YYT-C3001 Management of environmental data and information

Lecture 5: Spatial data model technologies

Aalto University
School of Engineering

Jussi Nikander

23.9.2020

Contents of this lecture

Data representation formats

Spatial reference systems, temporal data and metadata

Spatial data modeling in databases

GeoPackage

Learning goals for this lecture

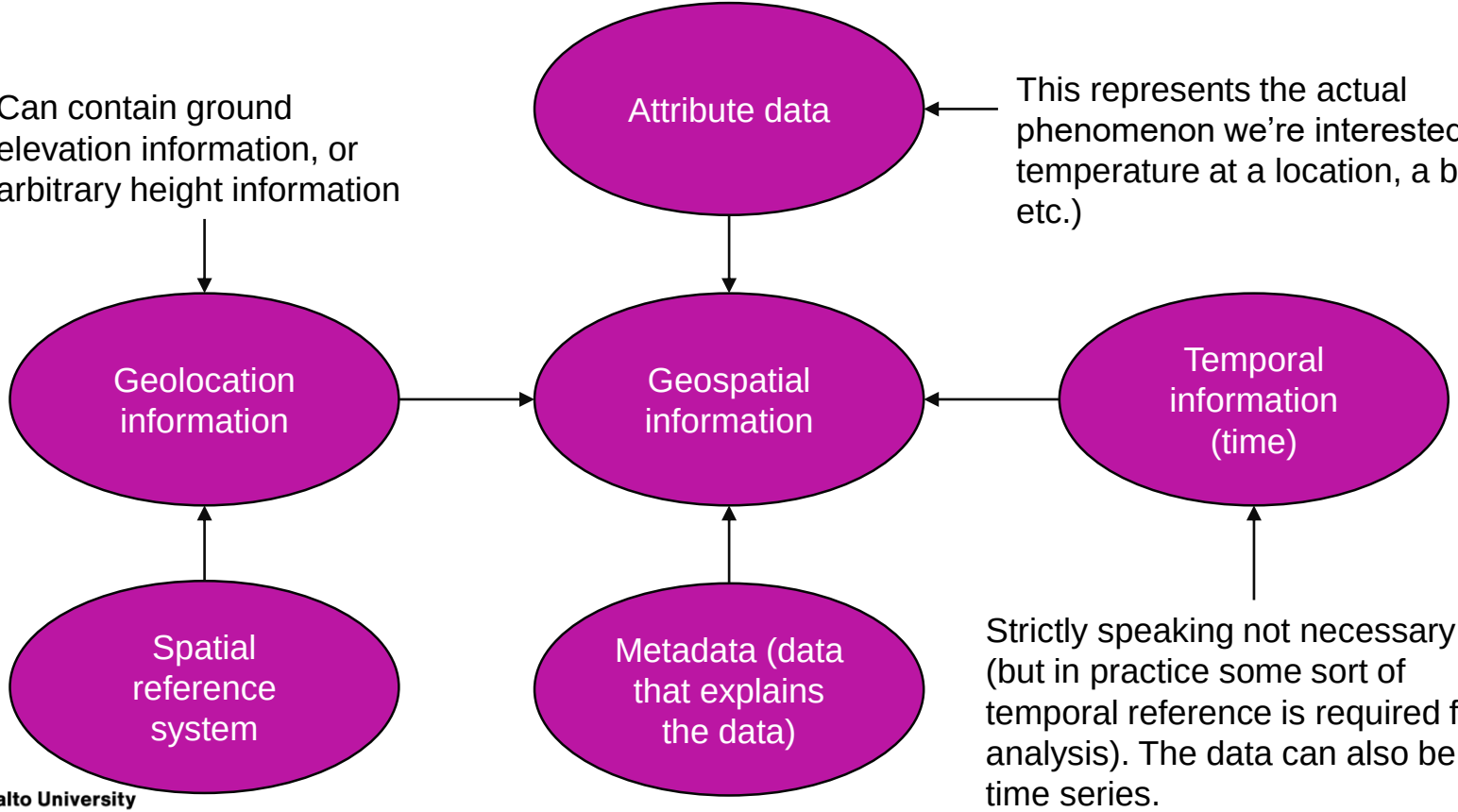
Reiterate your knowledge of raster and vector data (from MAA-C2005 Geoinformation in Environmental Modeling)

Understand what is required for a spatial dataset that can be provided to other users

Understand the contributions of the different parts of the data sets to interoperability

Familiarize yourself with the idea of GeoPackage data format

Structure of geospatial information



Geolocation information and attribute data

Lecture exercise

I will divide you into breakout rooms in Zoom

In breakout rooms consider the following

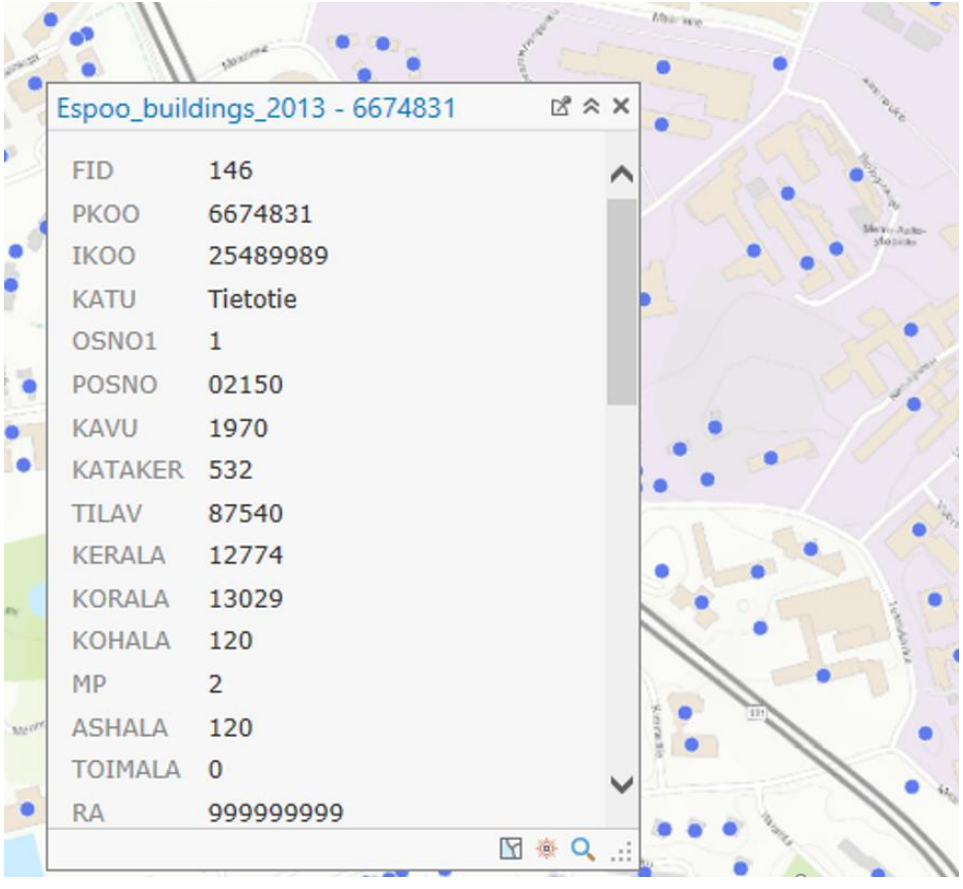
- **What are vector and raster data formats? What sort of data are they made to represent? What similarities and differences are there between the data formats?**
- **Consider representing waters (lakes, rivers, the sea) as a raster and as a vector layer. Come up with use cases where vector would be better than raster and vice versa. What in these use cases makes one data representation better than the other?**

Let's use 15 minutes for this

https://jamboard.google.com/d/1Lx86vEov4kgrPVEL0-cwIX6iWaTEv9eFXoIzTGvpQx0/edit?usp=sharing

Geospatial data example

- Each blue dot in the example represents a building
- The dot is a point data element; a coordinate value pair
- In this dataset, the shape of the building is not stored
- Each building has many attribute data elements
- Together, buildings become a spatial data layer
- In this case the layer is stored as a single file on a computer

