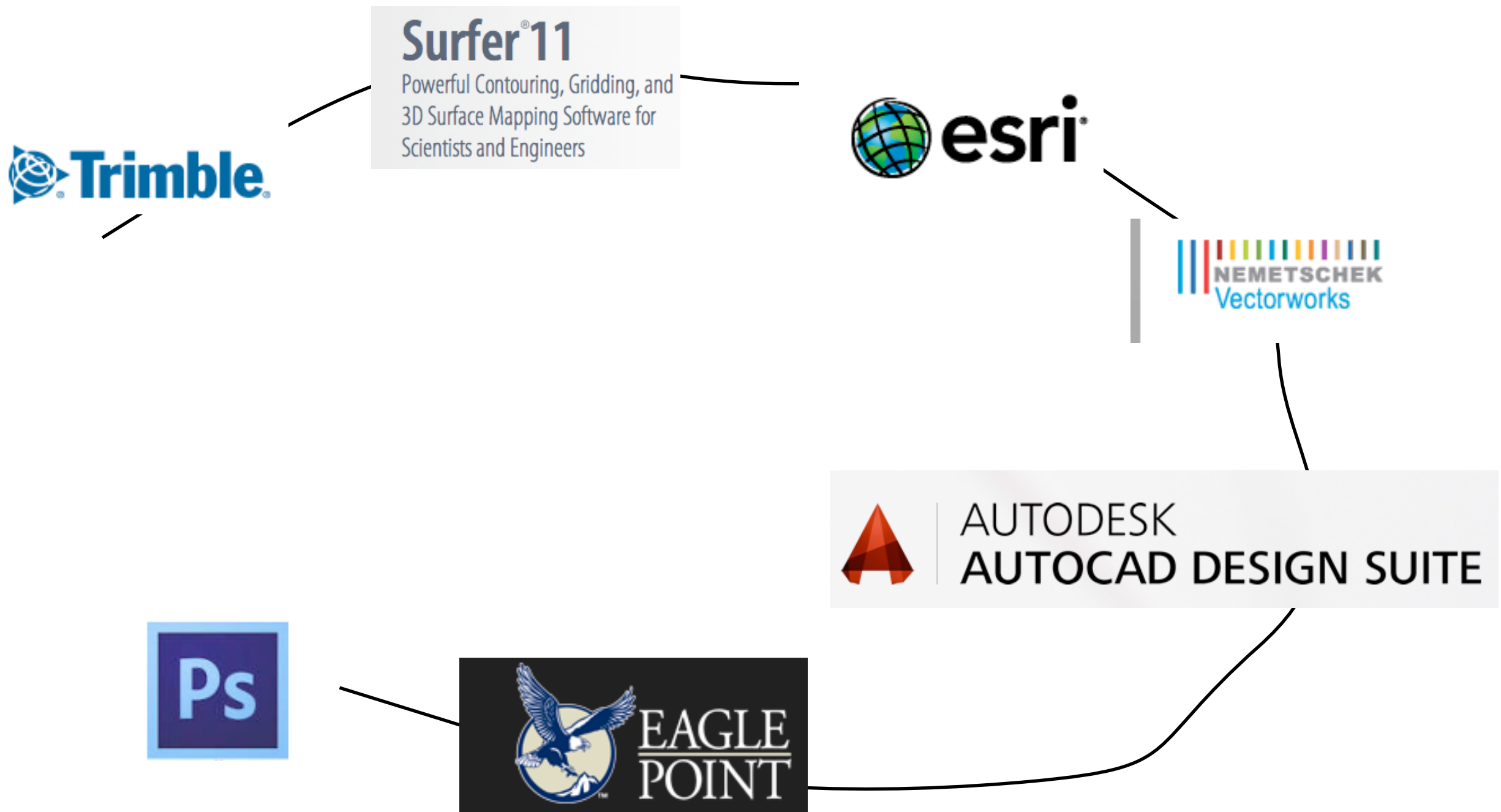
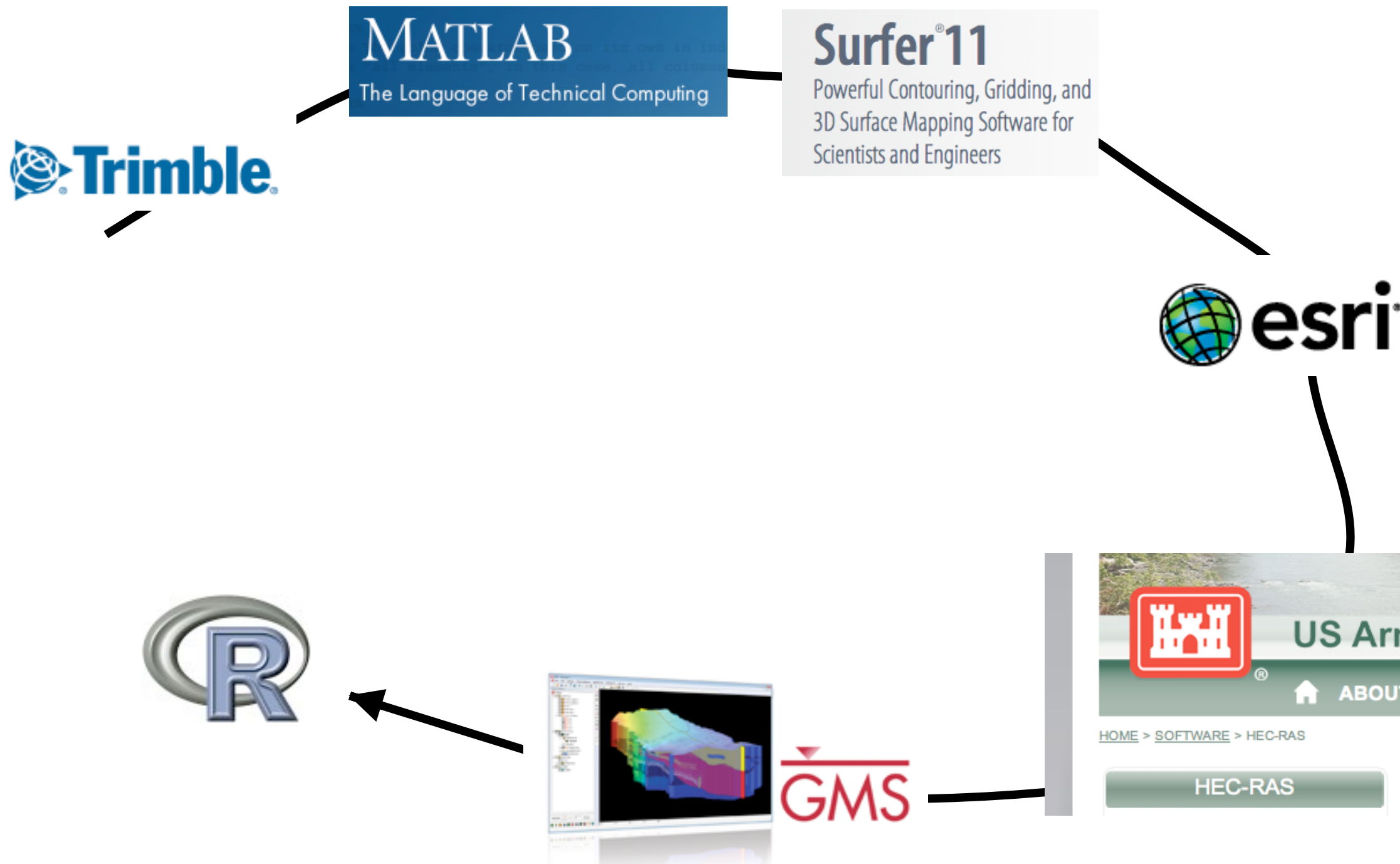


