JERICO summerschool
Friday june 20th 2014

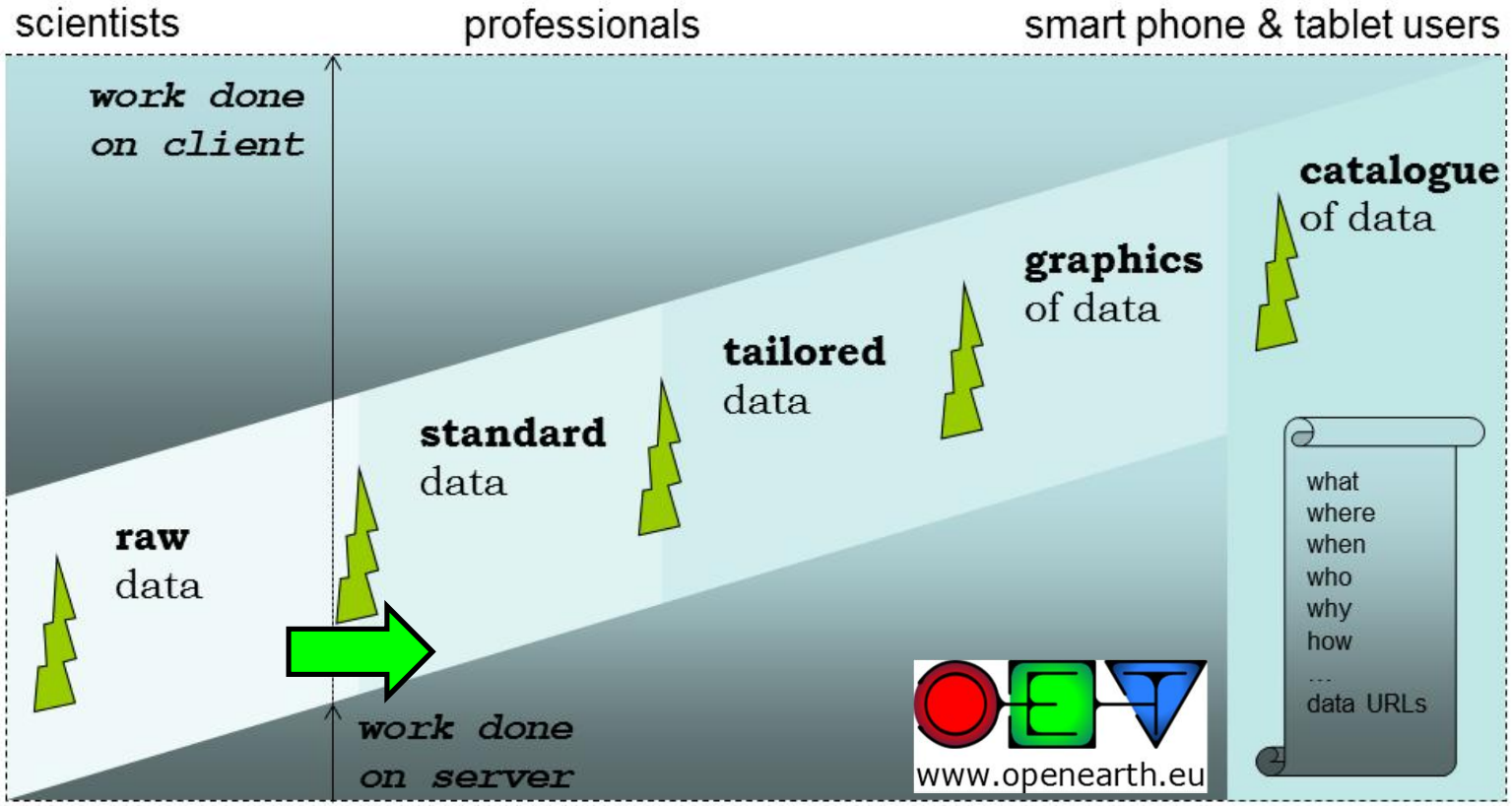
OpenEarth 4 JERICO

Gerben de Boer (middle →)

3TU.Datacentrum

5 step geospatial standards: only web services

inspired by 5-stardata of Tim-Berners Lee



The step from raw to standard is the hardest step: it is completely manual.
This is called ETL by database professionals: Extract, Transform, Load.
Once if is standard, the rest is only configuration (can still be time consuming)