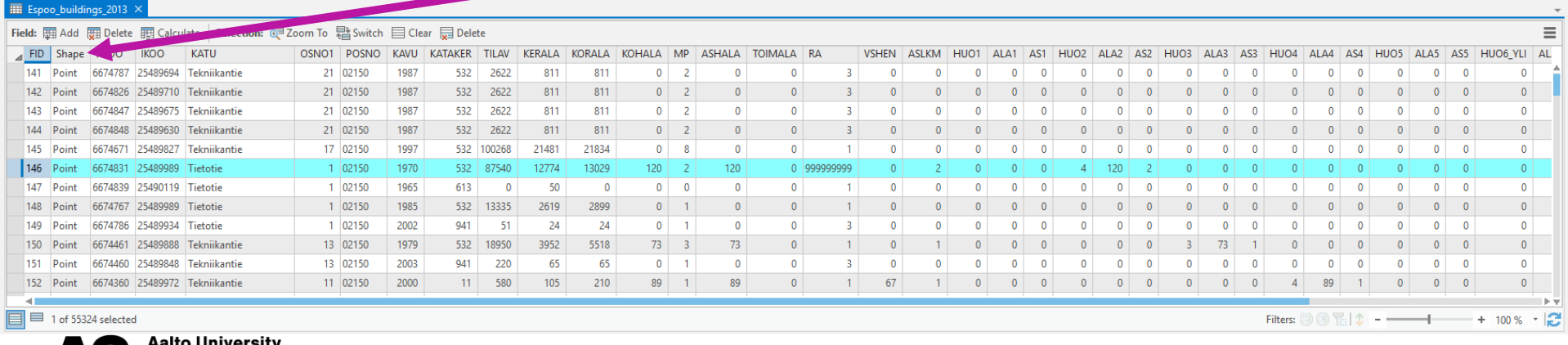


Geospatial data representation: vector data

The building data in the example is in vector data format

In vector format the data consists of discrete features that have attributes attached to them

- **Features can be points, lines, or areas (polygons)**



FID	Shape	IKOO	KATU	OSNO1	POSNO	KAVU	KATAKER	TILAV	KERALA	KORALA	KOHALA	MP	ASHALA	TOIMALA	RA	VSHEN	ASLKM	HUO1	ALA1	AS1	HUO2	ALA2	AS2	HUO3	ALA3	AS3	HUO4	ALA4	AS4	HUO5	ALA5	AS5	HUO6_YLI	AL
141	Point	6674787	25489694	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
142	Point	6674826	25489710	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
143	Point	6674847	25489675	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
144	Point	6674848	25489630	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
145	Point	6674671	25489827	Tekniikantie	17	02150	1997	532	100268	21481	21834	0	8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
146	Point	6674831	25489989	Tietotie	1	02150	1970	532	87540	12774	13029	120	2	120	0	999999999	0	2	0	0	0	4	120	2	0	0	0	0	0	0	0	0	0	0	0
147	Point	6674839	25490119	Tietotie	1	02150	1965	613	0	50	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
148	Point	6674767	25489989	Tietotie	1	02150	1985	532	13335	2619	2899	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
149	Point	6674786	25489934	Tietotie	1	02150	2002	941	51	24	24	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	Point	6674461	25489888	Tekniikantie	13	02150	1979	532	18950	3952	5518	73	3	73	0	1	0	1	0	0	0	0	0	0	3	73	1	0	0	0	0	0	0	0	0
151	Point	6674460	25489848	Tekniikantie	13	02150	2003	941	220	65	65	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
152	Point	6674360	25489972	Tekniikantie	11	02150	2000	11	580	105	210	89	1	89	0	1	67	1	0	0	0	0	0	0	0	0	0	4	89	1	0	0	0	0	0

Discrete objects and geometry types

- **In most spatial models, three spatially distinct types of discrete objects are used**
- **Point has one set of coordinate values and represents single location**
- **Polyline consists of two or more points that are linked together**
- **Polygon consist of an area enclosed by a closed polyline**

Point data (traffic accidents 2016)

Polyline data (road network)

Polygon data (municipal borders)

Geospatial data representation: vector data

In a vector data set each data element (such as building) can have numerous attribute data elements

Below is a table representation of the data set, showing all the attributes for each building

The Shape column contains spatial data, all other columns contain attribute data

Espoo_buildings_2013

Field: Add Delete Calculate Selection: Zoom To Switch Clear Delete

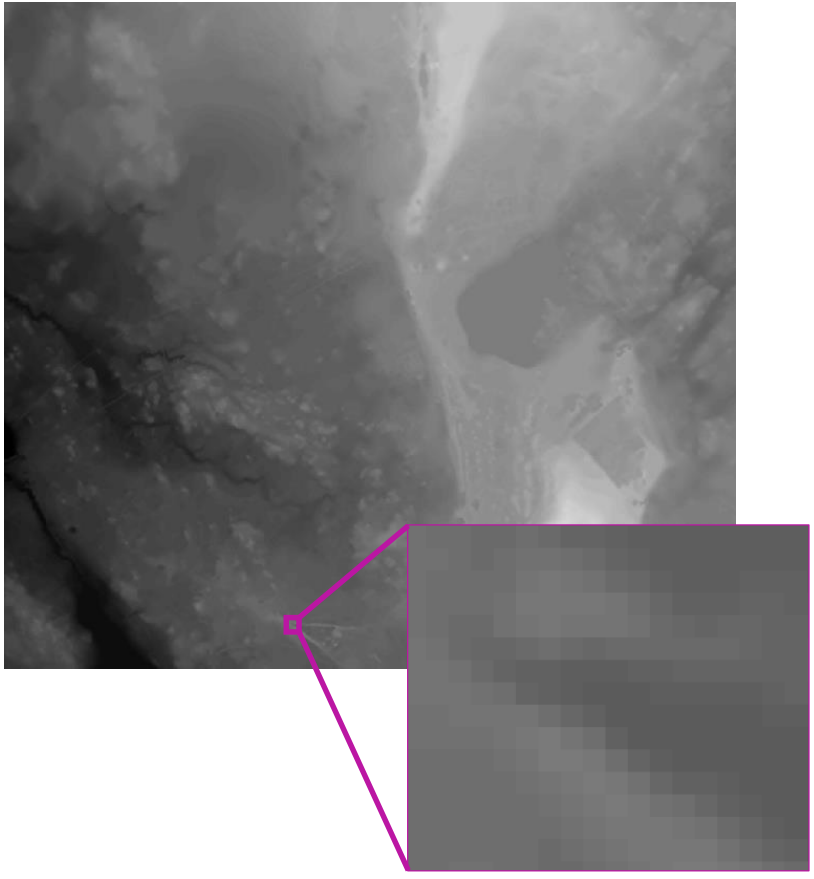
FID	Shape	PKOO	IKOO	KATU	OSNO1	POSNO	KAVU	KATAKER	TILAV	KERALA	KORALA	KOHALA	MP	ASHALA	TOIMALA	RA	VSHEN	ASLKM	HUO1	ALA1	AS1	HUO2	ALA2	AS2	HUO3	ALA3	AS3	HUO4	ALA4	AS4	HUO5	ALA5	AS5	HUO6_YLI	AL
141	Point	6674787	25489694	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
142	Point	6674826	25489710	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
143	Point	6674847	25489675	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
144	Point	6674848	25489630	Tekniikantie	21	02150	1987	532	2622	811	811	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
145	Point	6674671	25489827	Tekniikantie	17	02150	1997	532	100268	21481	21834	0	8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
146	Point	6674831	25489989	Tietotie	1	02150	1970	532	87540	12774	13029	120	2	120	0	999999999	0	2	0	0	0	4	120	2	0	0	0	0	0	0	0	0	0	0	
147	Point	6674839	25490119	Tietotie	1	02150	1965	613	0	50	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
148	Point	6674767	25489989	Tietotie	1	02150	1985	532	13335	2619	2899	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
149	Point	6674786	25489934	Tietotie	1	02150	2002	941	51	24	24	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
150	Point	6674461	25489888	Tekniikantie	13	02150	1979	532	18950	3952	5518	73	3	73	0	1	0	1	0	0	0	0	0	0	3	73	1	0	0	0	0	0	0	0	
151	Point	6674460	25489848	Tekniikantie	13	02150	2003	941	220	65	65	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
152	Point	6674360	25489972	Tekniikantie	11	02150	2000	11	580	105	210	89	1	89	0	1	67	1	0	0	0	0	0	0	0	0	0	4	89	1	0	0	0	0	