Data and File Structures: Why Should I Care?



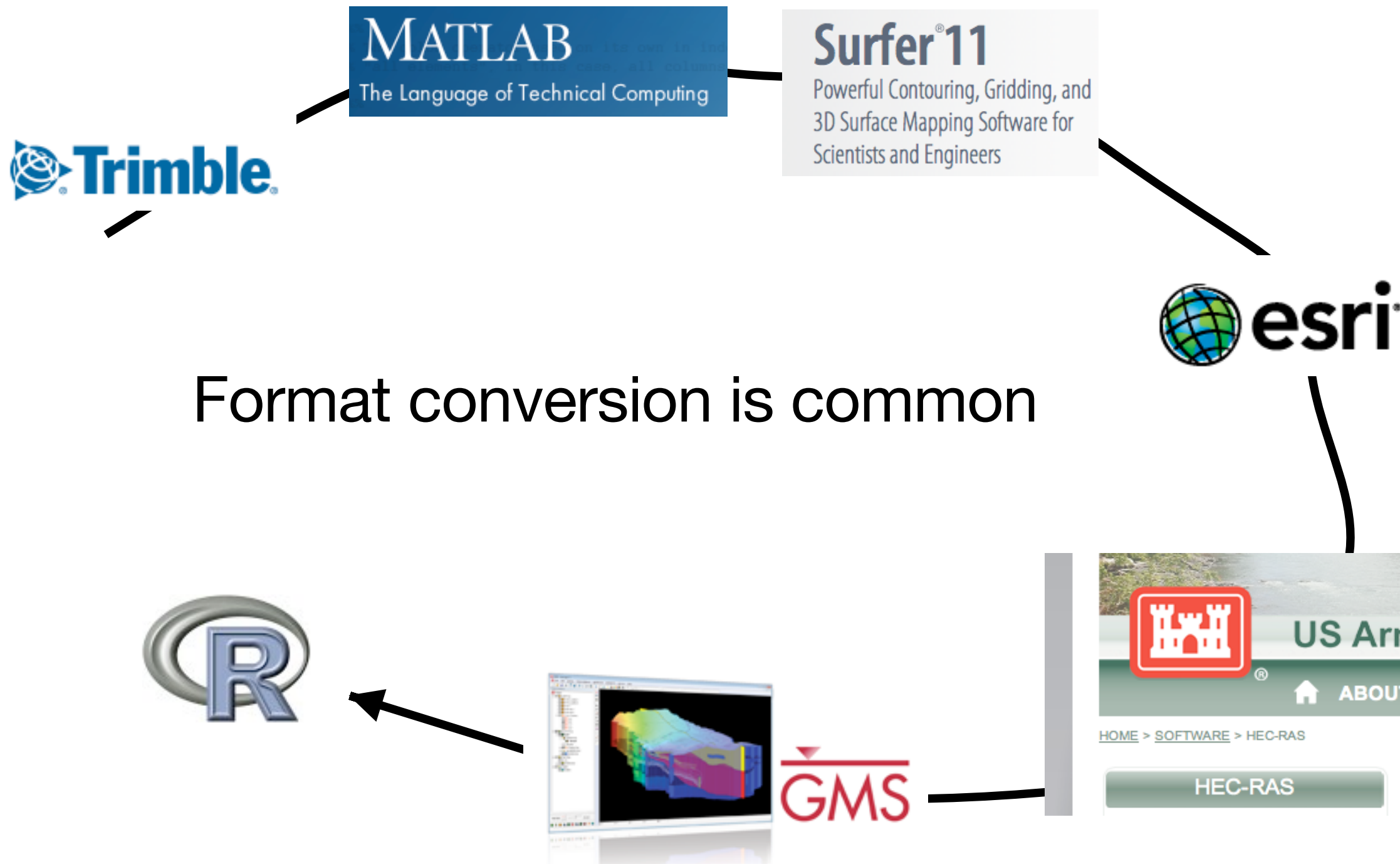
Data and File Structures: Why Should I Care?