Raw data + scripts = standard data

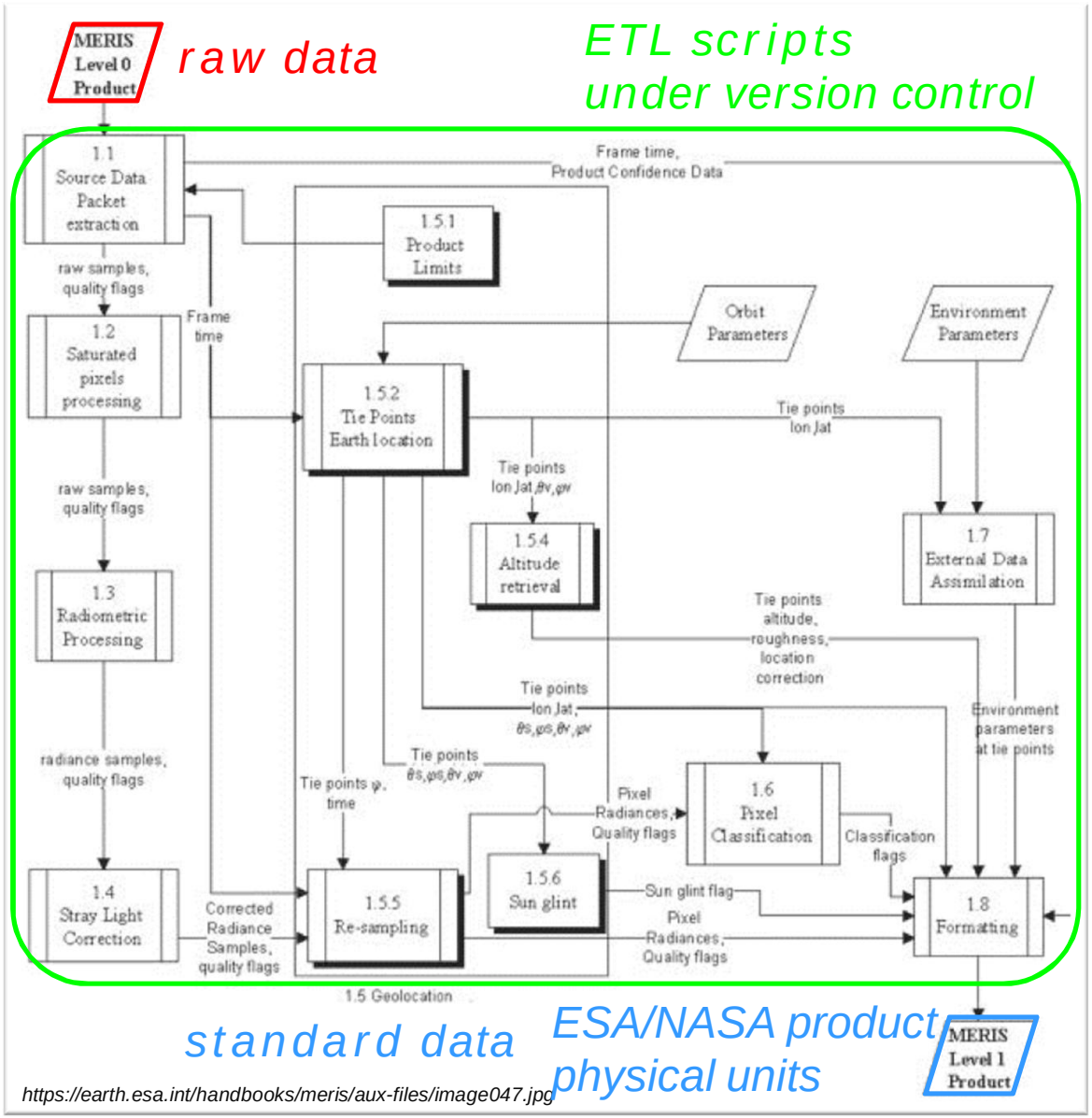
Compare with old-style and new-style Asian restaurants

raw data

scripts under version control

standard data

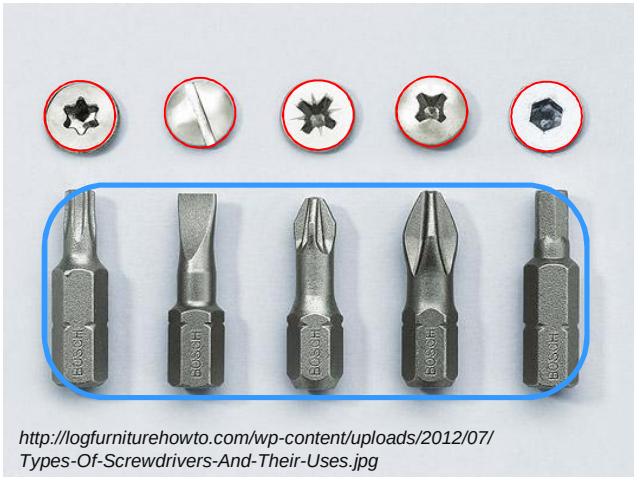
Satellites: L0 (raw) + script = L1,L2,L3,L4



https://earth.esa.int/handbooks/meris/aux-files/image047.jpg

data product =

raw data: volts counts → *+ software*



http://logfurniturehowto.com/wp-content/uploads/2012/07/Types-Of-Screwdrivers-And-Their-Uses.jpg