1 of 55324 selected

Filters: 100 %

Geospatial data example 2

- **The raster layer in this example represents elevation**
- **Each raster cell (pixel) covers a 10m x 10m area**
- **The cell value represents the elevation at that area**
  - Naturally, in real life elevation inside a pixel area can vary
- **The layer represents one phenomenon (~attribute value) and covers a large geographical area**
- **A background map is not visible since the raster covers the whole area**



Geospatial data presentation: raster data

Pixel value

Number of pixels with the given value

OID	Value	Count
0	55	6
1	56	266
2	57	2165
3	58	2710
4	59	3210
5	60	7978
6	61	7816
7	62	7517
8	63	5006
9	64	7423
10	65	11426
11	66	9232
12	67	11966
13	68	20004
14	69	29501
15	70	49254
16	71	32345
17	72	42959
18	73	34089
19	74	35847
20	75	41423
21	76	45539
22	77	44626
23	78	42234
24	79	73496
25	80	78952
26	81	72616
27	82	100344
28	83	92785
29	84	102384
30	85	124740
31	86	90894
32	87	123729
33	88	104687
34	89	115389
35	90	167639
36	91	108970

In raster data format, the data consists of a regular tessellation that covers an area

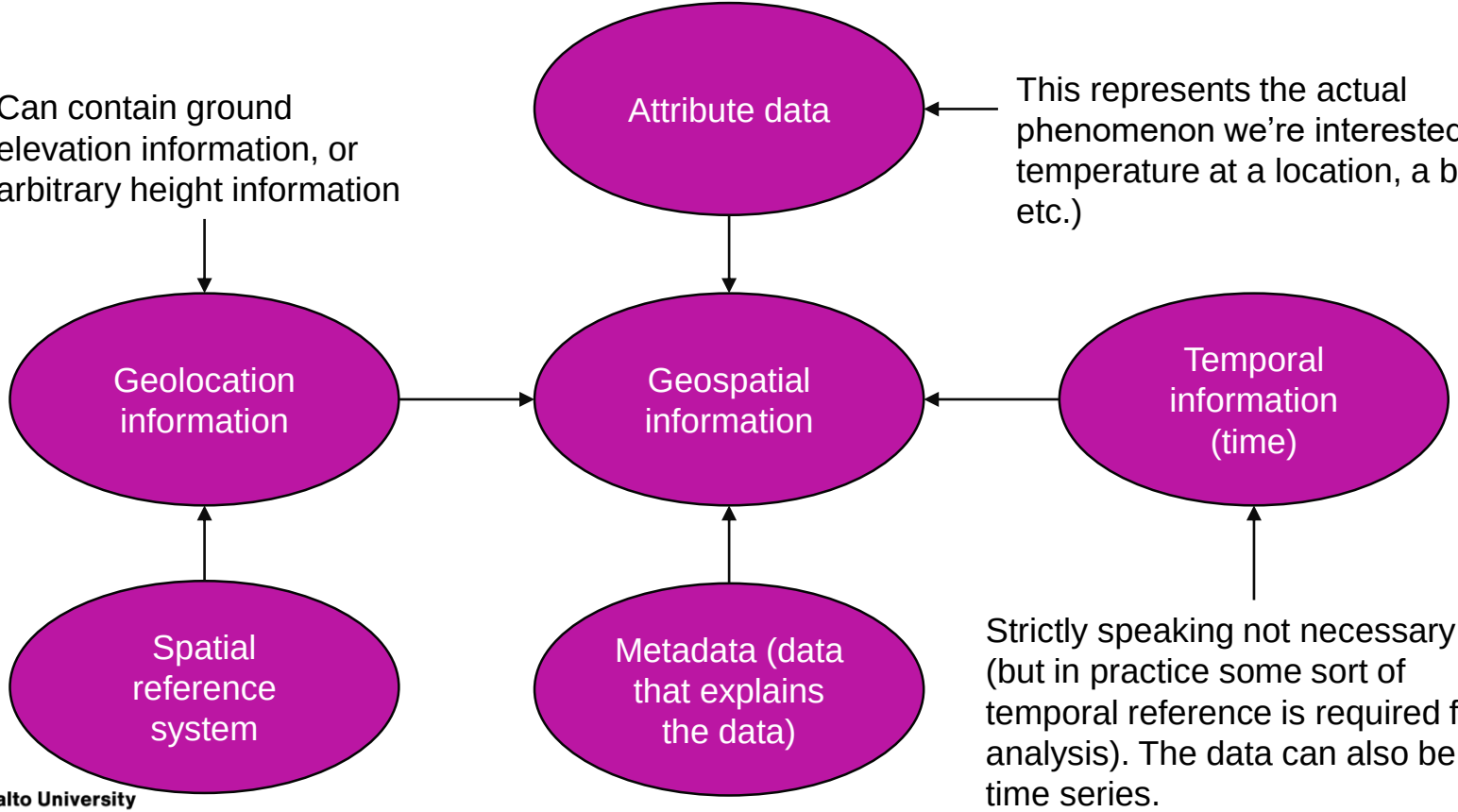
A raster layer represents a continuous phenomenon that can be measured

- **The area is divided into small, regular polygons, typically squares (triangles and hexagons can also be used)**

Each cell (pixel) in the raster represents the value of the phenomenon on the area covered by the cell

If many phenomena are to be covered, several raster layers are required

Structure of geospatial information



Spatial references, temporal data, and metadata

Aalto University
School of Engineering

Spatial reference systems

Coordinate systems

- **Geographic coordinates (spherical surface)**
  - E.g. ETRS89
- **Projected coordinates (planar surface)**
  - E.g. ETRS-TM35FIN

Map projections transform geographic coordinates (lat, lon) into planar coordinates (x, y)

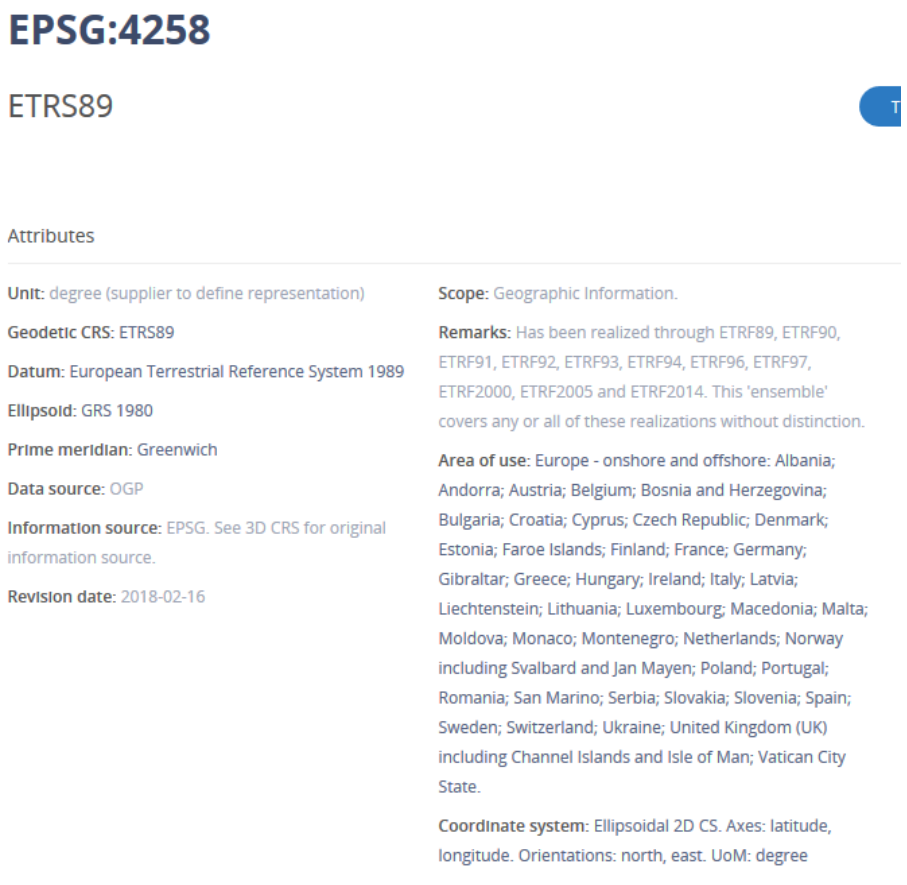
A spatial reference defines the parameters required for using a specific coordinate system

