processing and dissemination, methodology standardisation, and capacity development (UNSD 2018b). Thematically, UNSD covers topic such as development indicators (mainly Sustainable Development Goals – SDG), economy, environment, geospatial information, population and society. From a dataset perspective, UNSD lists the following main sources:

- UN Data
- Open SDG Data Hub
- Monthly Bulletin of Statistics Online
- SDG Indicators Global Database
- UN Comtrade
- UN ServiceTrade
- National Accounts Main Aggregates
- Joint Organisations Data Initiative
- Disability Statistics Database
- Population Census Datasets
- Population and Vital Statistics Report
- Demographic Yearbook System
- Minimum Set of Gender Indicators

The main data search tool provided by the UNSD is “UNdata” (<http://data.un.org/>), which serves as a primary search engine aggregates all the available UN data. According to UNSD (2018a), there are 32 databases with around 60 million records available. On that basis, only most general features of the UNdata portal will be mentioned. First, the user can search for data via full the text-search. The other option is to choose “Datamarts” feature, which brings the user to a data tree interface with the categorised dataset. In this environment, one can apply searching filters, select data (columns), order records, transpose rows and columns, share tables and also download data. When downloading, the user can choose between two main formats – XML and CSV. Moreover, detailed data description, including metadata, is available here.

It is also possible to look up for datasets that will be published by using the “Update Calendar” option. These options are complemented with a glossary and API (Application Programming Interface) helping users to understand indicators and/or use the data in their applications. Links to other UNSD specialised statistical databases as well as “popular statistics” are provided at the main webpage (Fig. 1.8).



Other UNSD Databases:



Fig. 1.8 UNdata main page with full text search, links to other UNSD databases and a popular search. (Screenshot from UNdata webpage, Copyright © 2018 UNSD)

1.2.4 WTO

World Trade Organisation (WTO) refers to itself as the only global international organisation dealing with the rules of trade between nations (WTO 2018). The main goal of WTO is to participate in trade negotiations and to help conclude trade contracts. WTO offers rich information sources, including the legal text of WTO agreements, economic analysis, publications, glossaries and terminology database, and statistical data. Main statistical databases are organised into thematic groups:

- Tariffs
- Trade in Services
- Global value chains
- Merchandise trade
- Non-tariff measures
- Trade in services
- Trade and tariff maps

A fundamental instrument for statistical data access is WTO Data portal (<http://data.wto.org/>). In the interface of Data portal search, it is possible to choose from more than 200 indicators, around 300 reporting economies (country profiles), about 200 products/sectors, up to 300 partner

economies, and time series with more than 70 years of history. The user can also filter the data based on topic, product classification, trade partner, and frequency. Moreover, user can exchange selector rows for columns (and vice versa) by dragging&dropping respective items and apply for changes in a resulting data table (Fig. 1.9). Once the selection is done, it is possible to download selected data and table composition as an XLS file and/or CSV. There is also an option to look into metadata with detailed information about the selected data. The user can also display the whole database inventory, where all the available indicators are listed and described.

1.2.5 World Bank

The World Bank was established in 1944 originally to offer low-interest loans for countries affected by World War II. Since then, The World Bank has grown into an organisation with 189 member countries. In general, The World Bank is a vital source of financial and

technical assistance to developing countries around the world (World Bank Group 2018). As regards data sources, The World Bank offers the World Bank Open Data (<https://data.worldbank.org/>) portal as the main proxy for various information sources. The main tool for a data search is a full-text window with two browsing options – by country and indicator. Both options take the user to a list of countries and indicators. When the searched indicator is chosen, the interactive tool will appear (Fig. 1.10) and the user can select the display method (line or bar chart, map), linked indicators, time span, check metadata, visit another data&visualisation tool (e.g. DataBank) and download the data in CSV, XML or XLS.

Besides the main search interactive tool, The World Bank Open Data portal provides links to other data resources:

- Open Data Catalog
- DataBank
- Microdata Library
- World Development Indicators

The screenshot shows the WTO Data portal interface. At the top, there are logos for WTO OMC and DATA, along with navigation links for 'Database Inventory | Technical Notes | User Guide' and language options 'English | Français | Español'. The main interface is divided into 'Selection' and 'Display' sections. The 'Selection' section on the left includes a 'Filters' area with an 'Indicator' filter and a 'Rows' area with 'Product/Sector', 'Partner Economy', 'Reporting Economy', and 'Period' filters. The 'Columns' section at the bottom of the 'Selection' panel has 'Year' selected. The 'Display' section on the right shows a 'Metadata' tab and download options for 'Excel' and 'CSV'. Below this, the 'Indicator' is set to 'Merchandise imports by product group - annual (Million US dollar)'. A data table is displayed with the following structure:

Product/Sector ↑	Reporting Economy ↑	Year --	2017	2016	2015	2014	2013
S13_AGG - TO - Total merchandise	World		18,065,140.6	16,298,573.7	16,794,444.2	19,136,180.3	19,021,196.
	Afghanistan		7,384.2	6,534.1	7,722.9	7,729.2	8,554.
	Africa		534,262.7	496,310.7	555,143.6	644,697.8	640,777.
	African, Caribbean and Pacific States (ACP)		413,348.4	383,129.9	433,923.7	502,867.8	501,337.
	Africa, CIS and Middle East		1,669,353	1,530,018.6	1,652,992.9	1,954,598.8	1,976,316.
	Albania		5,294.1	4,669.2	4,320.4	5,230	4,901.
	Algeria		46,059	47,089	51,702	58,580	55,02
	American Samoa		260	320	290	350	373.
	Andean Community (ANDEAN)		115,272.2	105,956.7	123,102.8	144,619.9	139,550.

Fig. 1.9 WTO interactive selection tool. (Screenshot from WTO Data portal)

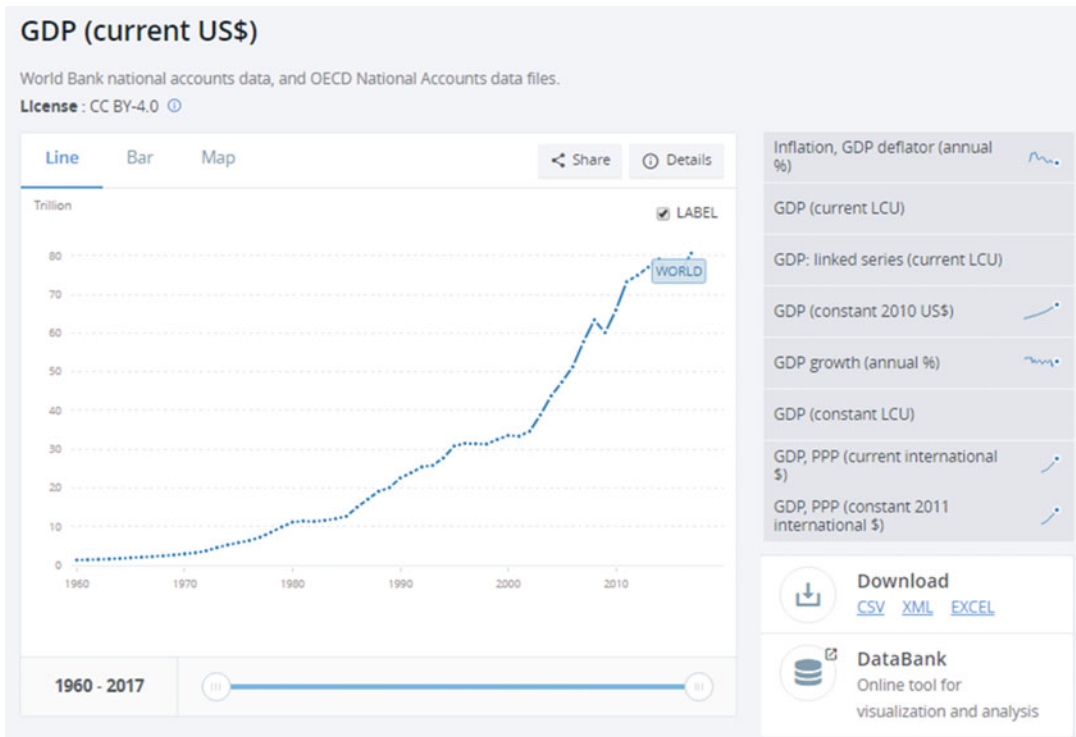


Fig. 1.10 World Bank interactive tool for data exploration. (Screenshot from The World Bank Open Data portal)

- Open Finances
- Projects & Operations
- Open Data Toolkit
- AidFlows
- Global Consumption Database

1.2.6 GADM

GADM is a database of global administrative areas available at the link www.gadm.org. It provides spatial administrative data and maps for all countries of the world. The spatial data can be download by country or for the entire world. There is not the same administrative detail for all states; for example, there are three levels available in the Czech Republic, three levels in Slovenia as well, and five levels in Germany. Several data formats are offered: shapefile, geopackages, KMZ and .rds (file for R software).

The coordinate system of downloaded data is WGS 84. Regarding the attribute data, only basic information is provided – the name of the administrative, unit, a code and type (state, region, district, municipality), both in English and local language. Unfortunately, some of the data is missing.

