

tools & data

Stewart C Russell - @scruss

#maptime TO

2015-11-18



happy GIS Day!

“be here now”

structure

1. Care & Feeding of Your Data

2. The Right Tools

3. A Tiny Little Example

*unless you collect it or
commission it,
data is in the
wrong format*



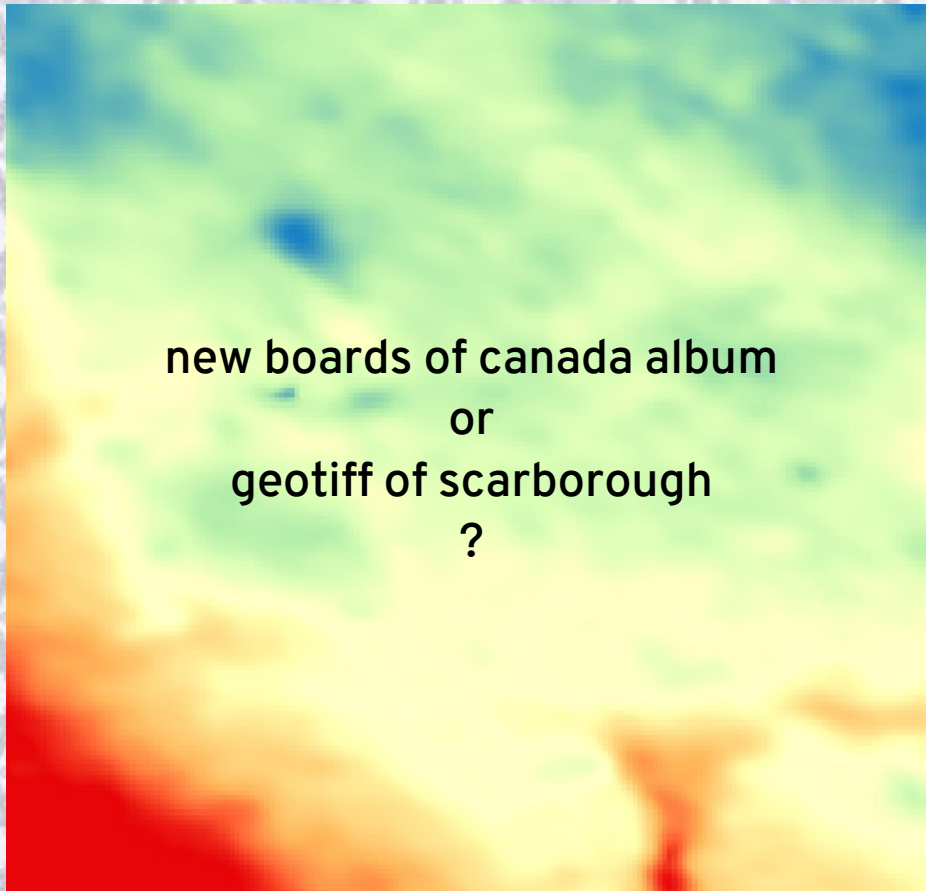
*be kind to your
metadata*

the undead Shapefile

```
$ ls -l  
Danube_GSGS_Grid.*  
Danube_GSGS_Grid.cpg  
Danube_GSGS_Grid.dbf  
Danube_GSGS_Grid.prj  
Danube_GSGS_Grid.qpj  
Danube_GSGS_Grid.shp  
Danube_GSGS_Grid.shx
```

- it's a file made up of lots of files
- based on a mid-80s database format
- too many limitations to name
- *will! not! die!*

the colossal GeoTIFF



- **extends the already baroque TIFF image format**
- **sometimes has aux files**
- **easily broken with graphics tools**
- **likely the biggest files you'll ever see**

check your licences

can I even publish this?

will it blend?

- **not all open data has compatible licences**
- **even Ontario municipalities aren't compatible**
- **spend all of your project budget on IP lawyers**

heid nip 1: map datum

Are we at

UTM 17T 629490 4833953

or

UTM 17T 629477 4833731 ?