Spatial reference system example: ETRS89

A reference system covering Europe, where the Eurasian plate is static

The EPSG number is a Spatial Reference ID, which unambiguously identifies this SRS

If you don't know what SRS your dataset has, you have a problem



EPSG:4258

ETRS89

Attributes

Unit: degree (supplier to define representation)

Geodetic CRS: ETRS89

Datum: European Terrestrial Reference System 1989

Ellipsoid: GRS 1980

Prime meridian: Greenwich

Data source: OGP

Information source: EPSG. See 3D CRS for original information source.

Revision date: 2018-02-16

Scope: Geographic Information.

Remarks: Has been realized through ETRF89, ETRF90, ETRF91, ETRF92, ETRF93, ETRF94, ETRF96, ETRF97, ETRF2000, ETRF2005 and ETRF2014. This 'ensemble' covers any or all of these realizations without distinction.

Area of use: Europe - onshore and offshore: Albania; Andorra; Austria; Belgium; Bosnia and Herzegovina; Bulgaria; Croatia; Cyprus; Czech Republic; Denmark; Estonia; Faroe Islands; Finland; France; Germany; Gibraltar; Greece; Hungary; Ireland; Italy; Latvia; Liechtenstein; Lithuania; Luxembourg; Macedonia; Malta; Moldova; Monaco; Montenegro; Netherlands; Norway including Svalbard and Jan Mayen; Poland; Portugal; Romania; San Marino; Serbia; Slovakia; Slovenia; Spain; Sweden; Switzerland; Ukraine; United Kingdom (UK) including Channel Islands and Isle of Man; Vatican City State.

Coordinate system: Ellipsoidal 2D CS. Axes: latitude, longitude. Orientations: north, east. UoM: degree

Source: epsg.io

SRS lecture exercise

I'll give you three pictures of a map view containing a background map and a population tile data. I'll also give you three spatial reference systems.

Connect the map view to the correct SRS. Consider why you answer the way you do.

Let's use at 10 minutes for this; it is individual work – after 5 minutes I'll put up a presemo where you can put your answers.

A

B

C

Spatial reference systems:
1. EPSG 3067 (ETRS-TM35FIN)
2. EPSG 3857 (WGS84 web pseudo-Mercator)
3. EPSG 4258 (ETRS89)

Bonus question: which of the above images represent appropriate use of the corresponding SRS?

Fill in your answers at https://presemo.aalto.fi/medi1

A - 2

B - 3

C - 1

Spatial reference systems:
1. EPSG 3067 (ETRS-TM35FIN)
2. EPSG 3857 (WGS84 web pseudo-Mercator)
3. EPSG 4258 (ETRS89)

Bonus question: which of the above images represent appropriate use of the corresponding SRS?

Bonus answer: EPSG 3857 covers the whole Earth, so it is used appropriately.

Temporal data

Spatial dataset should always have a temporal aspect

- **A dataset can consist of several time steps (e.g. weather forecast)**
- **A dataset can represent the situation at a specific moment of time**

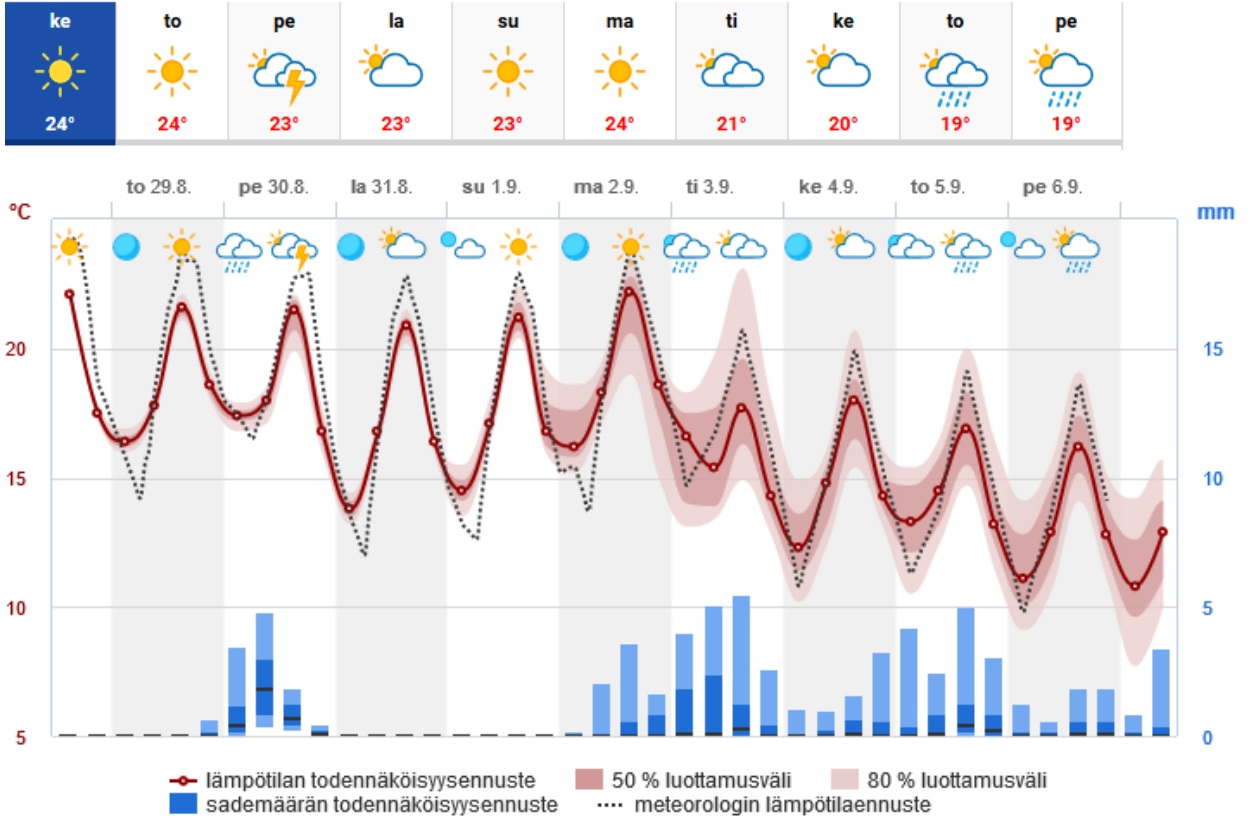
A dataset that consist of several time steps can be very sensitive to the length of the time step

- **Hourly weather forecast vs. daily weather forecast**

Temporal data example (non-spatial view)

A weather forecast for a 10-day period on both daily resolution, and on a 6-hour resolution

Notice especially the temperature on daily level compared to temperature in the graph