The second part of the GADM project is thematic maps. For almost every country, a set of maps is available. Main topics are average annual temperature, total annual precipitation, elevation and map of night light activity. Unfortunately, maps can be downloaded only with low resolution. To reach a better detail, a map for one of the sub-division unit can be generated but downloaded still just in the low resolution.

This dataset could be a great source of administrative boundaries for countries with difficulty available data (e.g. African countries). Attention should be paid on classification level – since data

is not provided by any government organisation, the user should always check if the classification follows the official administrative system of the country. The data are freely available for academic use and other non-commercial use. Redistribution or commercial use is not allowed without prior permission (GADM 2018).

1.2.7 Esri Open Data

Esri, as one of the most important GIS company worldwide, offers the collection of Data&Maps. This collection includes over 120 pre-symbolized vector data layers for North America, Europe, and the world. Datasets include several topographical data, demographic data, and transportation data. Access to the data is provided by the Esri Data & Maps Group on ArcGIS Online. Data can be downloaded in several GIS formats and can also be connected directly to Esri software products like ArcGIS or ArcGIS Online.

Second important od Esri data is their ArcGIS Open Data portal available at <http://opendata.arcgis.com>. This portal aggregates over 1000,000 datasets from over 5000 organisations worldwide. The idea of this portal is to offer the space and tools to share any spatial data as the open data. Data can be easily searched, visualised and downloaded in several GIS formats like KML, shh or GeoJSON, and can be accessed via several API (e.g. ArcGIS REST). Data covers many topics ranging from hydrology to criminality, depending on the users that published their datasets there.

1.2.8 OpenStreetMap

If the project needs free basic vector data for a certain area, the OpenStreetMap (OSM) project might be the first address to visit. OpenStreetMap was founded in 2004 as a free editable map of the world, inspired by the concept of Wikipedia where everybody who has something to contribute can participate and feed the OSM-databases from all over the world. To use these databases, one simply has to visit openstreetmap.org and browse through the maps in the interactive web

map. If the files should rather be opened in a GIS, e.g. to do some calculations, there can also always be defined a freely selectable subset on the web map to download then and import it into any standard GIS. The downloaded dataset contains all features that OSM provides, as there are points of interest, rivers, streets, outlines of buildings etc.

If the project area is not yet clearly defined or the project area requires complete datasets of whole states, countries or continents, then it's worth to take a look on geofabrik.de where there are direct download links that contain the same features as listed above for the chosen administrative area.

The whole project until now counts more than two million registered users while numbers are growing, which is one of the reasons why OSM-data is not the most trustworthy kind of data one can get. As the number of participants is steadily increasing, so does the number of people who may incorporate wrong datasets into the OSM-database – no matter if by accident or on purpose – what leads to a not directly recognisable inaccuracy in few areas. These inaccuracies exist as long as somebody detects and fixes them. So, if the project requirements demand a completely credible dataset and not just something that helps to get an overview, feed some background map or do some basic analysis in teaching classes, one has to take into account that OSM-data and its crowd-based digital modelling of the world's surface cannot fully replace the national datasets provided by governments and official releases that are mostly more reliable and trustworthy.

1.2.9 Urban Atlas

Urban Atlas is a service in the frame of the EU Copernicus program, the world's largest single earth observation program, and provides pan-European reliable, inter-comparable and high-resolution land use and land cover data for functional urban areas (FUA) and their surroundings. In the first reference year 2006 Urban Atlas included 319 FUAs with more than 100.000

inhabitants (as defined by the Urban Audit) classified into 20 different classes (e.g. urban fabric, agricultural, industrial/commercial, green urban areas, etc.). Since the second reference year 2012 Urban Atlas comprises 800 FUAs in sum, as the surroundings of the FUAs with more than 50.000 inhabitants were added to the database as well as various new classes for selected FUAs, like a Street Tree Layer (STL), the building height of core urban areas in European capitals or wetlands.

The classification is conducted by using a combination of statistical image classification and visual interpretation of Very High Resolution (VHR) satellite imagery. Finally, the Urban Atlas product is enriched with functional information (road network, services, utilities, etc.), using additional data sources such as local city maps or online map services. The access to the Urban Atlas database can be reached via land.copernicus.eu/local/urban-atlas. After creating a free account, all datasets of the demanded city/area are available for download.

1.3 National Data Sources

This section focuses on the three countries, from which the Spationomy project partners were drafted. To keep the logic of the previous section, both main statistical and geospatial bodies and their data sources will be mentioned. As regards the statistical offices, the situation in Czechia and Slovenia is rather simple – both countries have one official statistical institution – while in Germany, every single federal state runs its own statistical office. It is worth to mention that the standardisation level of indicators is strictly followed, so the datasets should be mutually comparable. Nevertheless, there exists an office on a national level in Germany that collect selected statistical indicators. The latter office will be a subject in this chapter. It is worth to mention that each EU member state is obliged to provide statistical data within the European Statistical System (ESS) via their national statistical offices. Complete list of official statistical bodies of EU member states is given in Table 1.2.

When talking about geodata sources in the three countries, the status quo is much more diverse. In each country, there are several institutions dealing with some geodata; therefore, only the main geoportals collecting the most important geodatasets will be described in this part.

1.3.1 Czechia

1.3.1.1 Czech Statistical Office

Czech statistical office (<https://www.czso.cz>) is the central authority for providing statistics in Czechia. It is also the main body to report statistics to Eurostat. Every product from the office is based on statistical data. Therefore, only the main data source – Public database – will be here described. The public database is an interactive search engine for most of the statistical data that the Czech statistical office produces. Within the Public database, there are three options on how to obtain data:

- Statistics – this options is indicator-based, i.e. the user can choose the individual indicator via data tree interface. Then, it is possible to sort the data, filter it, display in table, chart or map, or download.
- All about territory – this choice let the user select a territory or region and display is summarising statistic (territory profile). As in the previous case, the data can be sorted, modified, displayed in table, chart or map, printed in a template or downloaded.
- Customised selection – this represents the most advanced method of data selection. From the user perspective, it is necessary to know which indicator, territory and period is desired since this selected is stepwise (Fig. 1.11). Once the selection is ready, the user can modify a layout, download selected data, or go back and change the selection settings.

Regardless of the method, it is possible to select indicators from these main thematic groups:

Table 1.2 European Union member states statistical offices

Country	Statistical office	Webpage
Austria	Statistik Austria	https://www.statistik.at
Belgium	National Statistical Institute (STATBEL)	https://statbel.fgov.be/en
Bulgaria	National Statistical Institute of Bulgaria	www.nsi.bg/en
Croatia	Central Bureau of Statistics	https://www.dzs.hr
Cyprus	Statistical Service of the Republic of Cyprus (CYSTAT)	http://www.mof.gov.cy/mof/cystat/statistics.nsf/index_en/index_en?OpenDocument
Czech Republic	Czech Statistical Office	https://www.czso.cz/csu/czso/home
Denmark	Statistics Denmark, Statistics Greenland	https://www.dst.dk/en , http://www.stat.gl
Estonia	Statistics Estonia	https://www.stat.ee/en
Finland	Statistics Finland	https://www.stat.fi/index_en.html
France	National Institute of Statistics and Economic Studies (INSEE)	https://www.insee.fr/en
Germany	Federal Statistical Office (DESTATIS)	https://www.destatis.de/EN
Greece	Hellenic Statistical Authority (ELSTAT)	http://www.statistics.gr/en/home
Hungary	Hungarian Central Statistical Office	www.ksh.hu/?lang=en
Ireland	Central Statistics Office	https://www.cso.ie/en
Italy	The Italian National Institute of Statistics (ISTAT)	https://www.istat.it/en
Latvia	Central Statistical Bureau of Latvia	https://www.csb.gov.lv/en
Lithuania	Statistics Lithuania	https://www.stat.gov.lt/en
Luxembourg	National Institute of Statistics and Economic Studies (STATEC)	https://statistiques.public.lu/en/index.html
Malta	National Statistical Office	https://nso.gov.mt
Netherlands	Statistics Netherlands	https://www.cbs.nl/en-gb
Poland	Statistics Poland	http://stat.gov.pl/en
Portugal	Statistics Portugal (INE)	https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE&xlang=en
Romania	National Institute of Statistics	http://www.insse.ro/cms/en
Slovakia	Statistical Office of the Slovak Republic	https://slovak.statistics.sk
Slovenia	Statistical Office of the Republic of Slovenia (SURSTAT)	https://www.stat.si/statweb/en/home
Spain	National Statistical Institute (INE)	https://www.ine.es/en/welcome.shtml
Sweden	Statistics Sweden	https://www.scb.se/en
United Kingdom	Office for National Statistics (ONS)	https://www.ons.gov.uk/

Source: <http://www.unecce.org/stats/links.html>, authors survey

- Dwellings
- Prices, Inflation
- Tourism
- Foreigners
- Transportation, Inf. and Communication
- Energy
- Financial data, public budgets
- GDP, National Accounts
- Information Technologies
- Business Cycle Surveys
- Crime, Accidents, Fires
- Culture, sport
- Forestry
- Labour Costs and Earnings
- Retail Trade, Hotels and Restaurants
- Population
- Business Register Data
- Industry
- Living conditions, Household Income and Expenditure
- Population Census
- Services

Customize your own table

1. Indicator 2. Territory 3. Time period 4. Layout 5. Show data

Indicators, Time periods and Territories can be dragged to highlighted positions.