Hydrologic analysis data path

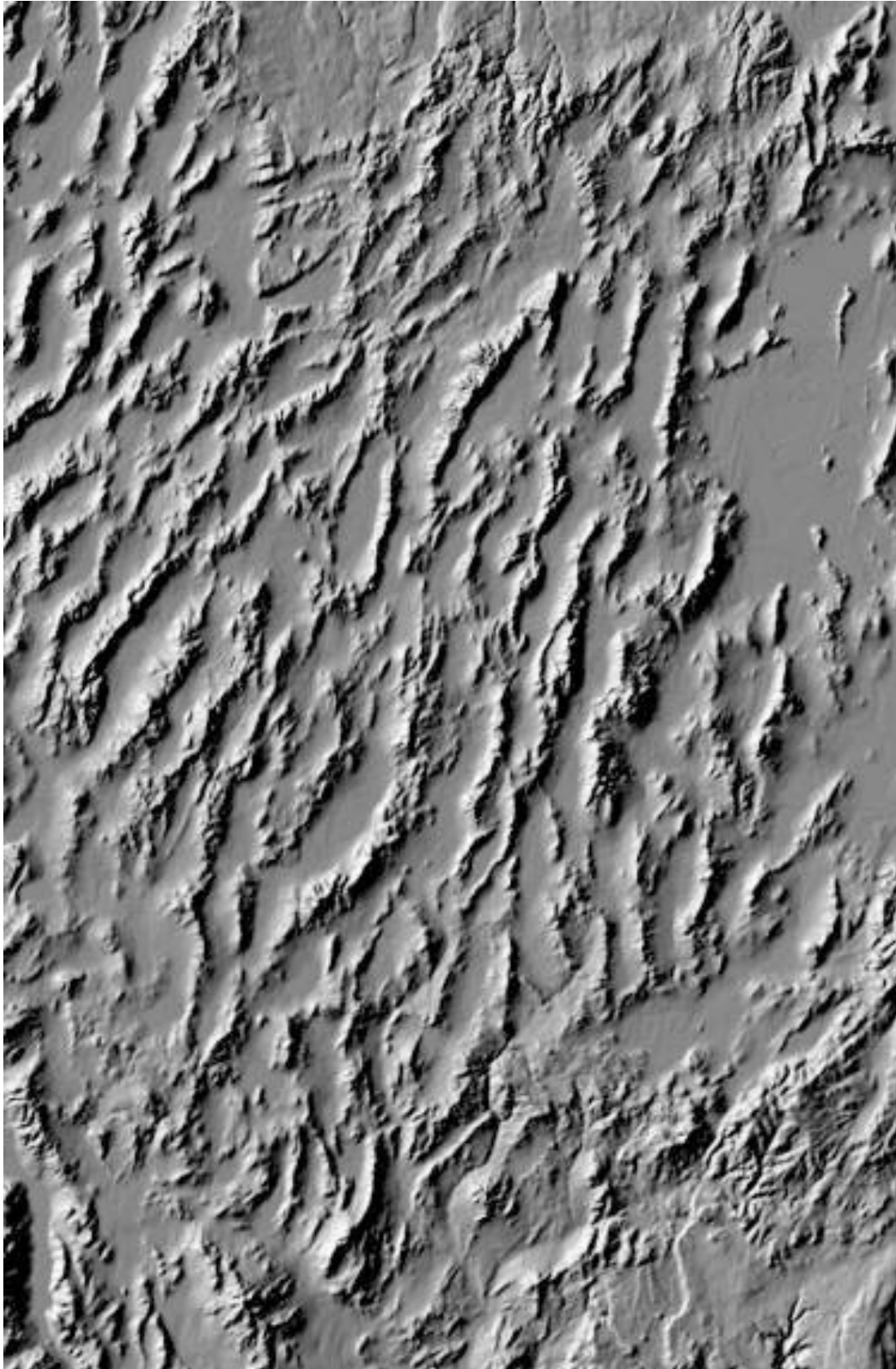


Data and File Structures: Why Should I Care?