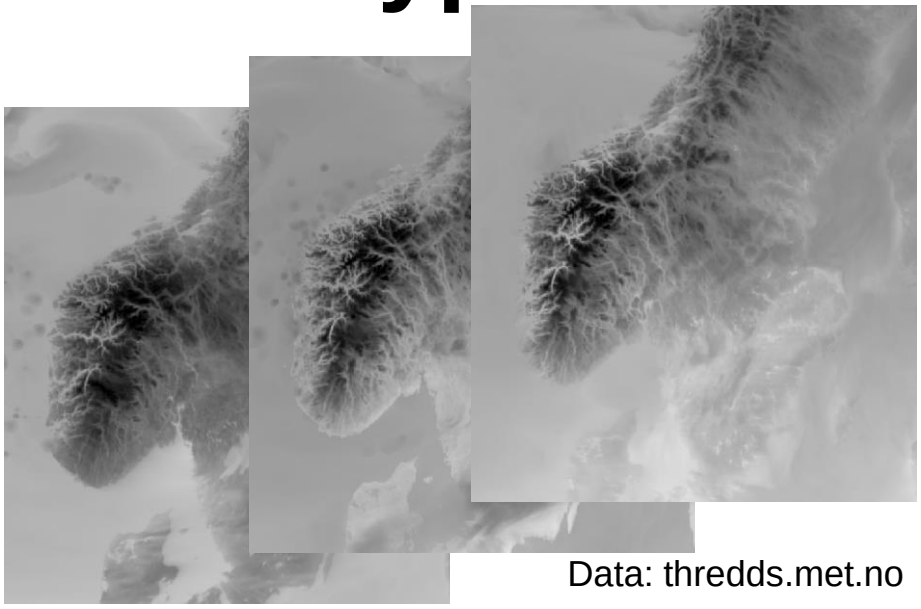
Aalto University
School of Engineering

Temporal data example (spatial view)

https://www.youtube.com/watch?v=RqmeJCBF_mg

Spatio-temporal data set types

- **Change with time in spatial data is typically expressed with several spatial layers that each represent the situation at a given point of time**
- **One common application are routes, which describe movement and consist of linked locations with associated time stamps**



Data: thredds.met.no

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., ... & Theodoridis, Y. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, *45*(4), 1-32.

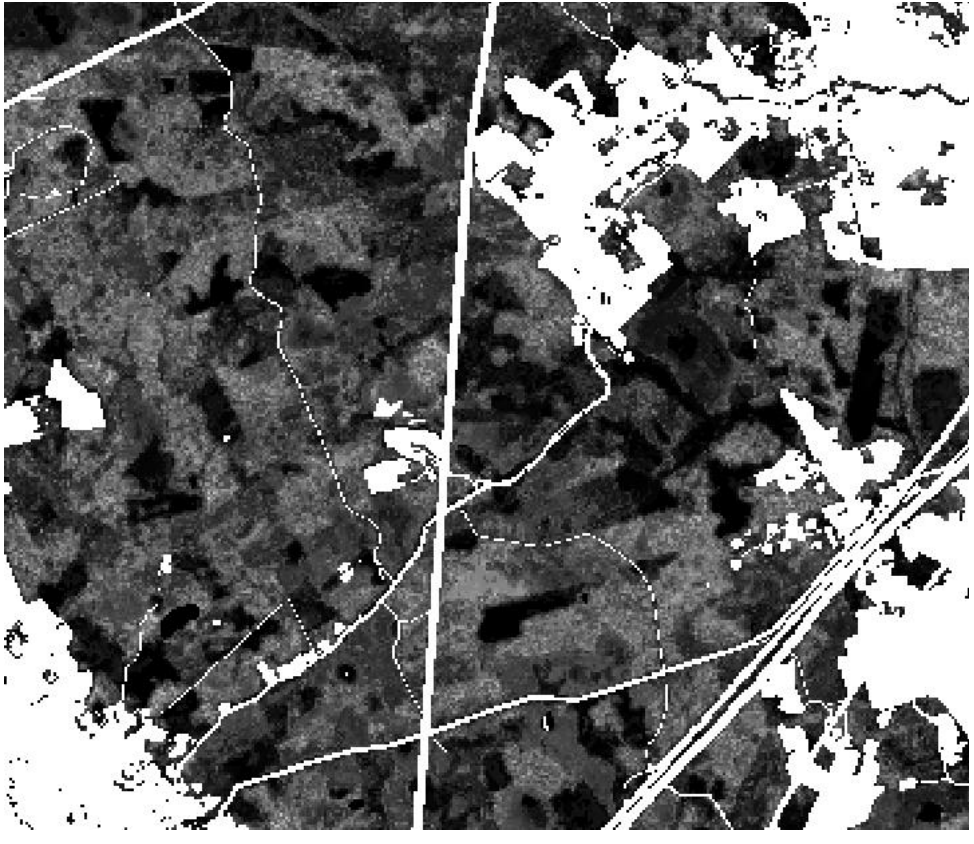
Metadata

- Spatial data typically has a complex structure
- This structure needs to be known (typically in a machine-readable format) in order to exploit spatial data
- In addition, the users need to know what a given spatial dataset represents, and what it therefore can be used for
- Therefore metadata (data about the data) is important for spatial datasets

Human-readable metadata: an example

“The estimation errors at pixel level are rather high but decrease when the area in question increases, i.e., when the area of interest consists of several pixels.”

- **Using this data for analysis on individual pixel level may be a bad idea**
- **The (potentially) large errors on single pixels will persist through the analysis**
- **When using larger areas in the analysis, the errors even out**



Data: Multi-source national forest inventory 2017, total volume of trees

Geospatial data modeling

Aalto University
School of Engineering

Spatial data model

A spatial data model defines how the real world can be approximated on a computer system

A model can be based either on the vector or the raster approach

- Raster models are typically conceptually relatively simple; a raster image and appropriate spatial metadata for spatial reference system, raster position, and data semantics
- Vector data models can be conceptually more complex, as they need to define the various geographic primitives used, as well as their connections, etc.

Defining a (spatial) data model

Data models need to be unambiguously and precisely defined

- **Everyone using a data model needs to implement it the same way**

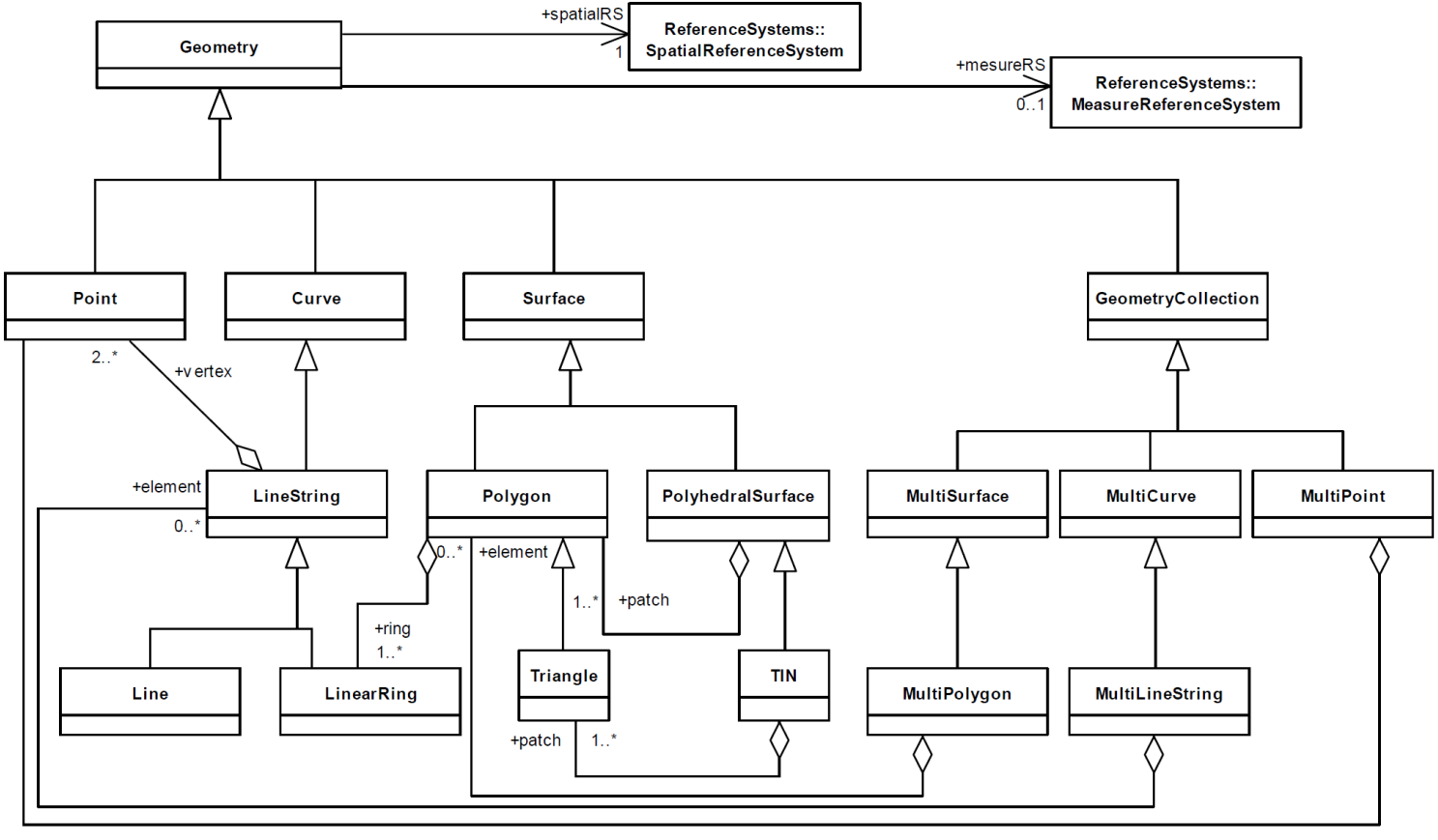
A data model needs to be put into the right context