⇒ “yes ...”

heid nip 2: projection

*who'd use
anything but
UTM?*

vs

*who'd use
anything but
WGS-84?*

heid nip 3: charset

I â™¥ UTF-8

vs

I ♥ UTF-8

be kind to everyone

(not just metadata)

**If you have wrangled a difficult public data set
and won, *please* write up how you did it**

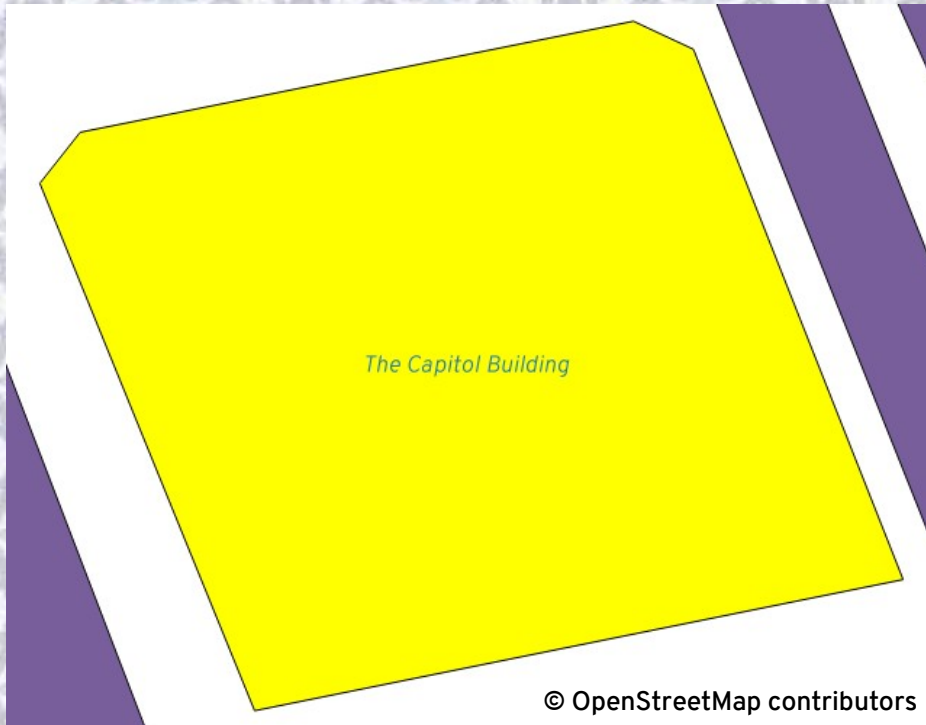


Photograph by Greg Stiller

The Studley Tool Chest

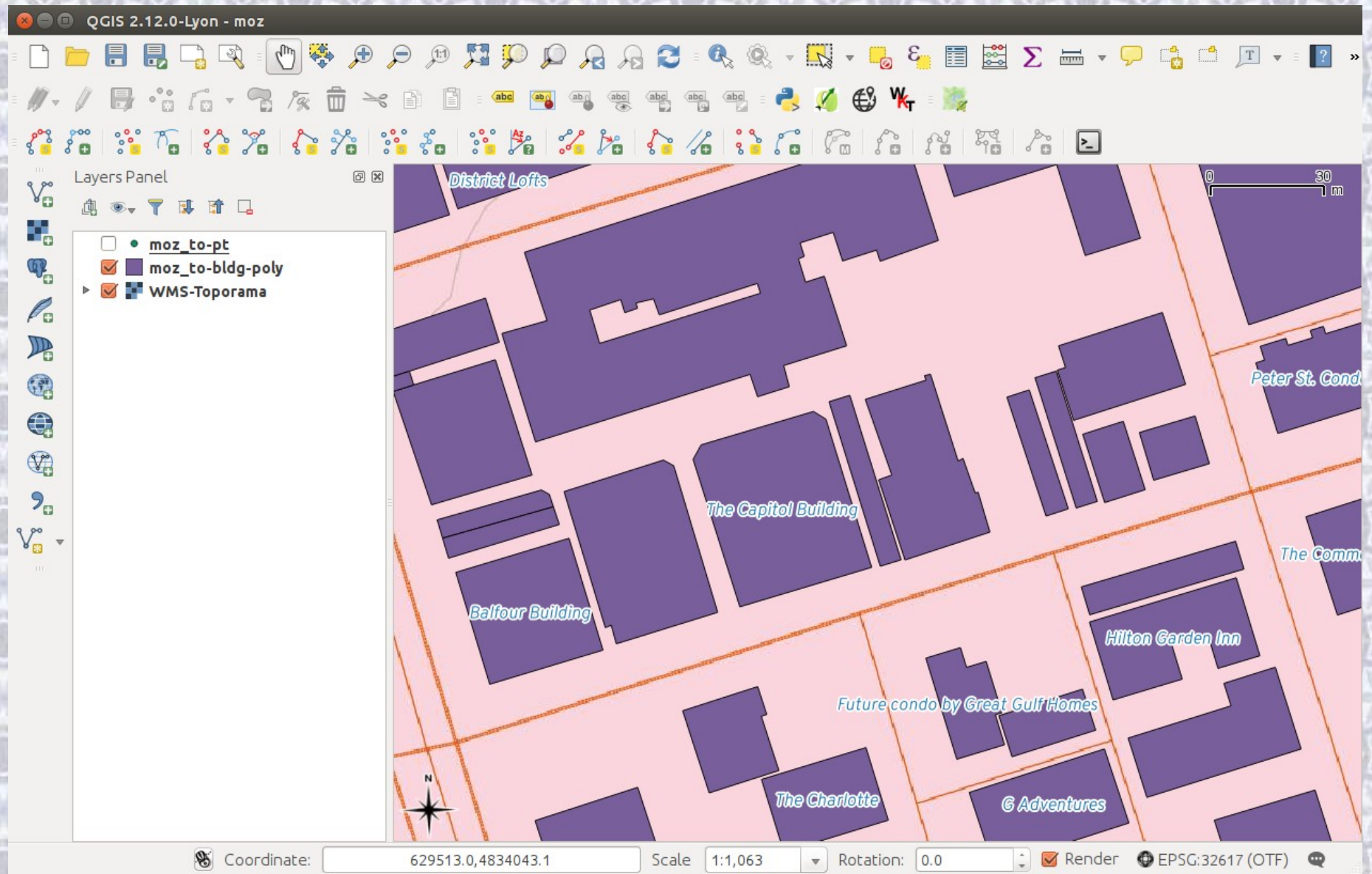
Fine
WoodWorking.com

visual or batch?



```
{ "crs" : { "type" : "name", "properties" : { "name" : "urn:ogc:def:crs:OGC:1.3:CRS84" } }, "type" : "FeatureCollection", "features" : [ { "geometry" : { "coordinates" : [ [ [ -79.39423, 43.64751 ], [ -79.39409, 43.64715 ], [ -79.39452, 43.64707 ], [ -79.39466, 43.64742 ], [ -79.39463, 43.64745 ], [ -79.39427, 43.64753 ], [ -79.39423, 43.64751 ] ] ], "type" : "Polygon" }, "type" : "Feature", "properties" : { "name" : "The Capitol Building" } } ] }
```


qgis



gdal and ogr

GDAL reads *raster* data and:

- **scales, warps, re-projects, clips, tiles, sharpens, merges, convolutes and munges[†]**
- **saves in pretty much any *raster* format you need**

OGR reads *vector* data and:

- **scales, translates, warps, queries, simplifies, re-projects, clips, and hits with a big stick[†]**
- **saves in pretty much any *vector* format you need**

[†]: not a term officially recognized by the OGC.

even lower level tools

- **proj.4 – the reprojection standard**
- **fiona – python vector data import library**
- **shapely – python geoprocessing library**
- **openrefine – unruly data tamer**
- **unfolding – beautiful interactive maps for Processing 2.n**



example:

war diaries

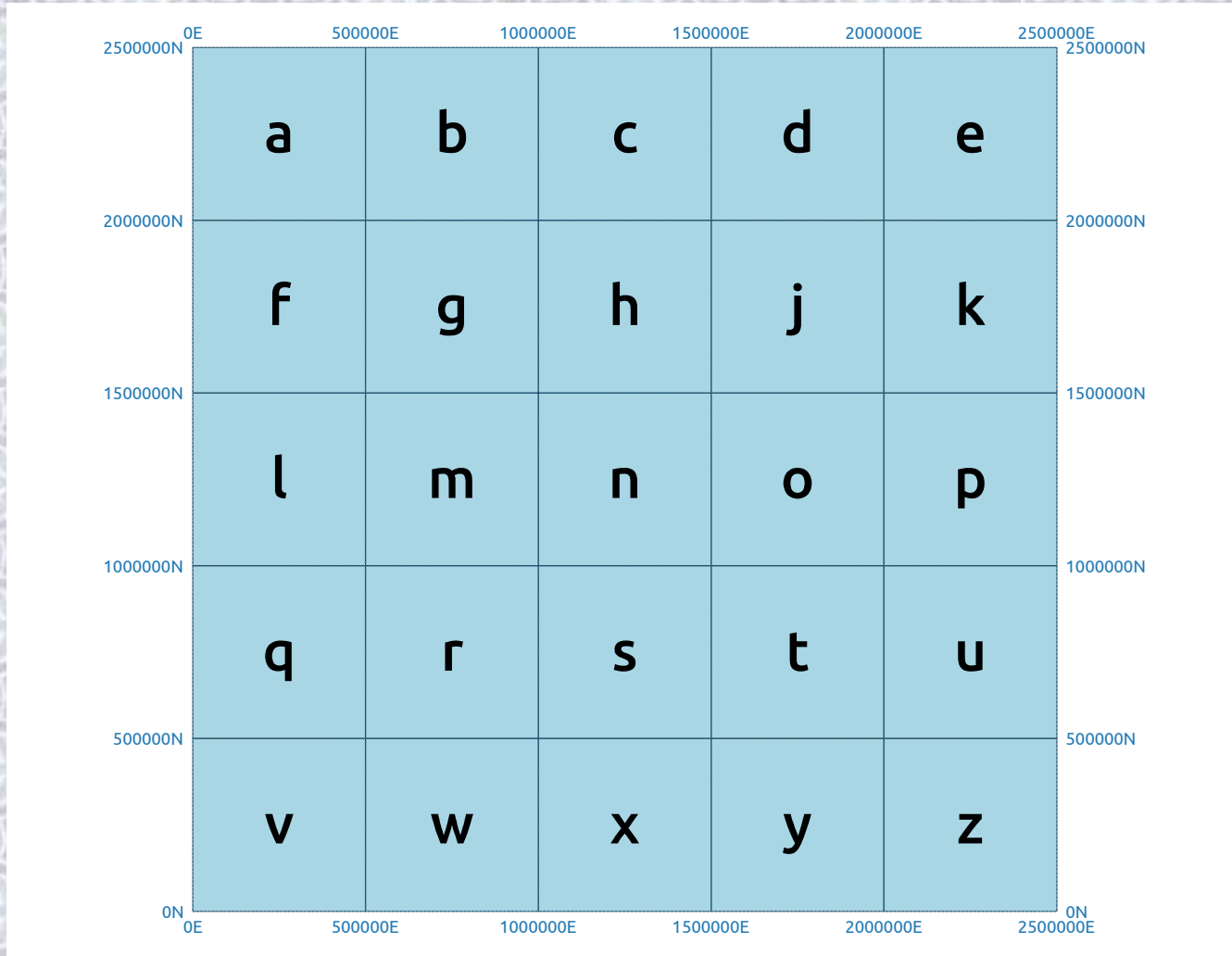
the data

20 Dec 43

The regiment, supported by 'C' Squadron, The Three Rivers Tank Regt, advanced under a barrage on a two Coy, two Troop front at 1200 hrs today. The start line was astride the ORTONA road 100 yards forward of the crossroads at **MR 322142. After very heavy fighting the regiment reorganized as follows: 'B' Coy the general area **MR 333164**, where liaison was established with 'C' Coy of the Seaforth H of C who had fought their way up the coast road; 'D' Coy to the northwest of 'B' Coy astride the main road; 'C' Coy at **MR 328158**; 'A' Coy and Bn Hq at **MR 332160**. ...**