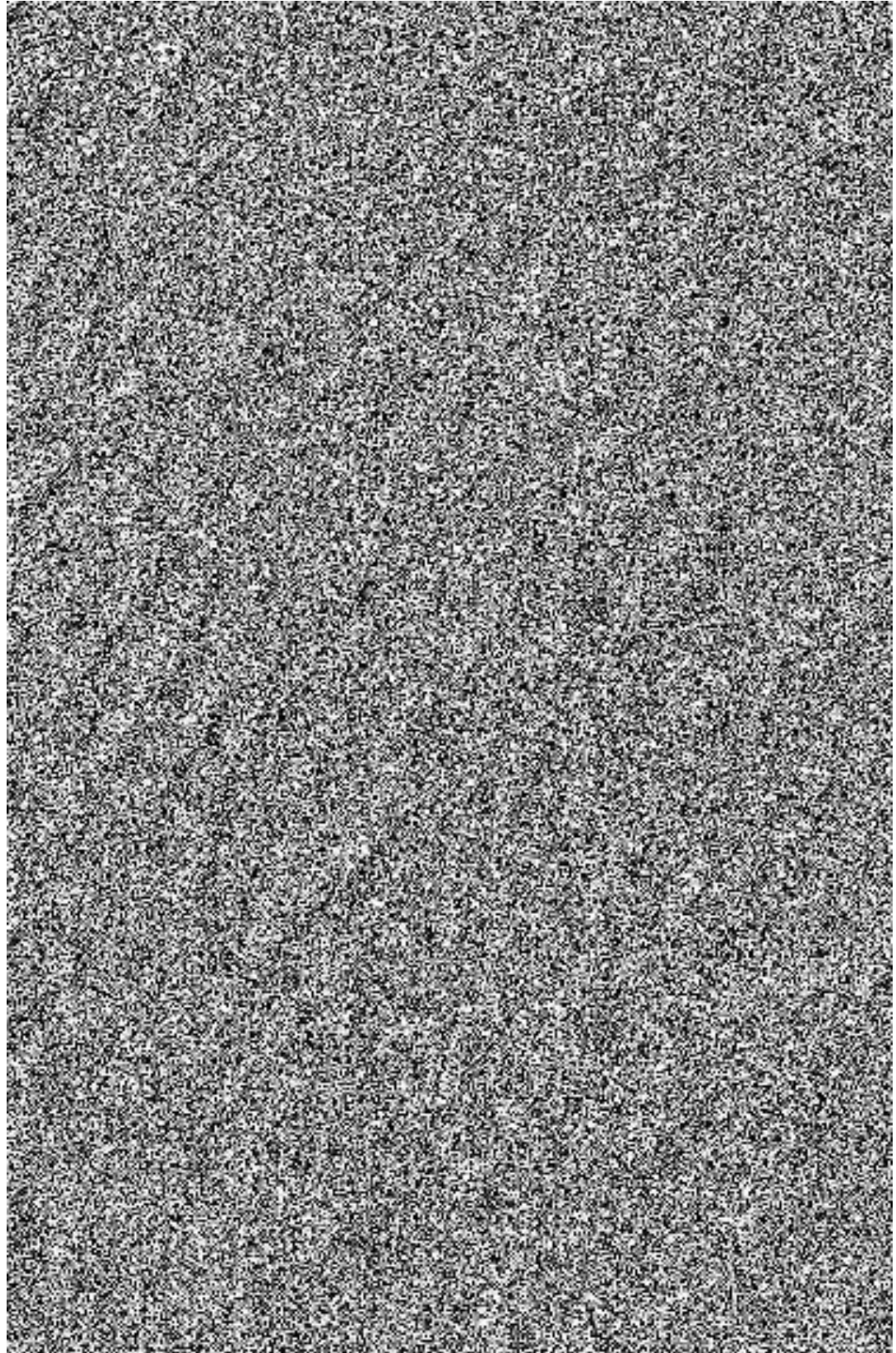
Hydrologic analysis data path



When you want this...



You get this:



Data and File Structures

Data are stored
as binary numbers

Bits are 0 or 1

Bytes are 8 bits

Data (e.g., raster cells)
are often references
as 1 byte, two byte,
etc.

Binary Columns

eights column	fours column	twos column	ones column
1	1	0	1

$$8 + 4 + 0 + 1 = 13$$

binary	decimal
00000001	1
00000010	2
00000011	3
00000100	4
00000101	5
00000110	6
00000111	7
00001000	8
00001001	9
00001010	10
00001011	11
....

Formats often specified in layer creation, e.g.

A raster may be assigned

float (real numbers) of a specified size, e.g., with 127 bits for the number and a bit to record if it is positive or negative,