- **A spatial data model defines how the real world is approximated on the computer in a specific universe of discourse (a specific domain)**

The model definition method needs to be (sufficiently) formal

- **UML (Unified modeling language)**
- **ER-diagram (Entity-Relationship diagram)**
- **Etc.**

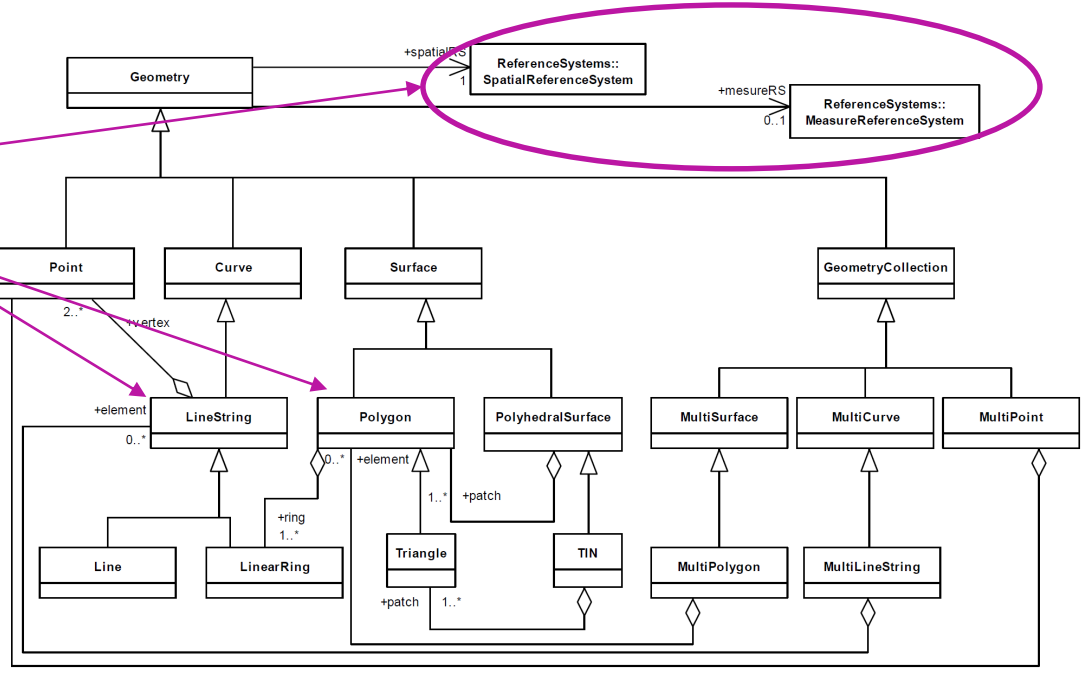
Spatial data model example: OGC Simple Features



A?

Spatial data model example: OGC Simple Features

- **Simple features defines explicitly**
  - Spatial reference
  - Location and shape
  - Some spatial relations
- **Simple features does not define**
  - Attribute data
  - Temporal information
  - Metadata
- **Implemented models need to include these elements**



From data models to database schemas

- **Spatial data typically resides in (often large) databases**
- **A vector model (e.g. Simple Features) can relatively easily be transformed into a database model**
  - Standardized in the ST_* types
  - ISO/IEC 13249-3
- **ER-model (Entity-Relationship) is often used to define SQL database schemas**

- **The Simple features and ST_* -type hierarchy define what sort of data elements can be used in any database that supports them**
- **An ER-model describes how a specific database is designed**