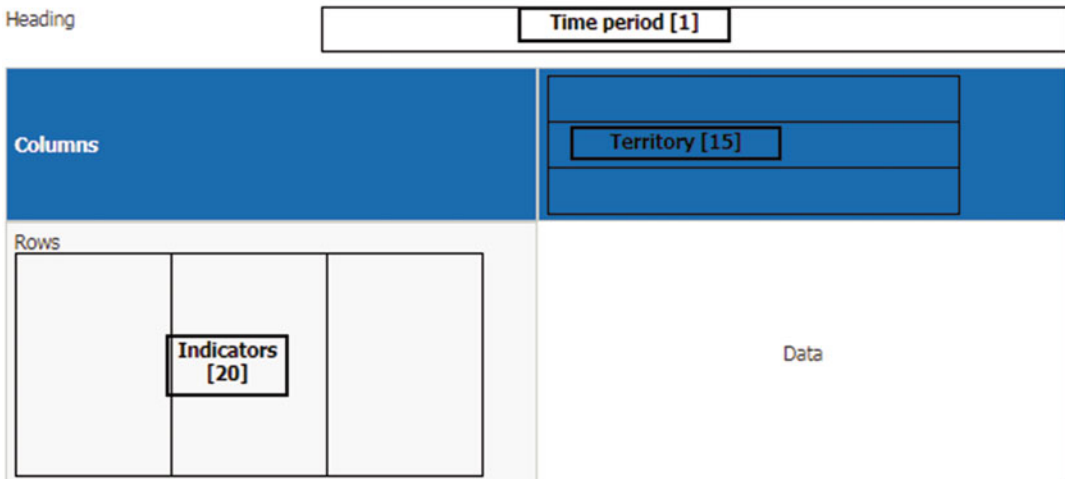


Fig. 1.11 Selection process via Customised selection option in the Public database. (Screenshot from Public database, Czech Statistical Office)

- Social Security
- Construction
- Territory, residential structure
- Science, Technology and Innovation
- Elections
- Education
- External Trade
- Employment, Unemployment
- Health Care, Incapacity for Work
- Agriculture
- Environment

Most of the indicators could be downloaded as XLS, PDF, XML and PNG (for maps) format. All datasets are complemented with metadata, and methodological guidelines are also available at the Czech Statistical Office.

1.3.1.2 Czech Office for Surveying, Mapping and Cadastre

Czech Office for Surveying, Mapping and Cadastre (ČÚZK) is the main state institution responsible for the production of spatial data. The main tasks of this office are, e.g. to complete administration of Czech cadastre, mapping of Czech

Republic in all scales, the creation of Fundamental Base of Geographic Data (ZABAGED), implementation of geodetic surveys or standardisation of geographic names.

ČÚZK offers access to all map and data products by Geoportal. It is a web interface to access the spatial data produced and updated by activities of the Czech Office for Surveying, Mapping and Cadastre (ČÚZK 2018a). The Geoportal is available at the web page www.geoportal.cuzk.cz. The Geoportal offers services of data sharing according to rules of the EU INSPIRE Directive. It allows to search for spatial data and other products, to access services based on the spatial data and to obtain the products via e-shop. Most of them are charged according to the amount of data user request. An overview of all products is also available on the web page, few of them is described on the following lines.

Orthophoto of Czechia

It is a periodically updated dataset of aerial images covering the whole republic. An orthophoto is a geo-referenced ortho-photographic display of the Earth surface. Orthophotos

show the photographic image of the Earth surface transformed in the way that image shifts generated during the acquisition of aerial images are removed. Since 2010 the photography has been carried out by a digital camera, which has caused an additional increase of product quality up to the spatial resolution of 0.2 m per pixel. This aerial images can serve as a suitable base map for use by for planning, project preparation, environmental protection, risk management and other applications done by organisations, state institutions and local governments (ČÚZK 2018b).

ZABAGED

The ČÚZK (ČÚZK 2018d) describes ZABAGED dataset as following: The geographic base data of the Czech Republic (ZABAGED®) is a digital vector model of the territory of the Czech Republic. ZABAGED® is a part of the surveying information system and belongs to information systems of the public service. It is maintained as a seamless database for the entire territory of the CR in a centralised information system managed by the Land Survey Office. Planimetric section of ZABAGED® contains two dimensional (2D) spatial information and descriptive information on settlements, roads, utility networks and pipelines, hydrology, administrative units and protected areas, vegetation and surface, terrain relief.

Both the orthophoto and ZABAGED database can be accessed as a Web Map Service (WMS). This can be easily added to any GIS software and then used for free. In total, ČÚZK offers almost 30 topics available as a free WMS, which is of course only for viewing, but sometimes this preview can sufficient as a base map for the project. A list of all services is available at the website in category Network services.

Registry of Territorial Identification, Addresses and Real Estates

Registry of Territorial Identification, Addresses and Real Estates (RÚIAN) is under operation since July 1st 2012 as an integral part of the whole system of public administration basic

registries. The administrator and operator of RÚIAN are ČÚZK. The main benefit of the entire set of basic registries is to create such a set of reference data, which is obligatory for the performance of public administration agendas. In this case, it means the administration of descriptive and localisation data about territorial elements, territorial inventory units, teleological territorial elements and address data and their mutual relations (ČÚZK 2018c).

A part of the RÚIAN project IS the public remote access, through which RÚIAN and data are freely available via the internet for viewing or downloading in RUIAN exchangeable format (VFR – derived from GML format). Free remote access is available publicly at <http://vdp.cuzk.cz/>, unfortunately only in Czech. There several features categories can be downloaded – administrative units and boundaries (regions, districts, municipalities, etc.) and detailed spatial information at the municipality level – parcels, address points, streets and buildings. This detailed information is beneficial for different economic applications, local government management and planning or development.

1.3.2 Slovenia

1.3.2.1 Statistical Office of the Republic of Slovenia

Statistical Office of the Republic of Slovenia (SURS) represents the main institution in Slovenia responsible for collecting, managing, and distribution of statistical data about the country. According to SURS (2018), SURS “is professionally independent government service with autonomy as regards professional and methodological issues. The mission of the Slovene statistical office is to provide to users statistical data on the status and trends in the economic, demographic and social fields, as well as in the field of environment and natural resources.”

As for a statistical data sources, SURS website (<https://www.stat.si/StatWeb/en>) offers four main option to access a statistical data – via dynamic search tool (although sensitive only for Slovenian

indicator names), A to Z browsing, main database (SI-STAT), and preset themes:

- Agriculture, Forestry and Fishery
- Construction
- Culture
- Development and Technology
- Earnings and Labour Cost
- Education
- Energy
- Enterprises
- Environment
- Foreign Economic Relations
- GDP and National Accounts
- Industry
- Labour Market
- Population
- Prices and Inflation
- Quality of Life
- Regional Overview
- Social Protection
- Territory
- Tourism
- Trade and Services
- Transport
- European data

All the indicators in the listed themes are available in the main database (SI-STAT), which offer broader functionality for searching, selecting, filtering, displaying and downloading a data. In the database system, data is organised in four main categories – Fields of statistics (e.g. demography, economy and others), Census data, Cross-sectional reviews, and Archive for discontinued tables. By choosing a specific topic within the category, a list of individual indicators appear, and the user can then select a particular settings of selected indicator (e.g. when choosing Gross Domestic Product in an Economy section, several variations of Gross Domestic Product are available; including a selection of a year and respective data description). A data download is available in PC-axis format, XLS, TXT, CSV, and as an HTML.

1.3.2.2 The Surveying and Mapping Authority of the Republic of Slovenia

The Surveying and Mapping Authority of the Republic of Slovenia comprises the Main office, the Real estate office, the Mass real estate valuation office, the Geodesy office and twelve regional surveying and mapping administrations. These have been set up for the reasons of effective operation and the accessibility of services implemented by the Surveying and Mapping Authority of the Republic of Slovenia (Surveying and Mapping Authority of the Republic of Slovenia 2018a).

The offices cooperate with the regional surveying and mapping administrations to implement the following tasks to:

- prepare the national land survey service annual program
- organise the work of the regional surveying and mapping authorities
- implement the development assignments about surveying and mapping activities;
- provide the implementation of international obligations of national land survey service.

There is the e-Surveying data portal on the following link: <http://egp.gu.gov.si/>. After a quick registration, a user can access the web portal, where all available themes (17) are listed. Themes are, for example, remote sensing data, basic topographic maps, digital elevation model, register of geographical names or land cadastre. All the data can be very easily downloaded. Several interesting layers supporting the synergy of an economy and spatial data are on offer here — for example, data from Public Infrastructure Cadastre. This is a centralised database of public infrastructure objects and networks (roads, railways, water supply, sewage network, etc.). Each element in the database has the information about its type, location, identification number and ownership. The infrastructure network owners or managers are obliged to provide up-to-date information.