or

Integer with up to 64 bits, with no bit for sign,

or

byte (eight bits), with no bit for sign

Each type has different limits on size and type of number that may be stored

You Need All Part

Example: ESRI Shapefiles

is a cluster of files,

wash_lc.shp - containing the coordinates

wash_lc.dbf - containing the attributes

wash_lc.shx - containing linkages, other info

wash_lc.prj - optional, containing projection information

wash_lc.sbn - an optional indexing file

Generally, all files are needed for useful data

If you mess with one file, you can “break” the data

Some formats all-inclusive, e.g., GeoPackages,
GeoDatabases

Data and File Structures

Compression

Reducing size – e.g., raster run-length coding

Raster

9	9	6	6	6	6	6	7
6	6	6	6	6	6	6	6
9	9	6	6	6	6	7	7
9	8	9	6	6	7	7	5

Run-length codes

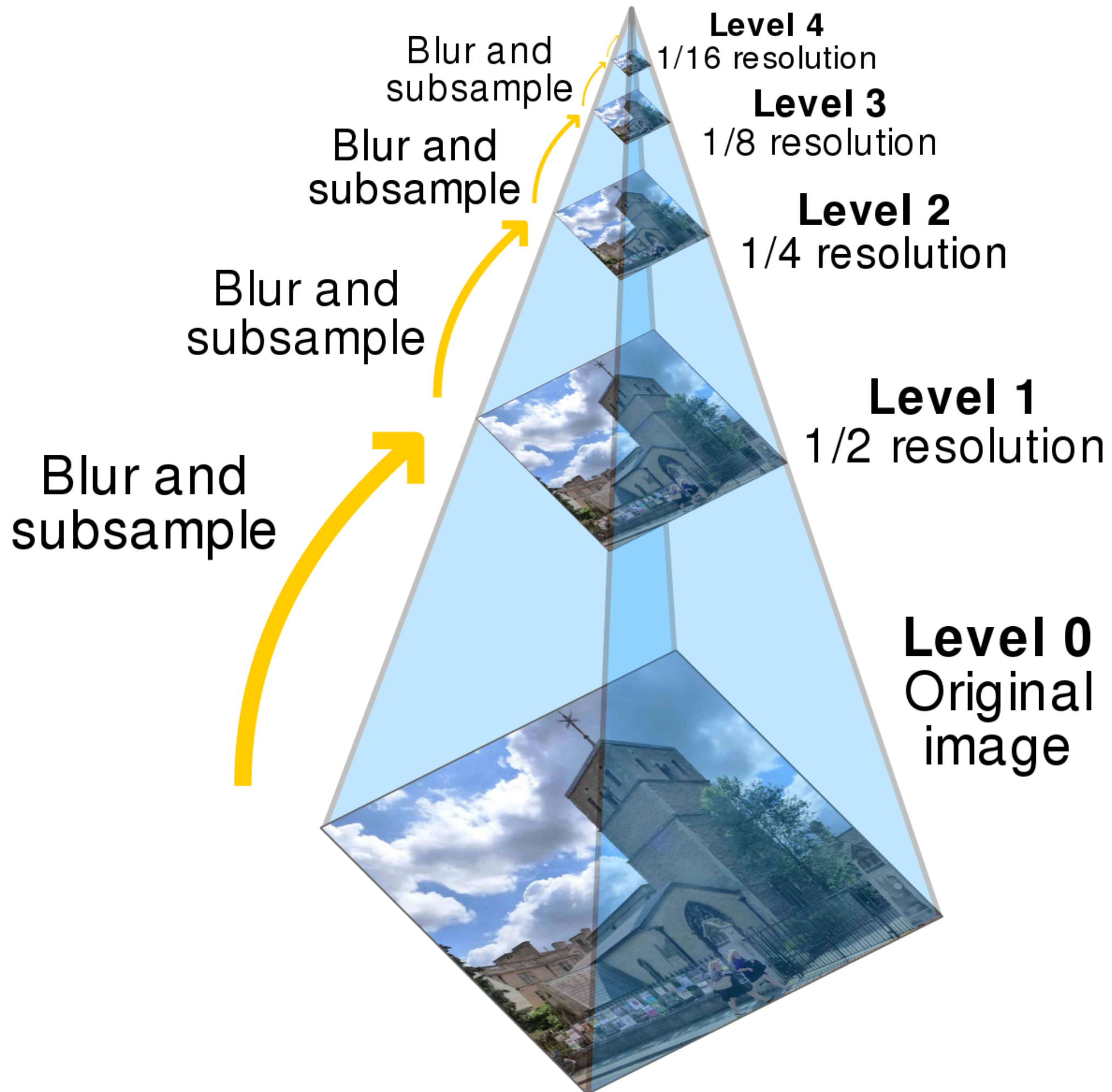
2:9, 5:6, 1:7

8:6

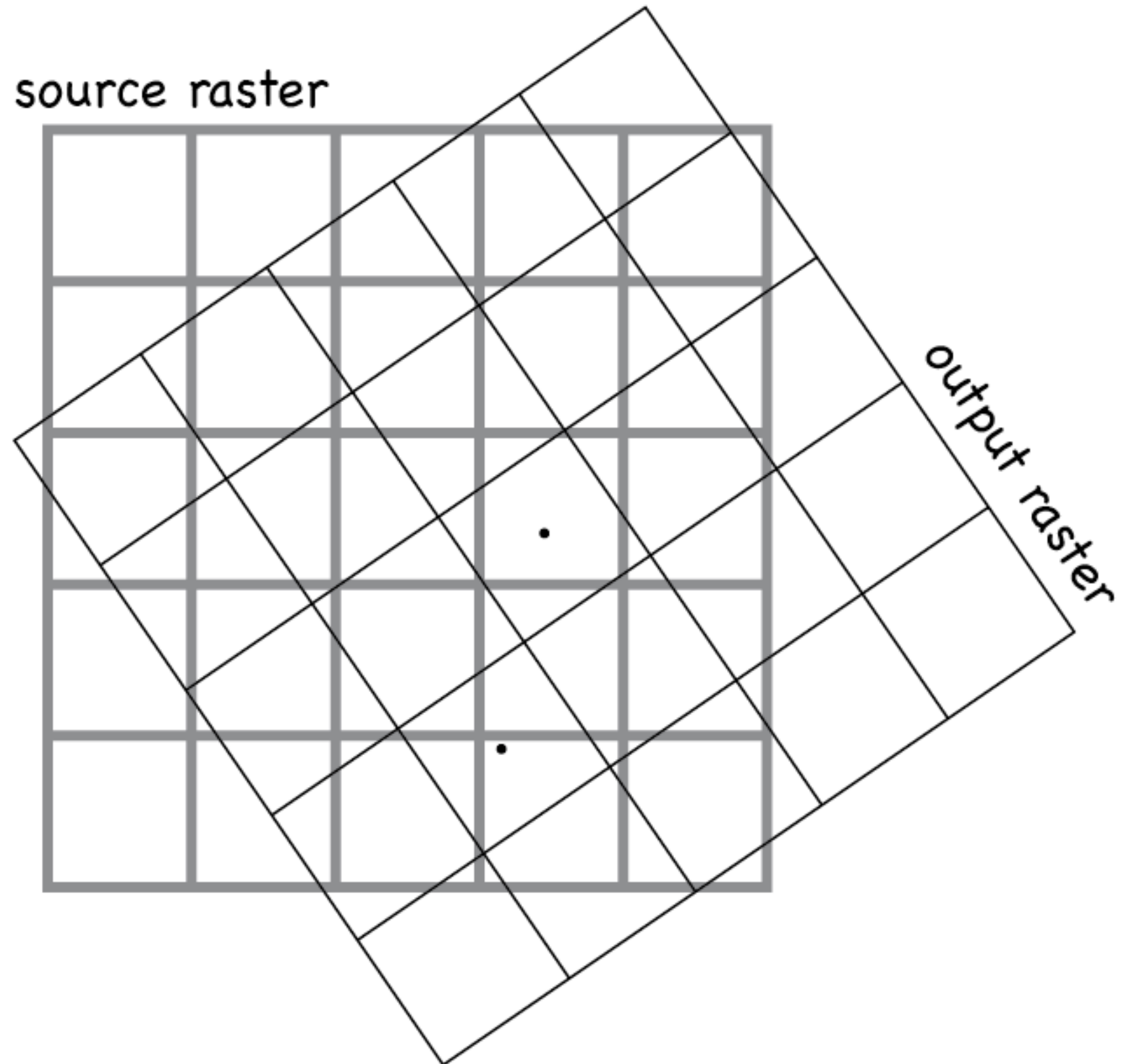
2:9, 4:6, 2:7

1:9, 1:8, 1:9, 2:6, 2:7, 1:5

Raster
Pyramids -
Resampling
at lower
resolution to
speed
display

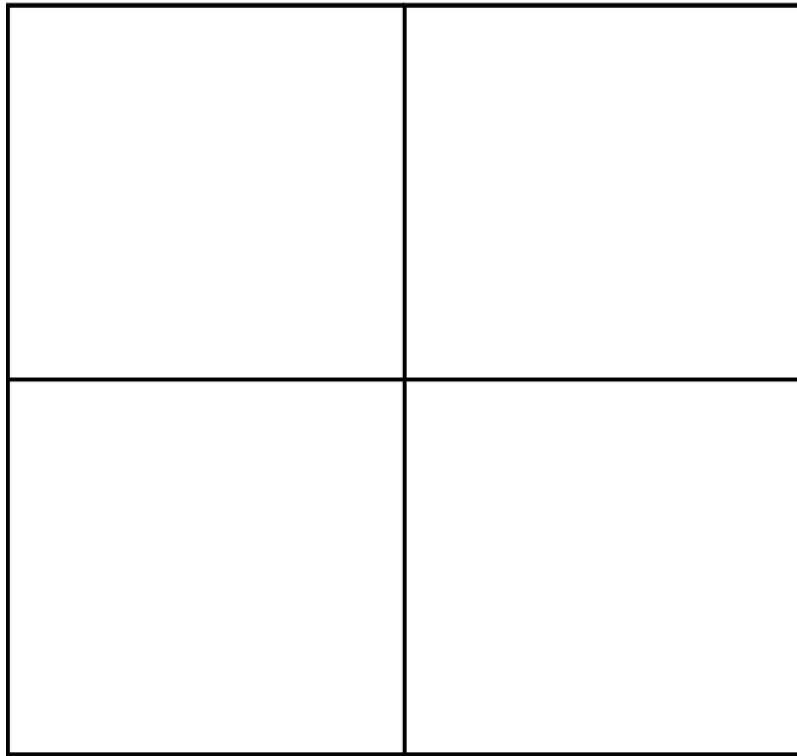


Raster Resampling - inter/intra-cell averaging



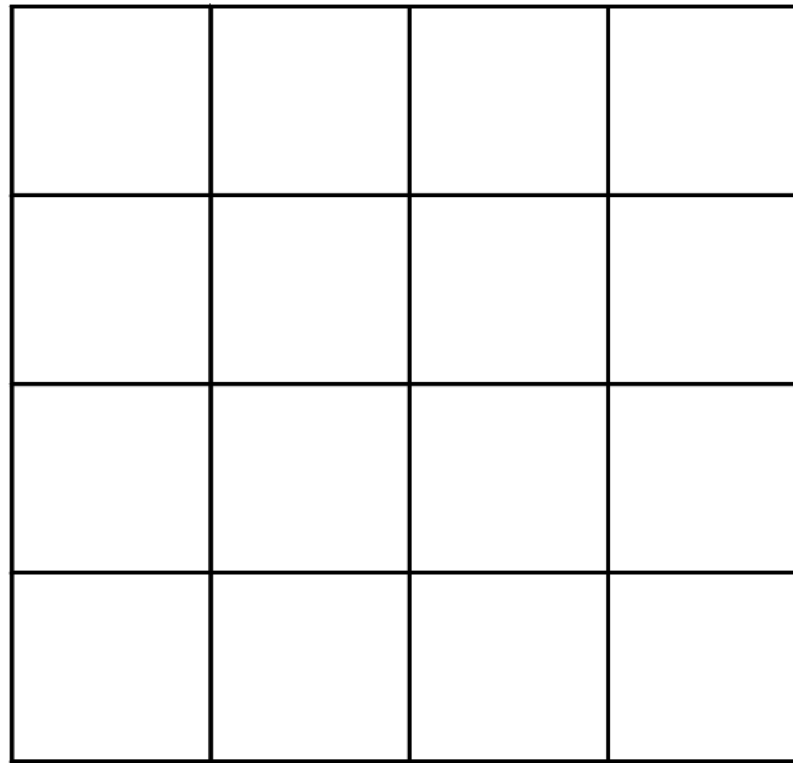
Raster resampling - our data have different grains

100 meter, 4 cells



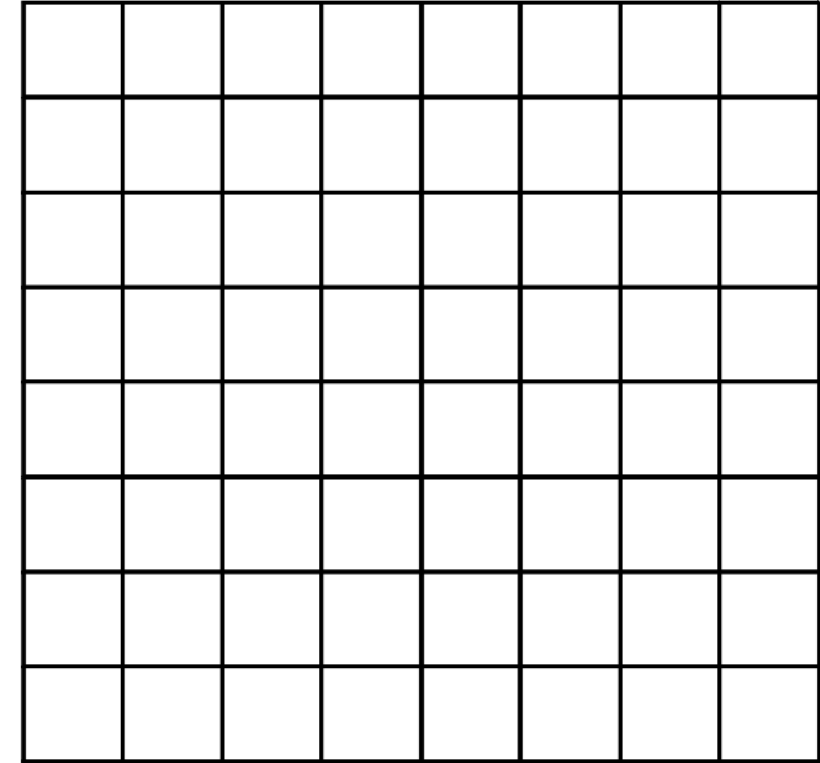
a)

50 meter, 16 cells



b)