Vector data in spatial databases: SQL-MM and the ST_* types

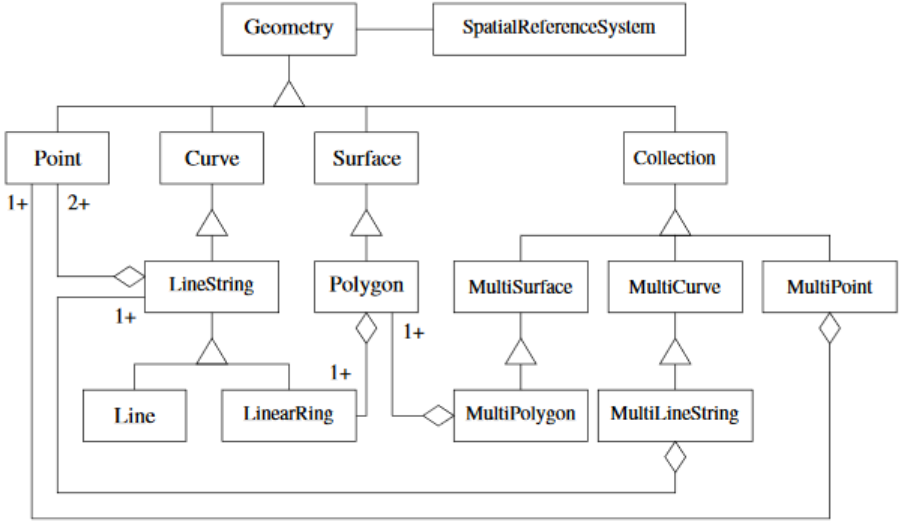


Figure 1: OpenGIS Geometry Class Hierarchy

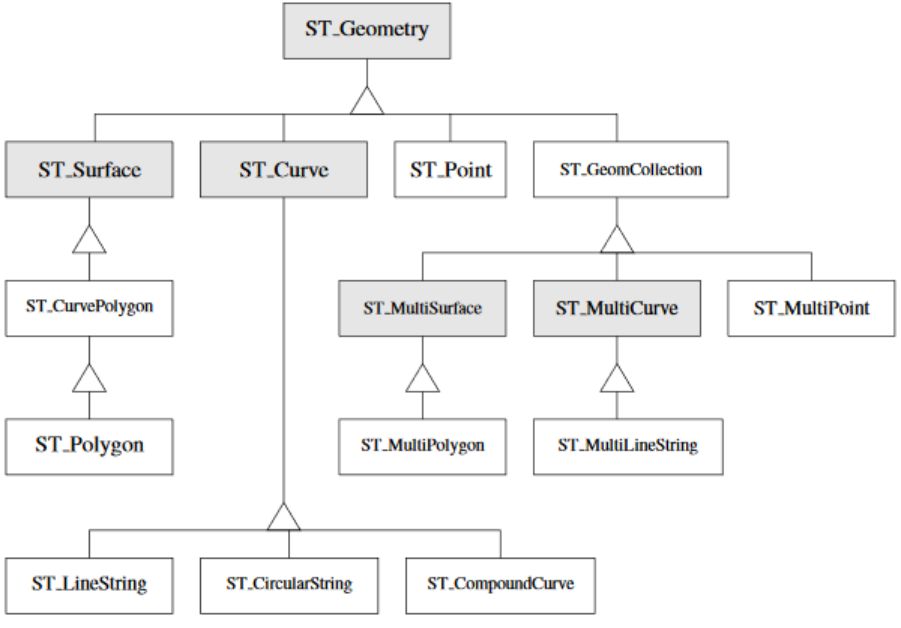
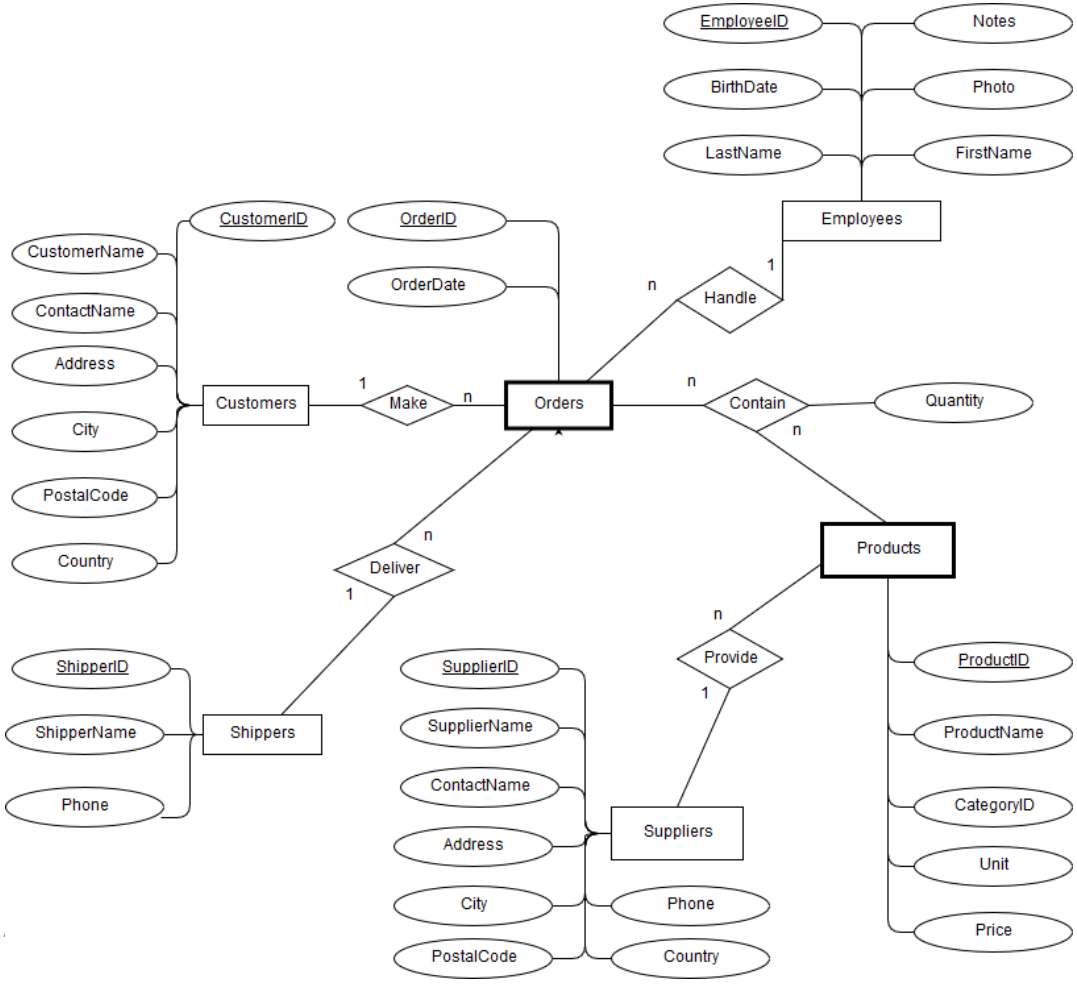


Figure 2: SQL Type Hierarchy

Aalto University
School of Engineering

Figures from Stolze, K., 2003, February. SQL/MM Spatial-The Standard to Manage Spatial Data in a Relational Database System. In *BTW* (Vol. 2003, pp. 247-264).

ER Diagram example

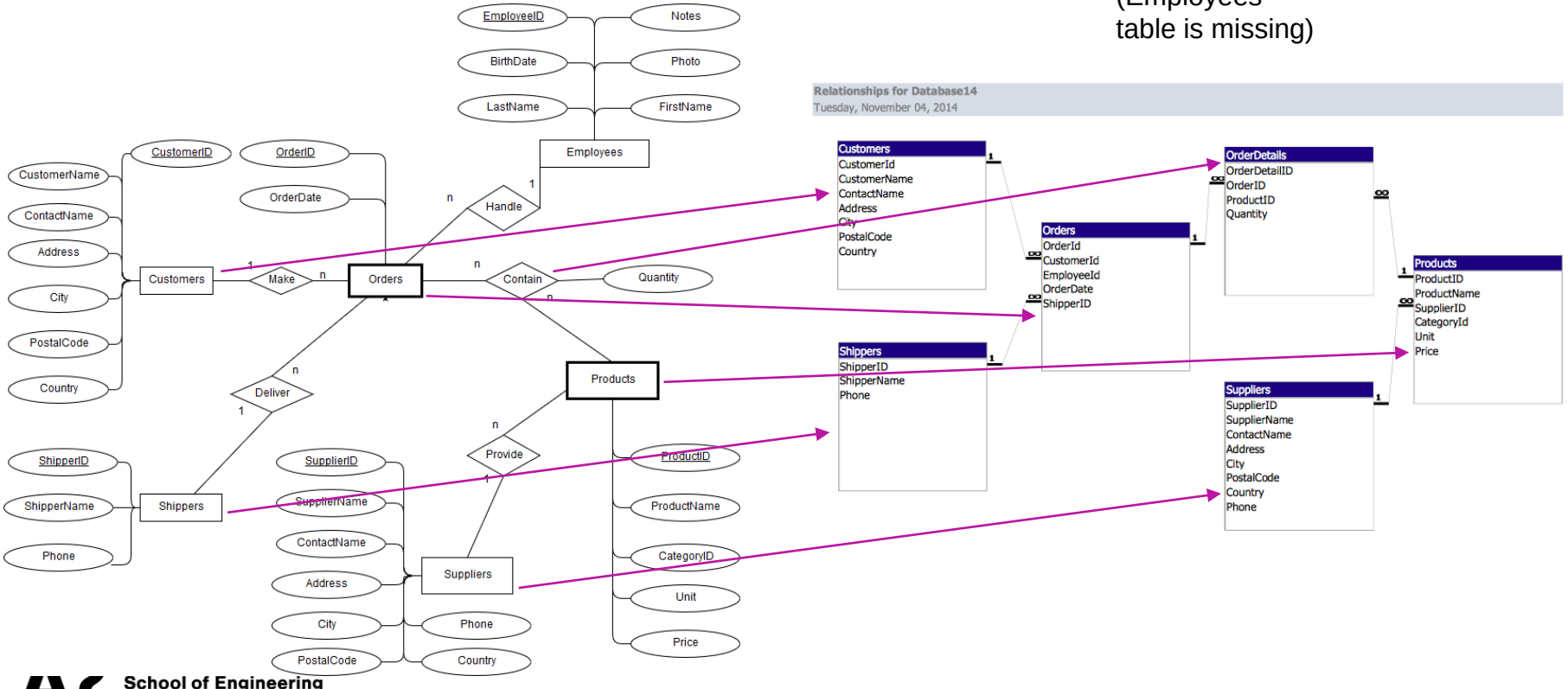


- Rectangles are entities
- Diamonds are relationships
- Ellipses are attributes

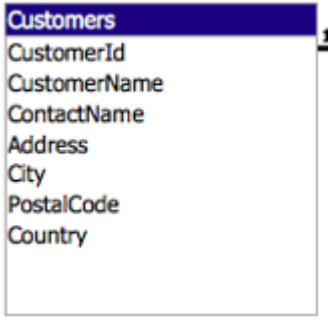
- Entities (and some relationships) are turned into database tables in a DB implementation
- Attributes are columns in a table
- Some of the attributes may be vector data
- Often in a spatial database there are very few, if any, relationships
- Raster layers can be a bit more complicated