As well as in the Czech Republic, also Slovenian most detailed cadastral data can be freely downloaded on this portal. The following information is kept in the Land cadastre: the parcel identification code, border, surface, owner, land under the building, land evaluation. The relation to the Register of Spatial Units, Building Cadastre and Land Registry is also provided. Information on ownership of physical persons is not available to the public (personal data protection rules). Personal data about ownership can be provided by the Data Issuing Department of the Surveying and Mapping Authority of the Republic of Slovenia only when the end-user has a special right to use this personal data defined in law (Surveying and Mapping Authority of the Republic of Slovenia 2018b). Attribute and spatial data can be downloaded separately; the elementary unit for download is a municipality.

1.3.3 Germany

1.3.3.1 The Federal Statistical Office

The Federal Statistical Office (DESTATIS) is responsible for providing and disseminating statistical information and based on the federal structure and administration in Germany, DESTATIS implements federal statistical surveys in cooperation with the statistical offices of the 16 federal states (DESTATIS 2018). This implies the importance of DESTATIS since it acts as the main coordinator, ensuring that the data are collected by federal states according to standards, methodology and is delivered in time.

Besides actual news and information on DESTATIS homepage (<https://www.destatis.de>), it is possible to search for a specific data by looking at Facts & Figures tab on the main webpage, but more importantly, DESTATIS runs database application “GENESIS”. Apart from the dynamic full-text search, it is possible to browse statistics by theme, which are grouped into nine main themes:

- Territory, population, labour market, elections
- Education, social security. Benefits, health, justice

- Housing, environment
- Economic sectors
- Foreign trade, enterprises, crafts
- Prices, earnings, income, consumption expenditure
- Public finances, taxes, public service personnel
- Economic accounts
- National and international indicator systems

Each of the themes contains several sub-themes in which individual indicators are available. Similar to other databases, the tree structure for data search is employed in the database interface. It is necessary to go through the tree structure down to the level with an individual table with an indicator (usually fifth level). In some cases, the tables are further split into the lower level of the hierarchy, for example:

4 Economic sectors – 47 Financial and other services – 473 Insurance – 47311 Statistics of insurance companies, pension funds – 47311-0001 Insurance companies’ key figures: Germany, years, economic activities

Once the table with the desired table is selected, it is possible to generate results (as a table or chart) with respective indicators in a table (Fig. 1.12). Download options are XLS, XLSX, CSV, and HTML.

1.3.3.2 Federal Agency for Cartography and Geodesy (BKG)

“The Federal Agency for Cartography and Geodesy is the central service provider of topographic data, cartography, and geodetic reference systems for the German government.” (BKG 2019a).

Its main tasks are to ensure a uniform coordinate system for the entire territory of the Federal Republic of Germany and to provide up-to-date spatial data of Germany via the internet. For this BKG integrates the official spatial data records of BKG and all sixteen federal states (Laender), as well as those of third-party suppliers. Their data is first edited and standardised by BKG before being made available in digital form.

Result - 47311-0001

Table

Options:

Insurance companies' key figures: Germany, years, economic activities

Statistics of insurance companies, pension funds
Germany

Year WZ2008 (2-5 digit codes): Insurances		Enterprises	Gross premiums recorded	Production value	Gross value added at factor cost	Gross operating surplus
		number	Mill. EUR	Mill. EUR	Mill. EUR	Mill. EUR
2016						
WZ08-65	Insurance, reinsurance, pension funding	1,300	268,829	99,662	23,555	12,569
WZ08-651	Insurance	1,087	199,729	65,528	13,704	4,287
WZ08-6511	Life insurance	584	86,025	17,573	1,294	-787
WZ08-6512	Non-life insurance	503	113,704	47,955	12,410	5,074
WZ08-65121	Health insurance	63	37,181	∆	∆	∆
WZ08-65122	Indemnity and accident insurance	440	76,552	∆	∆	∆
WZ08-652	Reinsurance	31	59,524	32,223	8,467	7,015
WZ08-653	Pension funding	182	9,576	1,911	1,384	1,267
WZ08-653-01	Retirement funds	153	6,880	∆	∆	∆
WZ08-653-02	Pension funds	29	2,695	∆	∆	∆

Result - 47311-0001

Diagram

Options:

Insurance companies' key figures: Germany, years, economic activities
Statistics of insurance companies, pension funds
Germany

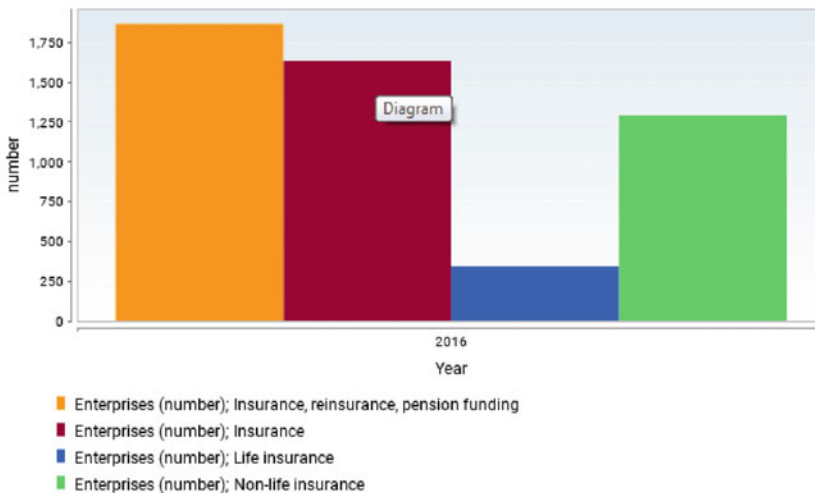


Fig. 1.12 Resulting table and chart obtained in GENESIS database. (Screenshot from GENESIS database, © Federal Statistical Office)

Furthermore, the authority supports the establishment and expansion of spatial data infrastructure, which in turn enables all citizens to search

for and take advantage of the spatial data offered by the federal government. BKG represents Germany's interests in international collaborative

entities and projects addressing the fields of geodesy and geoinformation. It also advises its customers and offers customer-oriented solutions. (BKG 2019b).

The BKG operates the “Service Centre of the Federal Government for Geo-Information and Geodesy” on the web (<http://www.geodatenzentrum.de>). Besides News and some descriptive content, it provides access to Web Applications, Online Shops and Open Data.

Under the category “Web Applications” the service centre provides web-map-applications and JAVA-applets as clients for the access to and the use of geodata provided by the BKG. The service “Maps of BKG” (“Karten des BKG”) allows an overview and browsing of geodatasets maintained by the BKG. The menu item “TopoPlusOpen Download” leads to a Java-Application that allows downloading tiles of the world-wide TopoPlusOpen-map.

The category “Online Shops” gives access to three specialised online shops. These allow ordering geodata, access geodata services, or to buy printed maps that are not free of charge and therefore not available for download.

In the section “Open Data” all datasets are available for download or are provided as WMS or WFS. They can be used free of charge according to different licenses specified in the metadata of the datasets. On the page “Free Data and Services of BKG” the following Products are on offer (Table 1.3).

The page “INSPIRE Themes” gives an overview of INSPIRE conformal services within the common spatial data infrastructure in Europe that are available free of charge. For each dataset, a description of its contents, downloads of PDFs holding the INSPIRE Data Specification and detailed documentation as well as the WMS- and WFS-URLs are provided. Datasets for the following INSPIRE themes are available:

- INSPIRE Hydro:
 - Physical Waters
 - Network
- INSPIRE Transport Networks:
 - Road Transport Network

- Rail Transport Network
- Water Transport Network
- Air Transport Network
- INSPIRE Administrative Units
- INSPIRE Protected Sites
- INSPIRE Geographical Names
- INSPIRE Land Cover

In addition to the “Service Centre” described above, the BKG operates the web-portal <https://www.geoportal.de> for the Spatial Data Infrastructure Germany (GDI-DE). The GDI-DE is an initiative of the German federal government, the states and its municipalities. It constitutes the German part of the European spatial data infrastructure implemented via the EU Directive INSPIRE (GDI-DE 2019).

Besides comprehensive information on the GDI-DE the portal guides the way to geodata related resources of many different entities within the German federal and decentralised administration (Geoportal → Service → Viewer und Portale). Direct links to the GDI-pages of the states are provided via a member list on the sub-homepage “GDI-DE”.