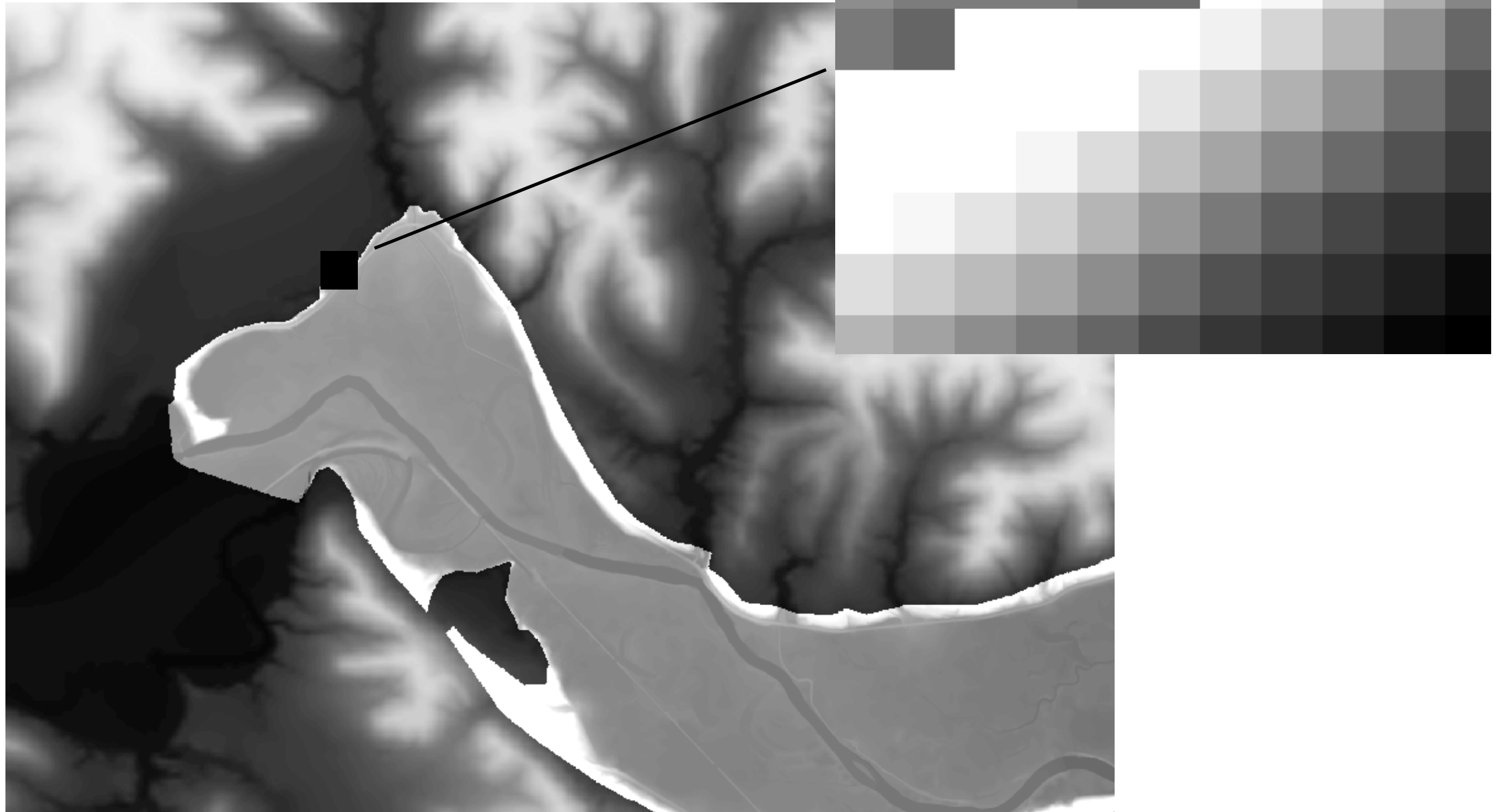
25 meter, 64 cells



c)