ER-diagram to database schema

(Employees
table is missing)



Database in practice



Number of Records: 91

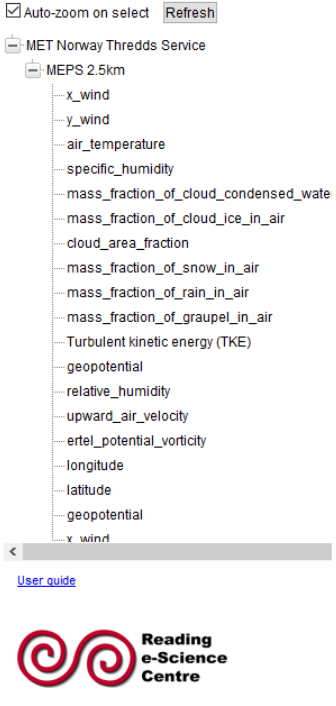
CustomerID	CustomerName	ContactName	Address	City	PostalCode	Country
1	Alfreds Futterkiste	Maria Anders	Obere Str. 57	Berlin	12209	Germany
2	Ana Trujillo Emparedados y helados	Ana Trujillo	Avda. de la Constitución 2222	México D.F.	05021	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mataderos 2312	México D.F.	05023	Mexico
4	Around the Horn	Thomas Hardy	120 Hanover Sq.	London	WA1 1DP	UK
5	Berglunds snabbköp	Christina Berglund	Berguvsvägen 8	Luleå	S-958 22	Sweden
6	Blauer See Delikatessen	Hanna Moos	Forsterstr. 57	Mannheim	68306	Germany
7	Blondel père et fils	Frédérique Citeaux	24, place Kléber	Strasbourg	67000	France
8	Bólido Comidas preparadas	Martín Sommer	C/ Araquil, 67	Madrid	28023	Spain

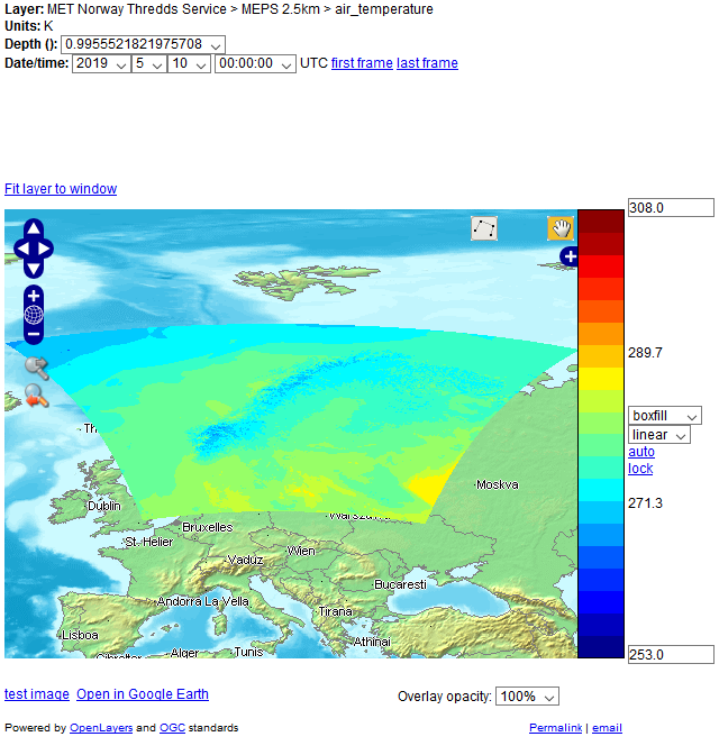
Spatial datasets

One or more spatial data layers create a spatial dataset

The layers may be explicitly connected to each other (e.g. database relationships)

Or the layers may have no explicit connections but create a thematic set





Data: thredds.met.no weather data; air temperature layer shown on map

Aalto University
School of Engineering

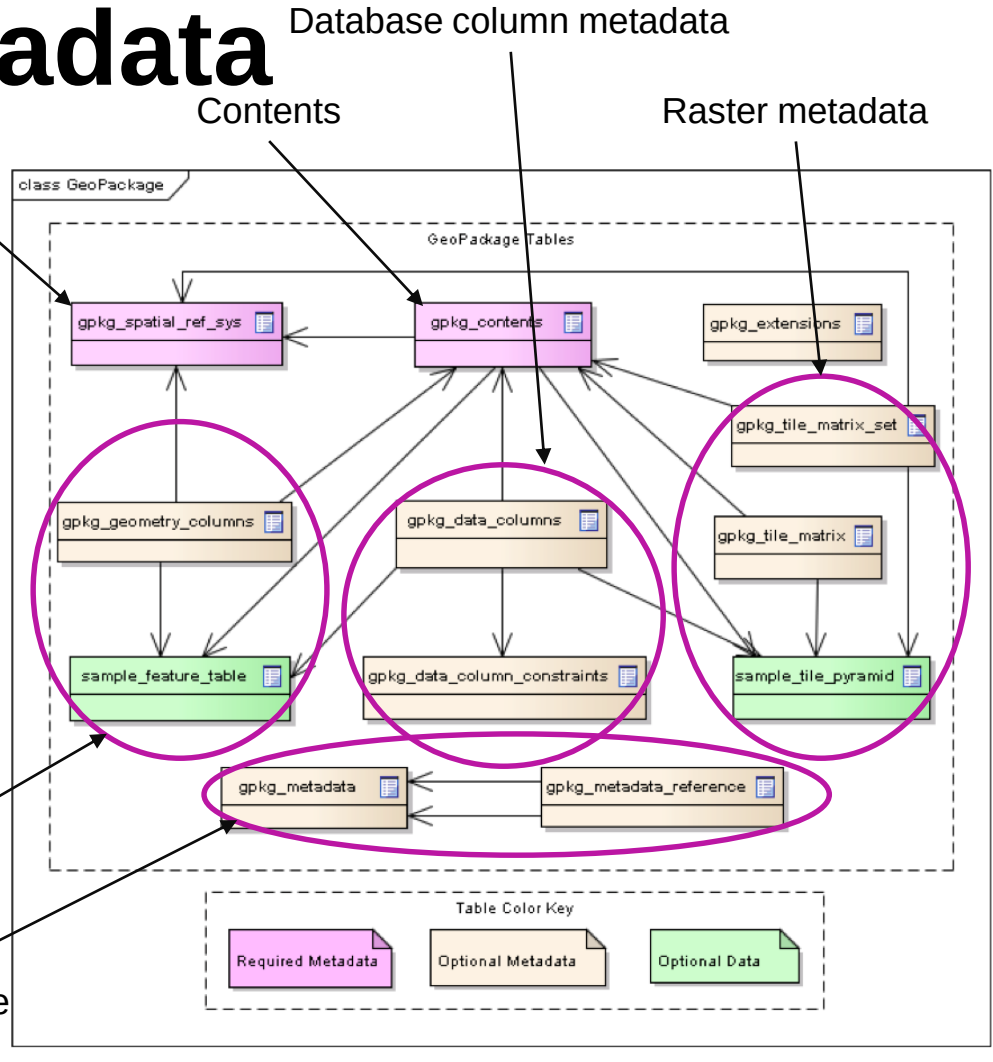
GeoPackage

GeoPackage

- **An open, non-proprietary standard format for geospatial data storage**
- **Goal is to be a general format for storing spatial data across different systems**
- **Can contain both vector and raster data**
- **Implementation is a SQLite database container**
- **Contains**
  - Metadata that specifies what a specific geopackage contains
  - The data

GeoPackage metadata

- **The GeoPackage metadata specifies the contents of a dataset**
  - Spatial reference & dataset index (contents)
  - Vector-specific metadata
  - Raster-specific metadata
  - Dataset metadata
- **The rest**
  - Extensions



For the next time…

Third exercise round has been published

Remember, next week you need to submit your learning diary for peer assessment.

Submit your diary on September 30th at the latest!

And remember, your learning diary will not be ready next week, so don’t stress about that. Just submit what you have ready, but submit it **on time**, so the diaries can be given for peer assessment.