1.4 Other Statistical Data Sources

This section highlights microdata sources covering most of the European countries. According to Eurostat (2018b), microdata is records containing information on individual persons, households or business entities. In many cases, due to the individual nature of such datasets, microdata is not publicly accessed to protect personal or other sensitive information about the entity. Moreover, microdata is usually collected as a sample of a given population, therefore are in a very specific topic, demographic or business sample, and not representing the whole population (e.g. entire business sector, or country/region). Similarly to microdata, commercial datasets possess the same characteristics as regards free access. Based on the business nature of commercial datasets, these are usually provided upon a purchase, which somehow limits their wide-range usage

Table 1.3 Open data available at the BKG

Product	Metadata	Download format	Service
TopPlusOpen	PDF	Web application for downloading of extracts and tiles of the web map tile archive	WMS/ WMTS
Digital landscape model 1:250000	PDF/ TXT	Shape-file	WFS
Digital landscape model 1:1000000	PDF/ TXT	Shape-file	WMS/ WFS
CORINE land cover	PDF/ TXT	Shape-file	WMS/ WFS
Digital terrain model grid width 200 m	PDF/ TXT	GRID-ASCII XYZ-ASCII	WMS
Digital terrain model grid width 1000 m	PDF/ TXT	GRID-ASCII XYZ-ASCII	WMS
Digital topographic map 1:250000	PDF/ TXT	TIFF	WMS
Digital topographic map 1:500000	PDF/ TXT	TIFF	WMS
Digital topographic map 1:1000000	PDF/ TXT	TIFF	WMS
Administrative map of Germany (Federal States, administrative regions, districts)	PDF/ TXT	PDF	
Administrative map of Germany (Federal States)	PDF/ TXT	PDF	
Administrative map of Germany (national border, Federal States, districts)	PDF/ TXT	AI, EPS, SVG	
Administrative areas and NUTS regions at different scales, with and without population numbers	PDF/ TXT	Shape-file	WMS/ WFS
Geographical names 1:250000	PDF/ TXT	Shape-file, CSV	WMS/ WFS
Geographical grids for Germany in UTM- and Lambert-projection	PDF/ TXT	Shape-file, CSV	
Landscapes of Germany	PDF/ TXT		WMS
Archives	PDF/ TXT	Raster & vector files	

(especially by the scientific community often with the constrained budget). However, it is necessary to mention some of the important data sources that are classified as microdata or commercial data.

1.4.1 Eurostat Microdata

1.4.1.1 Community Innovation Survey

According to Eurostat (2018c), Community Innovation Survey (CIS) is a survey of innovation

activity in enterprises as part of the EU science and technology statistics voluntarily (i.e. different countries contribute to the individual survey years). CIS uses harmonised questionnaire for all EU member states and as such presents unique and reliable source of data regarding innovation activities of enterprises of different size, age, and industry (Vaculík et al. 2017). As noted by Vaculík et al. (2017), the advantage of the CIS is the long-term experience with methodological issues related to the innovation activities involving data on technical types of innovation (product

and process) as well as on the long underestimated findings on non-technical innovations (marketing and organizational). The datasets are available for research purposes only upon request. First, the research organisation has to be recognised as a research entity; then it is possible to apply for data itself based on a research proposal. More information about the dataset could be found at <https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey>.

1.4.1.2 Eurostat Microdata – Other Sources

On the main Eurostat webpage about Microdata access (Eurostat 2018b), following microdata surveys and dataset are listed:

- European Community Household Panel
- European Union Labour Force Survey
- European Union Statistics on Income and Living Conditions
- Structure of Earnings Survey
- Adult Education Survey
- European Road Freight Transport Survey
- European Health Interview Survey
- Continuing Vocational Training Survey
- Community Statistics on information Society
- Micro-Moments Dataset
- Household Budget Survey
- Labour Force Survey
- Statistics on Income and Living Conditions

The latter two surveys are samples free for public in order to allow general public and researchers to become familiar with such microdata and to prepare software programmes (e.g. statistical computing tools) for the possible use of “full” microdata.

1.4.2 Global Entrepreneurship Monitor

Global Entrepreneurship Monitor (GEM) represents an international project on entrepreneurship data collection and research. The project monitors two aspects of entrepreneurship

– individual behaviour and attitudes to entrepreneurship, and general conditions and the entrepreneurial context in each participating country. The former aspect is monitored by Adult Population Survey (APS), which looks at the characteristics, motivations and ambitions of individuals starting businesses, as well as social attitudes towards entrepreneurship (Global Entrepreneurship Monitor 2018). The latter aims at the national context in which individuals start businesses (Global Entrepreneurship Monitor 2018), and is based on experts reports within National Expert Survey (NES). For both datasets, a national team responsible for surveys needs to be formed, mostly voluntarily. Therefore, freely available datasets vary from country to country and from time to time. Detailed information about the project can be found at www.gemconsortium.org.

1.4.3 Amadeus – Bureau Van Dijk

Amadeus database provides information about companies across Europe and is maintained by a Moody’s analytics company Bureau van Dijk (BvD). Amadeus database is one of the BvD’s international databases with comprehensive information on private companies. The database includes basic information about the company (e.g. postal address, IDs, number of employees, industry category), company financials and their indicators, detailed corporate structures, and many more. All data are collected yearly, and according to BvD’s webpage about the product (Bureau van Dijk 2018), the Amadeus database contains information on about 21 million companies across Europe. It has to be noted that besides the Amadeus database (and global database), BvD also maintains special database about Asia-Pacific region, insurance companies, intellectual properties, as part of international datasets. National datasets covering specific countries and specialist products are also provided by the company. Although very information-rich, databases offered by BvD are commercial and need to be purchased.

1.5 Earth Observation Data (by Carsten Jürgens)

Earth observation data, also called remote sensing data, are characterised to be pictorial representations (images) in raster format of the earth's surface acquired by sensors on board of associated platforms. Earth observation systems are characterised by up-to-date image data capture that is useful for various applications.

The principle of remote sensing relies on the electromagnetic energy, which is the transmitter of any information between the earth's surface objects and the image generating sensor on board of a platform (Jürgens 2003). Passive sensors are characterised by their dependency on emitted or reflected electromagnetic energy. Since the sun illuminates the earth during the daytime, passive sensors can capture the reflected portion of the incoming radiation. Therefore these sensors rely on the sun's illumination during daytime and are affected by cloud cover, which obscures the earth's surface (Campbell and Wynne 2011). In contrary, active systems have their own source of illumination and allow to capture images also during night time (Albertz 2013). Active sensors are LiDAR/Laser Scanning and RADAR, the latter case is also relatively unaffected by clouds. Due to the long microwave wavelengths used, clouds can be penetrated, and one can get image data of the earth's surface without any missing area due to cloud coverage.

1.5.1 Platforms

There is a variety of different imaging sensors that need to be mounted on-board of a flying platform. Platforms are subdivided into earth-orbiting satellites, airplanes, helicopters and unmanned aerial vehicles (UAV). Airplanes, helicopters and UAV's are able to acquire image data on demand and therefore are very flexible. For airborne systems, a flight plan has to be prepared to assure that the system covers the complete area of interest during the planned flight. Often aerial

systems capture stereo images which can be used for 3D-interpretation. Therefore an overlap of at least ca. 50% is needed between adjacent images and around ca. 15–35% between flight lines. Aerial systems can be started almost at any time and adjusted to special needs by equipping them with specific sensors according to those needs. The flying height can be adapted to the mapping scale requested for an image campaign. Earth observation satellites instead move on fixed near-polar sun-synchronous orbits with fixed revisit rates. Since the sensor on a satellite platform cannot be changed like on an airborne platform, the user has to decide which earth observation satellite system is most applicable for a specific task. This means in contrary to flexible platforms like airplane, helicopter and UAV's, which can carry different sensors, in the case of satellites, one has to decide which system serves the user needs best according to its specific sensor system characteristics.

1.5.2 Sensor Types

As indicated earlier, there are active and passive sensors. The active sensors are characterised by their ability to acquire images during day and night and in the case of RADAR their robustness against clouds and unfavourable weather conditions due to their long microwave wavelengths (1 mm–1 m) that can penetrate clouds. The passive sensors operate in the so-called optical domain of the electromagnetic spectrum with much shorter wavelengths, ranging from the visible part of the electromagnetic spectrum (blue, green, red) to the short, middle, thermal infrared and passive microwave wavelengths (see Fig. 1.13).

Some specialised sensors are able to capture thermal emissions and passive microwave emissions during the night as well.

1.5.3 Types of Resolution

Earth observation sensor systems are characterised by different types of resolution.