Raster Resampling -

We sometimes combine data from different resolutions

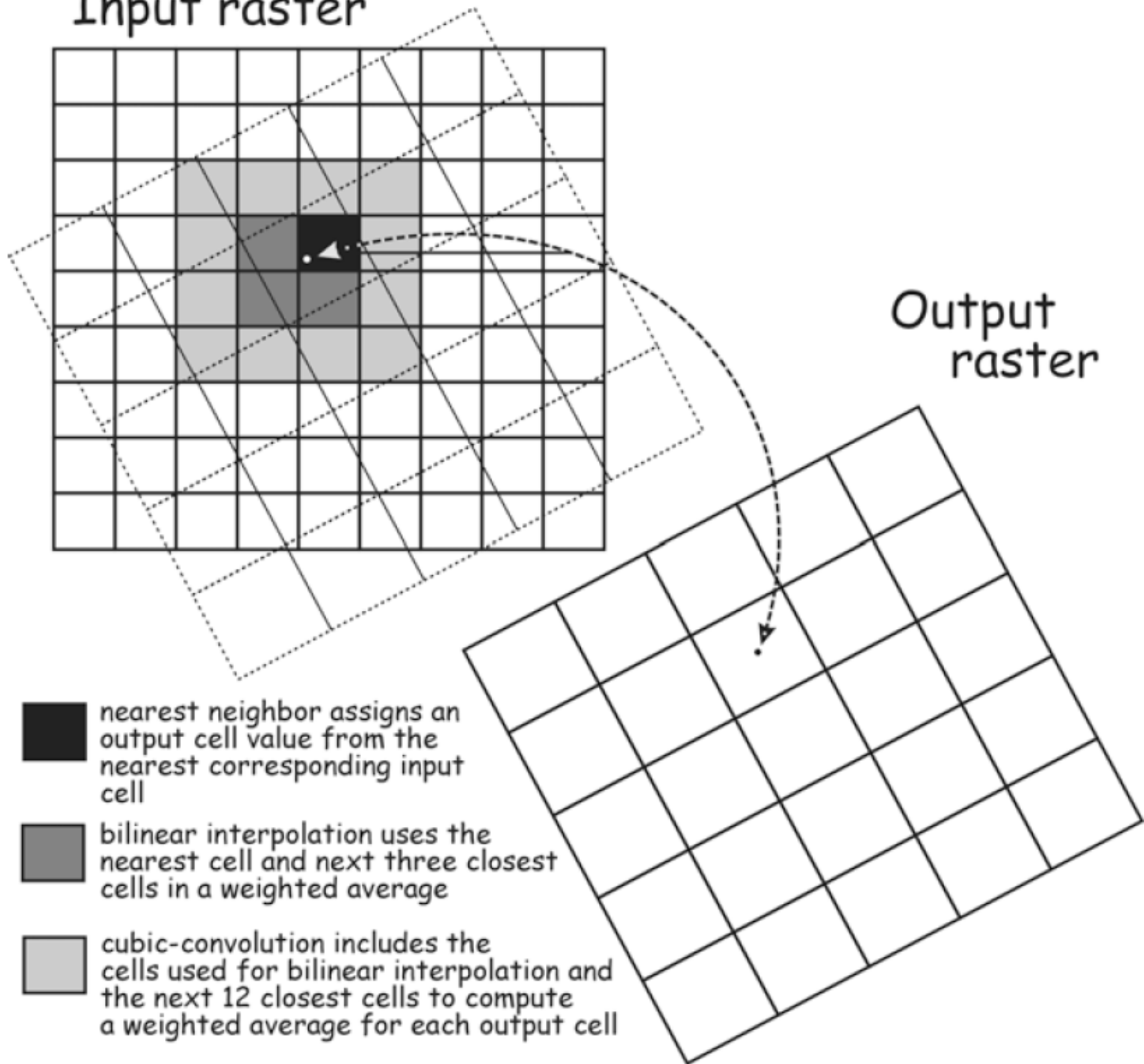


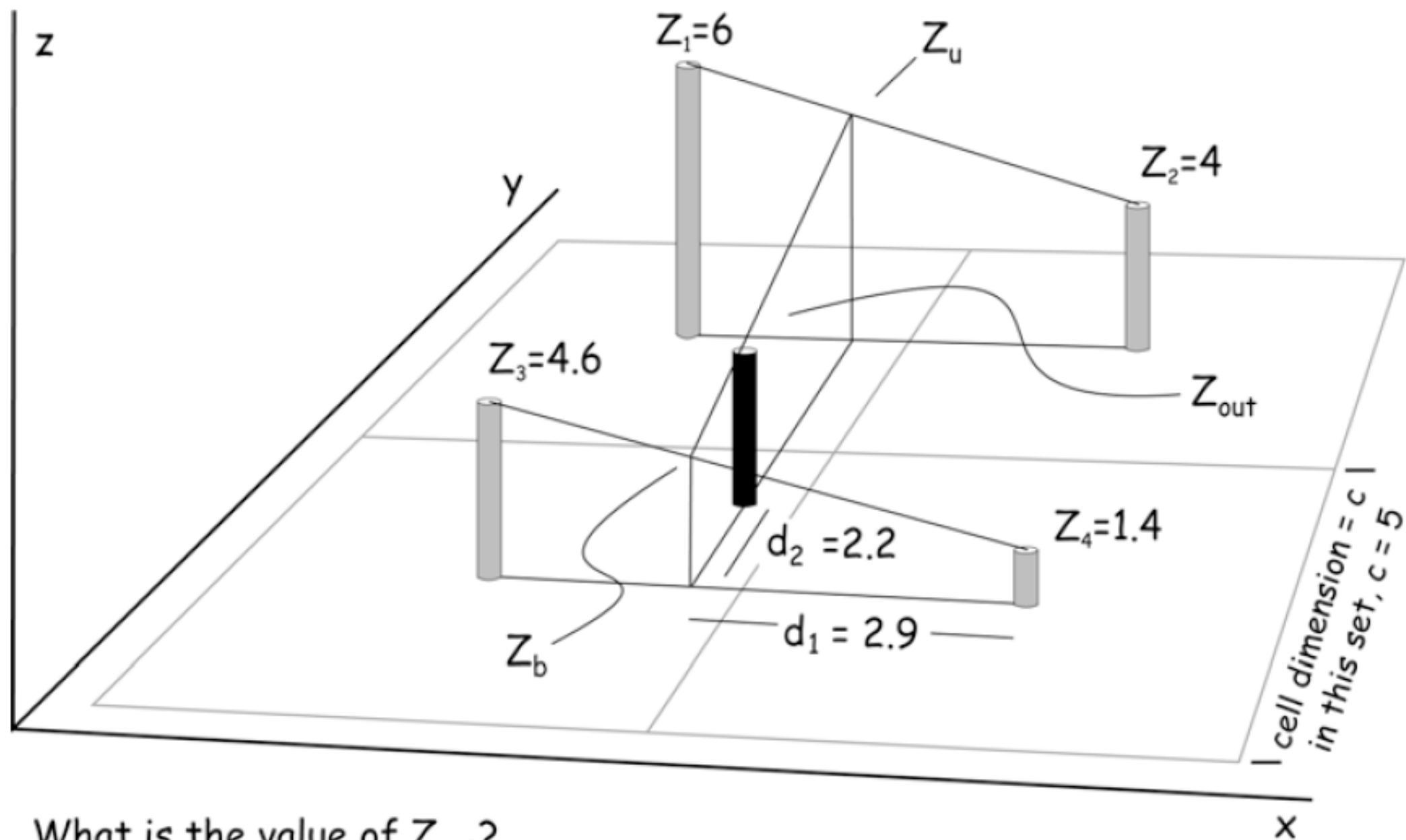
Does it make sense to resample nominal data?

Does it make sense to resample ordinal data?

Does it make sense to resample interval/ratio data?

Input raster





What is the value of Z_{out} ?

$$Z_b = Z_4 + \frac{(Z_3 - Z_4) * d_1}{c}$$

$$Z_b = 1.4 + \frac{(4.6 - 1.4) * 2.9}{5} = 3.26$$

$$Z_u = Z_2 + \frac{(Z_1 - Z_2) * d_1}{c}$$

$$Z_u = 4 + \frac{(6 - 4) * 2.9}{5} = 5.16$$

$$Z_{out} = Z_b + \frac{(Z_u - Z_b) * d_2}{c}$$

$$Z_{out} = 3.26 + \frac{(5.16 - 3.26) * 2.2}{5} = 4.1$$

Resampling Ambiguity

2	2	1	1
2	2	2	2
1	2	1	2
4	5	3	null

real....average?
→
integer....truncate?
null....average, null?

●	●
●	●

a	a	a	a
a	a	a	b
a	b	a	b
c	c	c	d

nominal....most common?



●	●
●	●