– *War Diary, The Loyal Edmonton Regiment, vol. 50 (Dec 1943)*

modified british grid



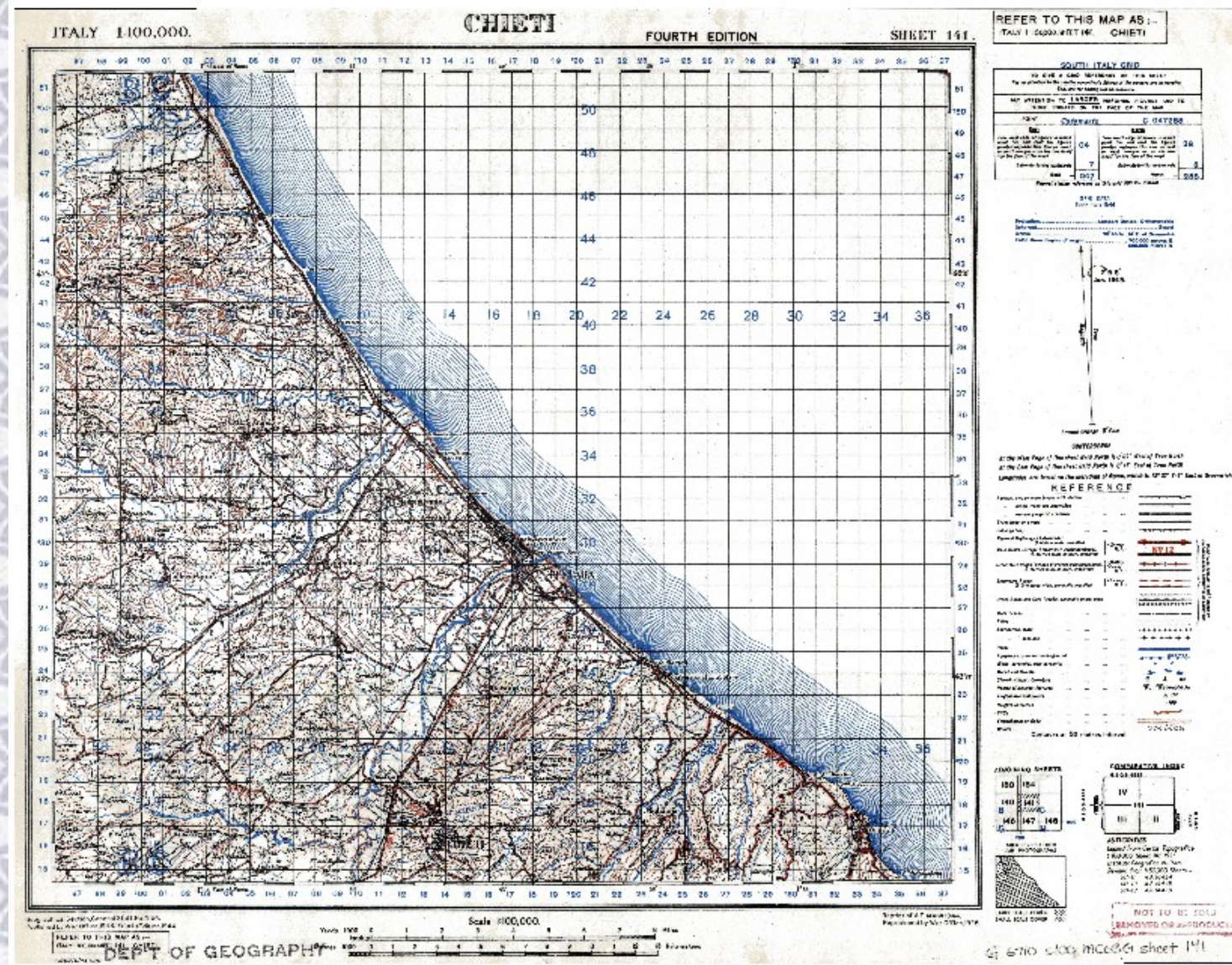
decoding the grid

The grid is a 5x5 matrix of cells, each containing a letter. The horizontal axis is labeled 'OE' (Easting) with values 0E, 500000E, 1000000E, 1500000E, 2000000E, and 2500000E. The vertical axis is labeled 'ON' (Northing) with values 0N, 500000N, 1000000N, 1500000N, 2000000N, and 2500000N. The letters are arranged as follows:

OE \ ON	0N	500000N	1000000N	1500000N	2000000N	2500000N
0E						
500000E		v	w	x	y	z
1000000E		q	r	s	t	u
1500000E		l	m	n	o	p
2000000E		f	g	h	j	k
2500000E		a	b	c	d	e

```
result = SEARCH(letter,  
"VWXYZQRSTULMNOPFGHJKABCDE") -  
1  
easting = result MOD 5  
northing = INT(result / 5)
```


the map



the map details

SOUTH ITALY GRID			
TO GIVE A GRID REFERENCE ON THIS SHEET <i>Pay no attention to the <u>smaller</u> co-ordinate figures at the corners and in margins. They are for finding full co-ordinates.</i>			
PAY ATTENTION TO <u>LARGER</u> MARGINAL FIGURES AND TO THOSE PRINTED ON THE FACE OF THE MAP.			
POINT		C 047288	
Casamarte			
<u>East</u> Take west edge of square in which point lies and read the figures printed opposite this line on north or south margin or on the line itself (on the face of the map) Estimate tenths eastwards East ...	04 7 047	<u>North</u> Take south edge of square in which point lies and read the figures printed opposite this line on east or west margin or on the line itself (on the face of the map) Estimate tenths northwards North ...	28 8 288
Nearest similar reference on this grid 500 Km. distant.			
GRID DATA South Italy Grid			
Projection..... Lambert Conical Orthomorphic			
Spheroid..... Bessel			
Origin..... 39° 30' N., 14° E. of Greenwich			
False Co-ordinates of origin..... 700,000 metres E. 600,000 metres N.			

georeferencing ...

Georeferencer - chieti_ortona-italy_south.tiff

Geographical Section, General Staff No. 4164.
Published by War Office, 1943. Fourth Edition, 1944

REFER TO THIS MAP AS:—
ITALY 1:100,000 SHEET 141 CHIETI

GCP table

Visible	ID	Source X	Source Y	Dest. X	Dest. Y	dX (pixels)	dY (pixels)	Residual (pixels)
<input type="checkbox"/>	2	4215.57	-414.967	730000	951000	0	0	0
<input checked="" type="checkbox"/>	3	4923.01	-412.663	736000	951000	0	0	0
<input type="checkbox"/>	4	5042.27	-4653.47	737000	915000	0	0	0
<input checked="" type="checkbox"/>	5	4214.14	-4650.58	730000	915000	0	0	0
<input checked="" type="checkbox"/>	6	328.751	-4646.84	697000	915000	0	0	0
<input checked="" type="checkbox"/>		4025.78	-4646.84	702000	915000	0	0	0

Transform: Not set 1199,-5005 None

the magic of proj.4 ...

Custom Coordinate Reference System Definition

▼ Define

You can define your own custom Coordinate Reference System (CRS) here. The definition must conform to the proj4 format for specifying a CRS.

Name	Parameters
Danube ...	+proj=lcc+lat_0=45.9+lat_1=45.9+lon_0=29+k_0=0.99906+x_0=15000...
* Gener...	+proj=lcc+lat_0=39.5+lat_1=39.5+lon_0=14+k_0=0.99906+x_0=70000...
* Gener...	+proj=lcc+lat_1=39.5+lat_0=39.5+lon_0=14+k_0=0.999059999999999...
* Gener...	+proj=lcc+lat_1=45.9+lat_0=45.9+lon_0=29+k_0=0.999059999999999...
* Gener...	+proj=longlat+ellps=sphere+no_defs
* Gener...	+proj=merc+ellps=sphere+units=m
* Gener...	+proj=merc+lon_0=0+k=1+x_0=0+y_0=0+a=6370997+b=6370997+uni...
* Gener...	+proj=merc+lon_0=0+k=1+x_0=0+y_0=0+a=6378137+b=6378137+uni...

+

−

📄

Name

_0=14+k_0=0.99906+x_0=700000+y_0=600000+ellps=bessel+units=m+no_defs)

Parameters

+proj=lcc+lat_0=39.5+lat_1=39.5+lon_0=14+k_0=0.99906+x_0=700000
+y_0=600000+ellps=bessel+units=m+no_defs

▼ Test

Use the text boxes below to test the CRS definition you are creating. Enter a coordinate where both the lat/long and the transformed result are known (for example by reading off a map). Then press the calculate button to see if the CRS definition you are creating is accurate.