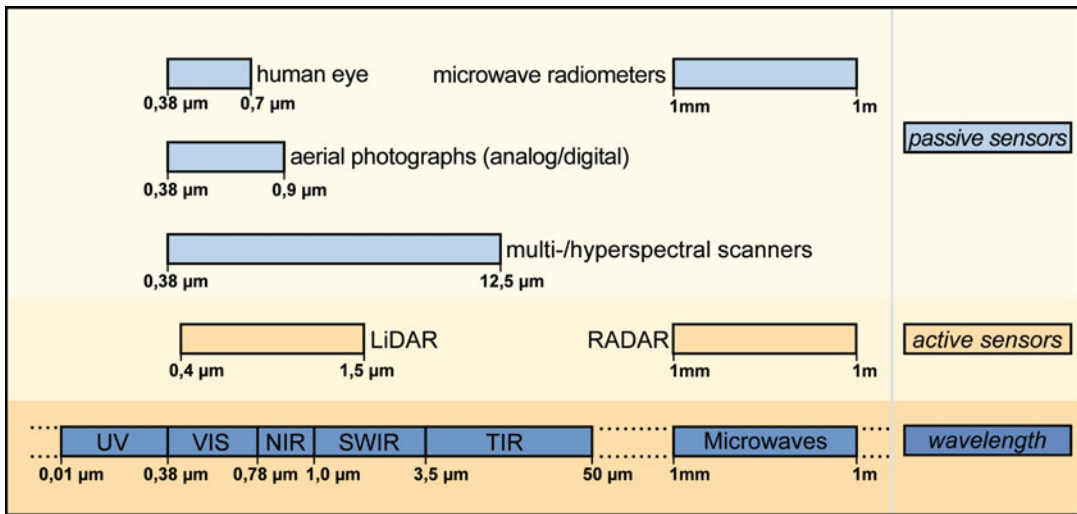


Fig. 1.13 Electromagnetic spectrum and selected wavelengths used for earth observation systems

One has to distinguish between four types of resolution, namely (Lillesand et al. 2015):

- *Temporal resolution*, which is the revisit rate or repetition rate, meaning in which time interval acquires the system another image of exactly the same area on the ground. In the case of earth observation satellites, this is dependent on the orbit parameters.
- *Spatial resolution*, which refers to the ground sampling rate or pixel size on the ground. One can imagine a pixel as the footprint on the ground, which gives a single „echo“ of reflected energy, neglecting on how it is composed of different materials. The smaller a pixel is, the more details one can identify in the image. This is especially meaningful in urban areas with a lot of material variety within short distances, which requires larger scales for better discrimination of objects and materials. In contrary to that, in the countryside with fields and forests, for example, one does not necessarily need very small pixels due to the relatively homogeneous land cover. So coarser or medium spatial resolution pixels are acceptable for studies in such landscapes associated with smaller scales. Spatial resolutions vary in the range of cm and dm pixel size for airborne systems and ca. 0,30 m and ca. 1000 m for satellite systems (see Fig. 1.14).
- *Spectral resolution* refers to the width and number of different spectral bands. Digital images are captured and stored in separate spectral bands, meaning that individual portions of the electromagnetic spectrum are separated from other parts of the spectrum. The more individual bands are available the smaller is the individual bandwidth of each band, which refers to a high spectral resolution.
- *Radiometric resolution* refers to the sensitivity of a sensor to capture the photons of the reflected energy. The radiometric resolution is better/higher in systems that can differentiate more and therefore finer levels of reflectance. Early satellite systems typically had a radiometric resolution of 8 bit, meaning that 256 different intensities of reflected energy (0–100%) could be separately stored for each pixel. Modern systems have improved sensitivity (e.g. 12 bit) and can separate finer intervals. This allows to separate 4096 individual radiation intensities and therefore much more detailed discrimination of the reflected energy.

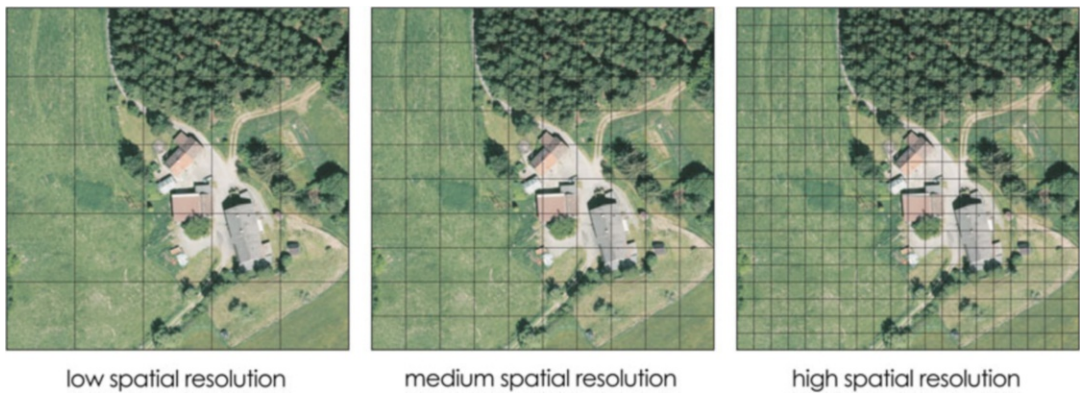


Fig. 1.14 Different examples of the spatial resolution of earth observation sensors. Larger pixels capture most likely more land cover types in one pixel than finer/smaller pixels. (Image source: dl-de/by-2-0)

For practical work with image data, one has to decide with which type of data one will work. This implies to define the scale of the investigation, and then one can decide on the necessary pixel size, which determines the level of detail. The pixel size mostly defines also the areal extent of images. As a rule of thumb, aerial images by UAV's and planes or helicopters cover a lot less ground per image than satellite images. And satellite images with a very high spatial resolution which means small pixels, cover less area per image than satellite images with coarser resolution. For large areas of interest, one has to consider to eventually stitch together images of different orbits or flight paths to cover the ground completely in an image mosaic.

One more aspect is the spectral characteristic of a sensor. Can that provide the information one is looking for? For instance, if one is interested in plant characteristics, the sensor should at least be able to capture near-infrared information, since in this portion of the electromagnetic spectrum essential information about plant conditions is located. In addition to that one has to consider the revisit rate, meaning how often one can get another „fresh“ image of the same area of interest, which is necessary in the case of heavy cloud cover on an earlier image. Satellites have fixed orbits and depending on the systems constellation, satellites can reach daily coverage or acquire images of the same area of interest after a couple

of days. All airborne platforms can acquire images on demand, as long as one has cloud-free conditions. Satellite images in the optical domain are much more affected by clouds. Due to fixed orbits, one cannot change the time of the satellites overpass. Since clouds disturb the direct observation of the ground, the user of these image data has to search in the satellite archives of the image provider for cloud-free images for the specific area of interest and specific period of the year.

An alternative chance for cloud-free images offers satellite constellation systems, which consist of a group of identical satellites that use the same orbit and fly one after the other in a fixed equal distance. With this technique (e.g. used by 5 German RapidEye satellites) the temporal resolution is reduced and the chance of a cloud-free image is increased (see Fig. 1.15).

Airborne and some satellite systems have the ability to be tilted along a track or across a track. This allows in the case of across-track, that the satellite points to an area that is not in nadir direction underneath the satellite, but aside of that. Doing the across-track tilt from two different orbits (with a time lag of a few days between the orbits and respective images) to the same target area, one gains two images with different perspectives that could be used as a so-called stereo pair. This allows the images to be used for a 3D representation of the earth. In the along-track version, one uses sensors with

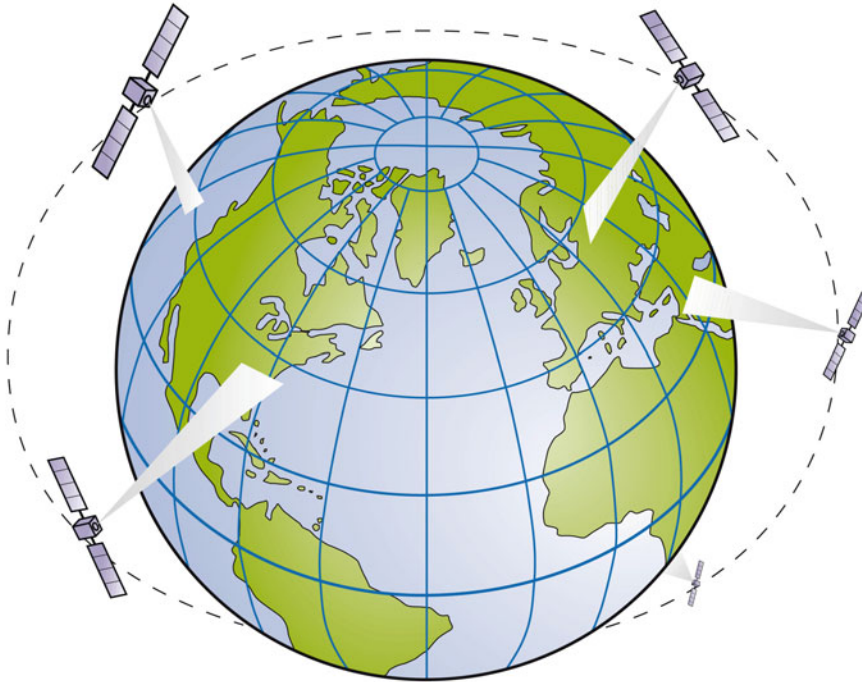


Fig. 1.15 Satellite constellation of five satellites

different viewing direction along the orbit pass: one sensor looks forward, another looks backwards. By using images of these two directions, also allows to use two different perspectives on the same area on the ground, but with almost simultaneous image capture and also allows a 3D representation of the ground. The 3D representation of the ground enables to generate digital height and digital surface models, which are requested in a number of applications (Crespi and Jürgens 2016). Due to the high repetition rates of modern earth observation systems, one can also update the 3D information in short intervals, e.g. in the case of changes mining of built-up structures on the ground.