Geographic / WGS84

Destination CRS

North

42.33048

914,200.1678

East

14.39061

732,199.7882

Calculate

Help

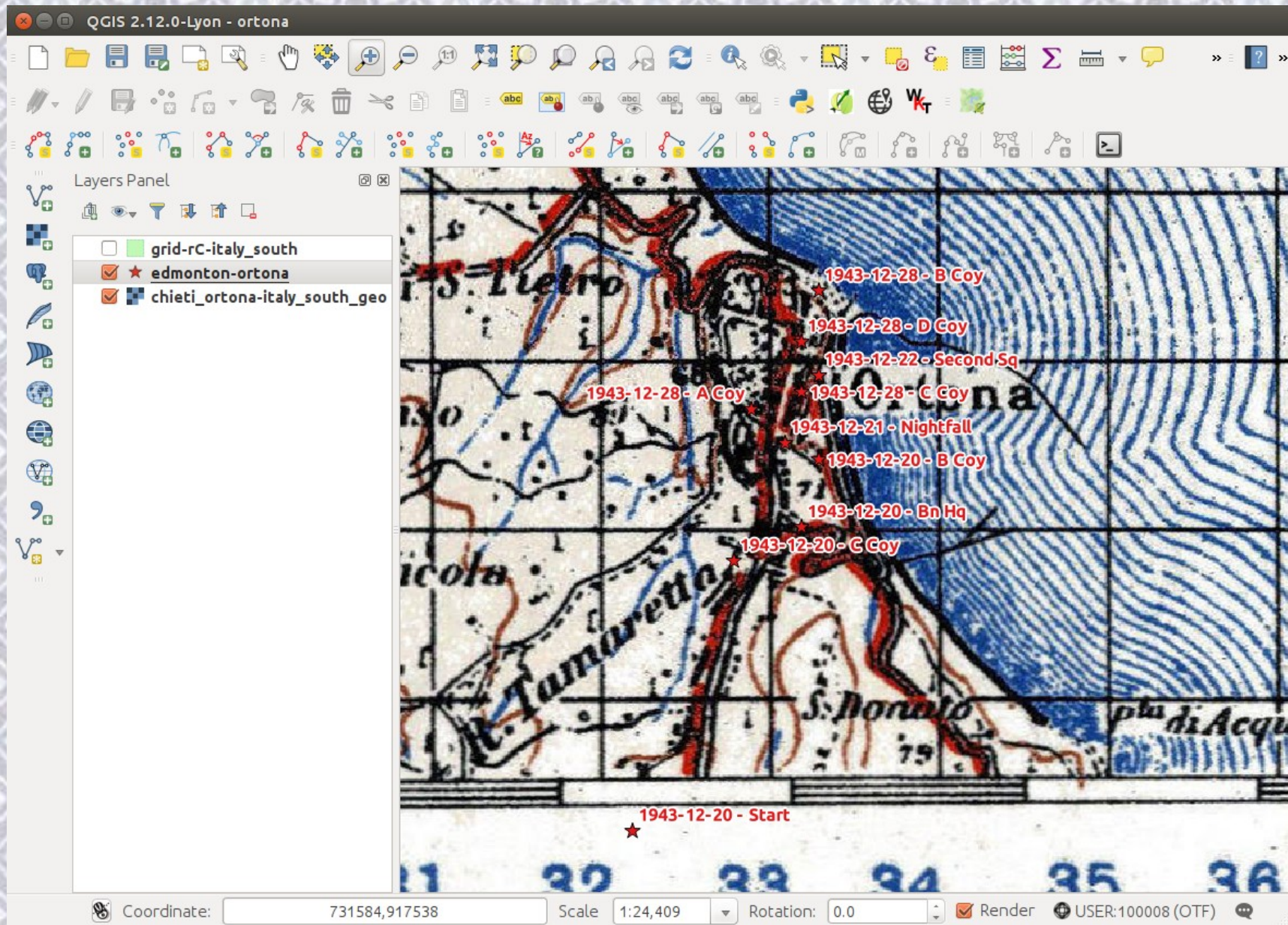
Cancel

OK

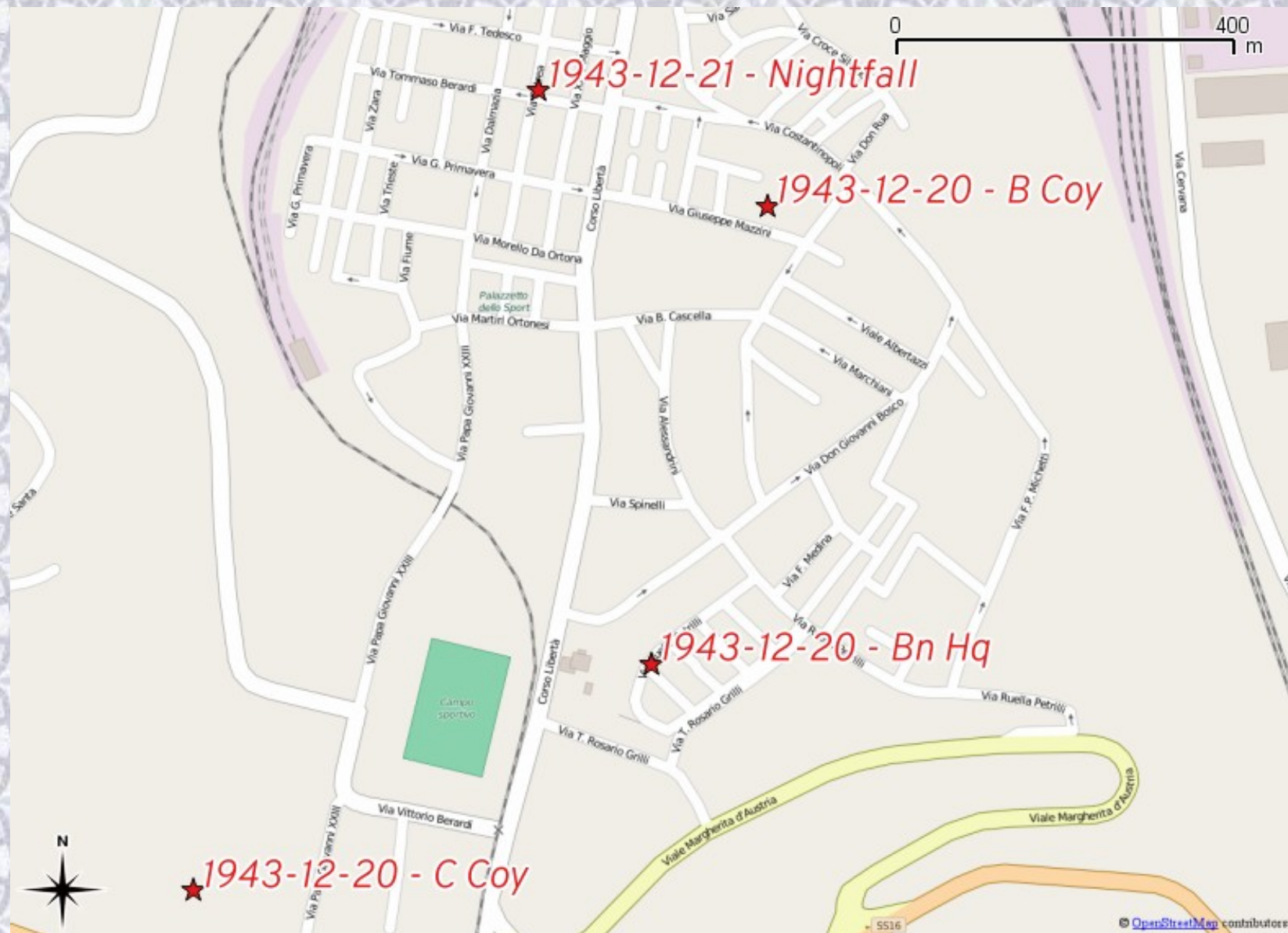
the data, transformed

MR	X	Y	Label
322142	732200	914200	1943-12-20 - Start
333164	733300	916400	1943-12-20 - B Coy
328158	732800	915800	1943-12-20 - C Coy
332160	733200	916000	1943-12-20 - Bn Hq

in historical context



in modern context



credits

- **YOU** – for showing up
- OpenStreetMap
- McMaster University Library
- The Taunton Press, Inc / finewoodworking.com
- The Loyal Edmonton Regiment Museum
- Thierry Arsicaud, echodelta.net
- GIS StackExchange
- Malcolm Hamilton
- All the GIS developers who can do hard sums so we don't have to