Of course, one can also extract 2D information from the image data. Due to the fact that earth observation image data capture the specific ground situation at the time of overpass, the land cover in the resulting image reflects this specific situation. The changing characteristics of plants during the growing season or the vegetation period affects their reflective responses that can be imaged by sensor systems. Due to such seasonal effects and

agricultural activities, the land cover/land-use varies between images of the same year and between years as well. Sometimes this information of change is the needed information, e.g. if one does a change detection study (Jürgens 2000 or Henits et al. 2016) for a multi-year period to document the real land cover changes in a certain region. Sometimes these changes disturb an analysis, especially if one seeks to document the average situation. For harvested fields, for instance, one is unable to determine the crop of that year. Or if one tries to average statistical information retrieved from the image data, harvested fields could be hindering. Also, artificial surfaces can vary if new objects are constructed in an area.

The described disadvantages of images representing the snap-shot situation on the earth's surface during image capture, is also an advantage, due to the fact that images are not generalised representations of the real situation at a certain point in time. All real objects are recorded, for instance, cars, trucks, ships etc. Therefore all images are historical records or documents. Another advantage compared to areal statistics is

the fact that the images themselves and maps derived from the images show the real land use in the specific location and not an averaged situation like in statistics. For example, if one is interested in cornfields of an area, it might be interesting that an administrative unit has 30 ha of cornfields. This would be an aggregated statistical information for the administrative unit under investigation. However, to exactly know the position and extent of every single cornfield is a much more sophisticated information, which can only be gained by field inspection, which is very time consuming and therefore costly, or by earth observation images in a short time.

Area-specific information like in this example is required for spatial analysis and modeling approaches, which will be described in Chap. 3.

Due to the snap-shot properties of earth observation images, one can easily conduct time series analysis to find out differences in land use/land cover for defined regions on the globe. The oldest satellite-based earth observation data are available since 1972 from the first LANDSAT satellite Jürgens (1998). Since this year more and more different earth observation satellites came into orbit so that there is a great variety of different systems available nowadays. The latest development is the European fleet of Sentinel satellites. Those images are free of charge and have a high temporal revisit rate, a high spatial, spectral and radiometric resolution in addition to large area coverage. In addition to those freely available earth observation data from space, there are a lot of commercial providers who sell very high spatial resolution data. Typically prices of these images are calculated per square kilometer.

In the airborne domain, many countries have archives with aerial images starting with greyscale analogue World War II images. Approximately in the mid 1990ies, the analogue aerial images were transferred to colour images. In most countries, the analogue image capture techniques for aerial photographs ended around the turn to the third millennium and were transferred into digital image acquisition systems. In

summary, in the airborne domain, there are even archives that cover a much longer time span than in the satellite domain, which is of importance for time series analysis approaches.

To be able to efficiently work with image data and derived products in a GIS system, one needs to geo-reference the image data. By default, there are no coordinates associated with image data, so that a common use with other spatial data is not possible since the images „do not know where they are”.

Besides georeferencing, corrections for terrain deformation is needed to have so-called orthoimage products to be GIS-ready.

1.5.4 Orthoimage Products

As described above orthoimages are essential for further GIS-analysis in combination with further spatial data sets. Aerial images in raw format have various disadvantages, due to relief effects, radial distortions and the central perspective of the camera. This results in displaced positions of objects in the images. After proper image correction and georeferencing, the images have properties like (image) maps. Similar deformations affect satellite images, resulting in similar corrections to produce ortho satellite images. One should never consider a raw stereo image or single image for any mapping purpose or use in GIS. This rule applies to all raw earth observation images.

The following illustration (see Fig. 1.16) shows a distorted regular grid in a raw image and the corrected representation in the corresponding orthoimage. One can easily imagine how many distortions and scale variations could be in the raw image material.

1.5.5 Use of Earth Observation Image Data

These orthoimage data sets can then be used to generate up-to-date information based on the situation captured in those images. Images are the

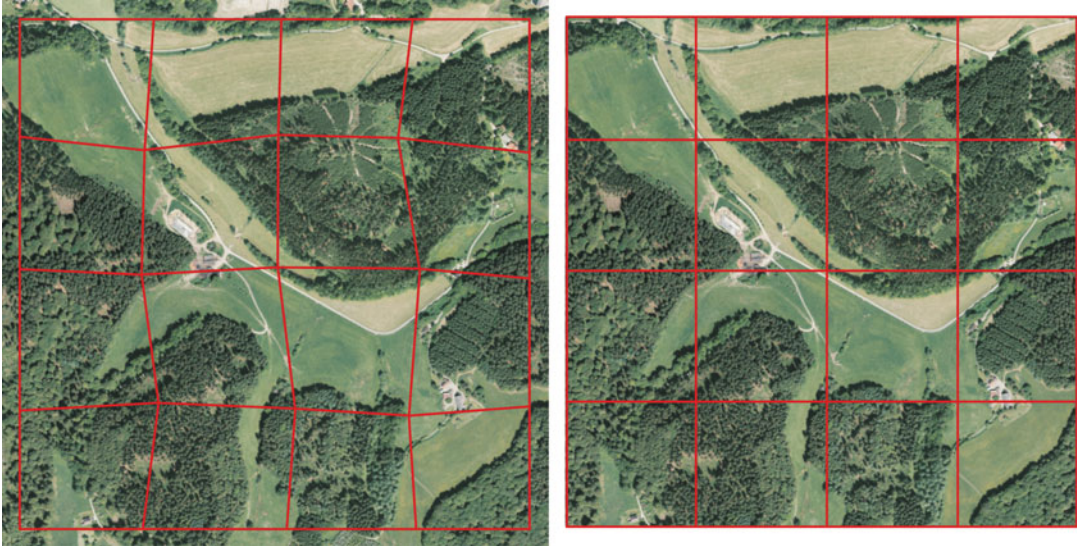


Fig. 1.16 Comparison between a raw aerial image and an orthoimage: the distorted regular grid in a raw aerial image and the corrected representation in the orthoimage shows

the deformations and scale implications in the uncorrected raw image. (Image source: dl-de/by-2-0)

basic data sets for timely information in information systems. With specialized image processing techniques the extraction of specific information as possible. Each image processing step can generate another thematic GIS layer, depending on the type of image processing analysis.

By application of classification algorithms, one can generate an up-to-date land cover/land-use map of an area of interest. For the extraction of that kind of thematic information out of the potentially ambiguous image data, one has to define a proper nomenclature to describe the land cover/land-use classes for a clear separation among each other (Thunig et al. 2011). Normally a hierarchically structured nomenclature serves most needs and allows to merge classes of lower hierarchy to a class of the next level hierarchy if needed in generalisation processes. Automated and semi-automated classification procedures then use the prior knowledge regarding the land cover/land-use in the area of interest and extract the information regarding the defined classes based on the implemented algorithms.

Very often one needs a second or even third image of the same season to account for the

seasonal variances in the crop and plant surfaces. The additional images can help to get reliable information in the sense that the classified areas are most likely true. For the accuracy assessment, one uses field data, which was not used for the classification and is tested against the classification result to determine its reliability. As a result, one can calculate the overall accuracy, users and producers accuracy. These give clear identification of how reliable each result for each individual class of the nomenclature is. It is also possible to get a map representation of the reliability of individual parcels. For further spatial analysis, the spatial reliability of data sets is a crucial point, since further analysis using the classification results or a resulting thematic map could benefit from that information if it is reliable and also could suffer from it in the case of poor quality.

1.5.6 Available Earth Observation Satellites

Due to the fact that aerial images normally are to be obtained at national agencies, these data are

not considered in this section. Therefore in this section, you will find a selection of earth observation resources, including further information on the parameters of individual imaging systems. Due to the many and dynamically changing earth observation systems, one cannot describe all systems in detail. The following links will support your search for the optimal earth observation data source for your specific application.

1.5.6.1 Sentinel Satellite Fleet

The European Space Agency (ESA) is operating a fleet of different next-generation earth observation satellites that offer images at no cost to the user. The Sentinel satellites (<https://sentinel.esa.int/>) are characterised by a high temporal revisit rate due to a concept of twin missions for each type of Sentinel satellite in one orbit. One can find all technical parameters and technical guides describing the systems in depth for proper use of the images.

1.5.6.2 Landsat Satellites

The first earth observation satellite started in 1972 with Landsat-1. The Landsat fleet of satellites (<https://landsat.usgs.gov/>) was increased with time, and today Landsat-8 is in operation. The technical parameters (e.g. resolution) have changed several times since 1972 and can be investigated on the web portal. The US Geological Survey (USGS) operates this web page and offers the longest continuously-acquired medium resolution data archive, which can be used to time series analysis and change detection.

1.5.6.3 European Space Imaging

The company European Space Imaging (<http://www.euspaceimaging.com/>) is selling spatially very high-resolution satellite images from a group of commercial satellites. Each system is described in detail and potential applications are outlined.

1.5.6.4 Additional Resources

For further reading one is encouraged to visit the following web pages to gain a deeper understanding of remote sensing.

1.5.6.5 ESA

The European Space Agency (ESA) offers a web portal (<http://www.esa.int>) that gives one a full view on the space activities of ESA. One section is dedicated to remote sensing of the earth and gives detailed information on specific missions.

1.5.6.6 ESA EOPortal

This portal (<https://eoportal.org>) lists available earth observation resources and describes the individual satellite systems.

1.5.6.7 ESA EDUSPACE

This portal (http://www.esa.int/SPECIALS/Eduspace_EN/SEM7YN6SXIG_0.html) offers in-depth background information about earth observation principles, history, and specific satellites as well.

1.5.6.8 SATIMAGINGCORP

This portal (<https://www.satimagingcorp.com/>) is dedicated to showing real-world applications in various fields. In addition to that one can find image examples of satellites in high and medium resolution. Descriptions of many earth observation satellite systems are also available.

1.5.6.9 Satellite Image Archives

The following links are some examples for satellite image archives for image search and free download:

<https://scihub.copernicus.eu/dhus/#/home>
https://lta.cr.usgs.gov/SPOT_Historical
<https://www.eodc.eu/services/data-services/>
<http://landcover.org/>
<https://earthexplorer.usgs.gov/>

References

- Albertz, J. (2013). *Einführung in die Fernerkundung. Grundlagen der Interpretation von Luft- und Satellitenbildern* (4th ed.). Darmstadt: Wissenschaftliche Buchgesellschaft.
- BKG. (2019a). *Tasks and organization*. Retrieved January 10, 2019, from <https://www.bkg.bund.de/EN/About-BKG/Tasks-and-Organization/tasks-and-organization.html>
- BKG. (2019b). *Maps and coordinates – BKG’s range of tasks*. Retrieved January 10, 2019, from <https://www.bkg.bund.de/EN/About-BKG/Tasks-and-Organization/Tasks-of-BKG/tasks-of-bkg.html>
- Bureau van Dijk. (2018) *Amadeus* [online – cit. 2018-12-10]. Available from: <https://www.bvdiinfo.com/en-gb/our-products/data/international/amadeus>
- Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing* (5th ed.). New York: Guilford Press.
- Crespi, M., & Jürgens, C. (2016). 3D remote sensing and urban remote sensing. *International Journal of Remote Sensing*, 37(15), 3437–3438. <https://doi.org/10.1080/01431161.2016.1205314>.
- ČÚZK. (2018a). *ČÚZK: Geoportál*. [https://geoportal.cuzk.cz/\(S\(2vb00ejqnx1uuiyzl4xvtokw\)\)/Default.aspx?mode=TextMeta&text=about&side=about&menu=2](https://geoportal.cuzk.cz/(S(2vb00ejqnx1uuiyzl4xvtokw))/Default.aspx?mode=TextMeta&text=about&side=about&menu=2). Accessed 28 Nov 2018.
- ČÚZK. (2018b). *Orthophoto of the Czech Republic*. [https://geoportal.cuzk.cz/\(S\(2vb00ejqnx1uuiyzl4xvtokw\)\)/Default.aspx?mode=TextMeta&text=ortofoto_info&side=ortofoto&menu=23](https://geoportal.cuzk.cz/(S(2vb00ejqnx1uuiyzl4xvtokw))/Default.aspx?mode=TextMeta&text=ortofoto_info&side=ortofoto&menu=23). Accessed 28 Nov 2018.
- ČÚZK. (2018c). *ZABAGED*. [https://geoportal.cuzk.cz/\(S\(2vb00ejqnx1uuiyzl4xvtokw\)\)/Default.aspx?mode=TextMeta&text=dSady_zabaged&side=zabaged&menu=24](https://geoportal.cuzk.cz/(S(2vb00ejqnx1uuiyzl4xvtokw))/Default.aspx?mode=TextMeta&text=dSady_zabaged&side=zabaged&menu=24). Accessed 28 Nov 2018.
- ČÚZK. (2018d). *RÚIAN*. [https://geoportal.cuzk.cz/\(S\(2vb00ejqnx1uuiyzl4xvtokw\)\)/Default.aspx?mode=TextMeta&text=dSady_RUIAN&menu=31](https://geoportal.cuzk.cz/(S(2vb00ejqnx1uuiyzl4xvtokw))/Default.aspx?mode=TextMeta&text=dSady_RUIAN&menu=31). Accessed 28 Nov 2018.
- DESTATIS. (2018). *Our mission – Statistics* [online]. Wiesbaden, 2018 [cit. 2018-11-27]. Available from: <https://www.destatis.de/EN/AboutUs/OurMission/StatisticsInfo.html>
- ESRI. (1998). *ESRI shapefile technical description*. White paper. Retrieved November 28, 2018, from <https://support.esri.com/en/white-paper/279>
- European Commission. (2014a). *INSPIRE data specification on coordinate reference systems – Technical guidelines*. Retrieved January 2, 2019, from <https://inspire.ec.europa.eu/id/document/tg/rs>
- European Commission. (2014b). *INSPIRE – An efficient way to share European spatial data!* Retrieved January 2, 2019, from <https://inspire.ec.europa.eu/documents/inspire-efficient-way-share-european-spatial-data>
- Eurostat. (2018a). *About – Overview* [online]. Luxembourg [cit. 2018-11-27]. Available from: <https://ec.europa.eu/eurostat/about/overview>
- Eurostat. (2018b). *Access to microdata – Overview* [online]. Luxembourg [cit. 2018-12-04]. Available from: <https://ec.europa.eu/eurostat/web/microdata/overview>
- Eurostat. (2018c). *Community innovation survey* [online]. Luxembourg [cit. 2018-12-04]. Available from: <https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey>
- GADM. (2018). *GADM data*. <https://gadm.org/data.html>. Accessed 28 Nov 2018
- GDI-DE. (2019). *GDI-DE: Networking data*. Retrieved January 10, 2018 from https://www.geoportal.de/EN/GDI-DE/gdi-de_artikel.html?lang=en
- Global Entrepreneurship Monitor. (2018). *Data* [online]. London [cit. 2018-12-10]. Available from: <https://www.gemconsortium.org/data>
- Henits, L., Juergens, C., & Mucci, L. (2016). Seasonal multitemporal land-cover classification and change detection analysis of Bochum, Germany, using multitemporal Landsat TM data. *International Journal of Remote Sensing*, 37(15), 3439–3454. <https://doi.org/10.1080/01431161.2015.1125558>.
- ISO. (2018a). *ISO 19115:2003. Geographic information – Metadata*. Retrieved January 2, 2019 from <https://www.iso.org/standard/26020.html>
- ISO. (2018b). *ISO/TS 19139:2007 Geographic information – Metadata – XML schema implementation*. Retrieved January 2, 2019 from <https://www.iso.org/standard/32557.html>
- Jürgens, C. (1998). *Satellitenfernerkundung – ein Überblick über derzeitige und zukünftige Aufnahmeinstrumente*. *Geoökodynamik, Jg.19* (Heft: 1–2): 119–137.
- Jürgens, C. (2000). *Change Detection – Erfahrungen bei der vergleichenden multitemporalen Satellitenbilddauswertung in Mitteleuropa*. *Photogrammetrie Fernerkundung Geoinformation*, 1/2000: 5–18.
- Jürgens, C. (2003). *Geo-Fernerkundung – Was ist das?* *Praxis Geographie*, 3/2003: 4–7.
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). New York: Wiley.
- OECD. (2018). *About the OECD* [online]. Paris, 2018 [cit. 2018-11-27]. Available from: <http://www.oecd.org/about/>
- Snyder, J. P. (1987). *Map projections – A working manual*. *U.S. geological survey professional paper 1395*

- SURS. (2018). *About statistical office* [online]. Ljubljana, 2018 [cit. 2018-11-27]. Available from: <https://www.stat.si/statweb/en/AboutUs/AboutStatOffice>
- Surveying and Mapping Authority of the Republic of Slovenia. (2018a). *Assignments and objectives | Surveying and mapping authority of the RS*. http://www.gu.gov.si/en/gu_vstopnastran/assignments_and_objectives/. Accessed 29 Nov 2018
- Surveying and Mapping Authority of the Republic of Slovenia. (2018b). *e-Surveying data*. <http://egp.gu.gov.si/egp/?lang=en>. Accessed 29 Nov 2018
- Thunig, H., Wolf, N., Naumann, S., Siegmund, A., Jürgens, C., Uysal, C., Maktav, D. (2011). Land use/land cover classification for applied urban planning – The challenge of automation. <https://doi.org/10.1109/JURSE.2011.5764762>
- UNSD. (2018a). *UNdata – A world of information* [online]. New York, 2018 [cit. 2018-11-27]. Available from: <http://data.un.org/>
- UNSD. (2018b). *Welcome to UNSD – About us* [online]. New York, 2018 [cit. 2018-11-27]. Available from: <https://unstats.un.org/home/about/>
- Vaculík, M., Pászto, V., & Švarcová, B. (2017). Spatial distribution of innovation activities in Czech Republic, 2010–2012. *Journal of International Studies Vol, 10*(1).
- World Bank Group. (2018). *What we do* [online]. Washington, DC, 2018 [cit. 2018-11-27]. Available from: https://www.wto.org/english/thewto_e/thewto_e.htm
- World Trade Organisation. (2018). *UNdata – The WTO* [online]. Geneva, 2018 [cit. 2018-11-27]. Available from: https://www.wto.org/english/thewto_e/thewto_e.htm

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