In situ should behave as satellite data

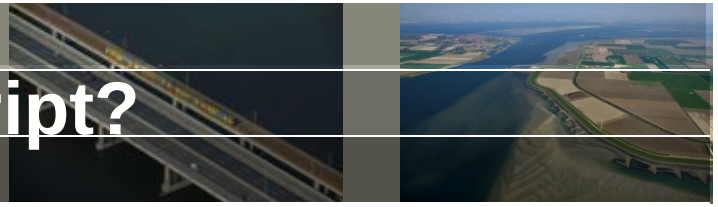
The peanut butter principle

- Store all raw data
- Make automatic processing routines
  - Store routines
  - Keep on updating all routines
- At regular intervals re-apply tools to historic data
  - Data should always have a version number, or not be used
  - Just like a jar of peanut butter always has an expiration date

http://www.food-info.net/nl/national/ww-pindakaas_clip_image001.jpg

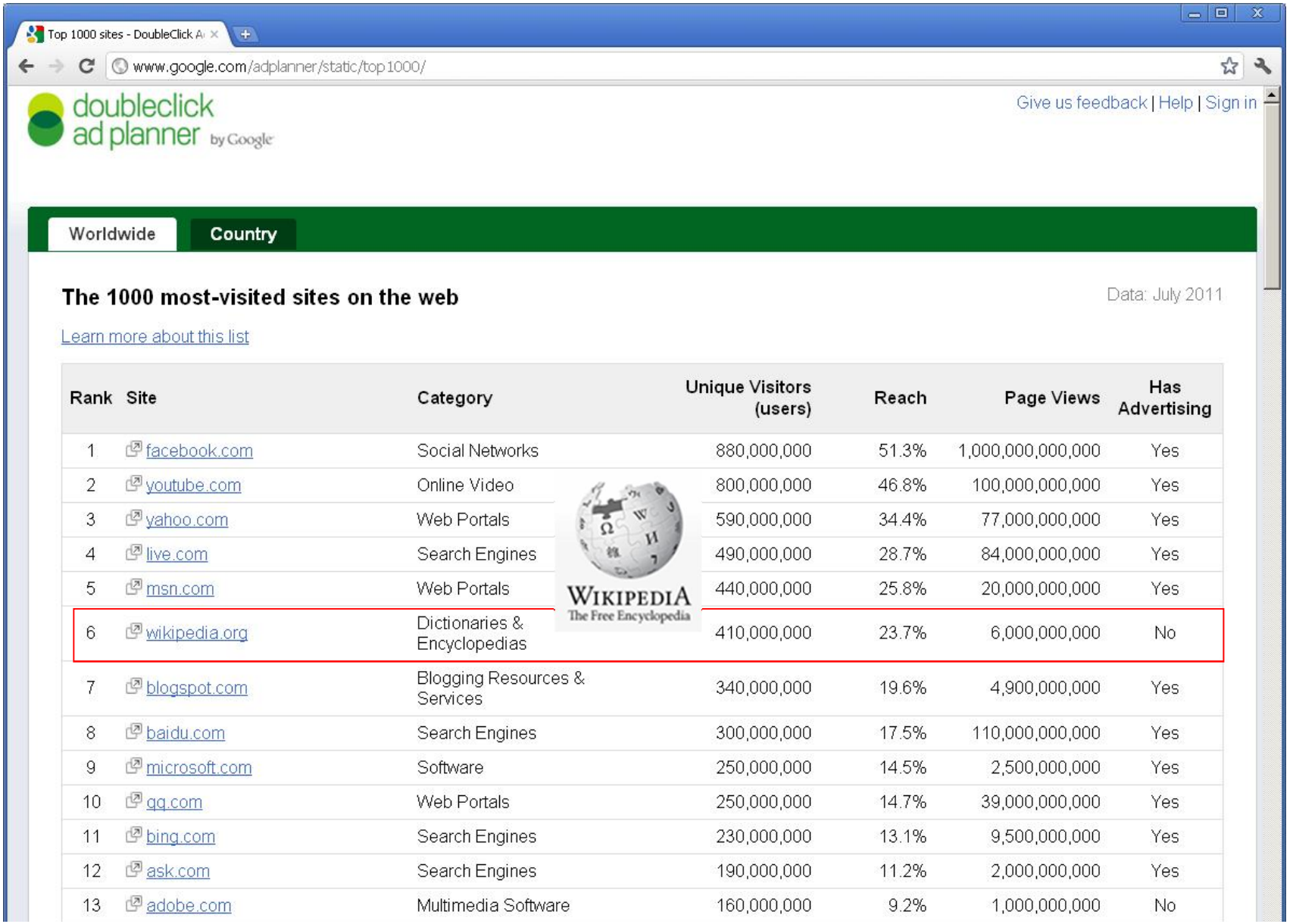
http://www.knups.nl/leuk_plaatje/56/Helaas_pindakaas.html

How to store raw data and script?



- Professional software is already stored in SubVersion
  (or alikes: newer: GIT, mercurial, older: vss, csv)
- Why not add the *data* to this version control repository too?
- Your raw data gets a url:
  - Image ftp with version control
  - Dropbox
- The version control system already exists
- Version control also allows others to help you edit your data (i.e. edit your data processing)
- Like wikipedia …

Communities: for science we can: 400 million



Top 1000 sites - DoubleClick A

www.google.com/adplanner/static/top1000/

doubleclick ad planner by Google

Give us feedback | Help | Sign in

Worldwide | Country

The 1000 most-visited sites on the web

Data: July 2011

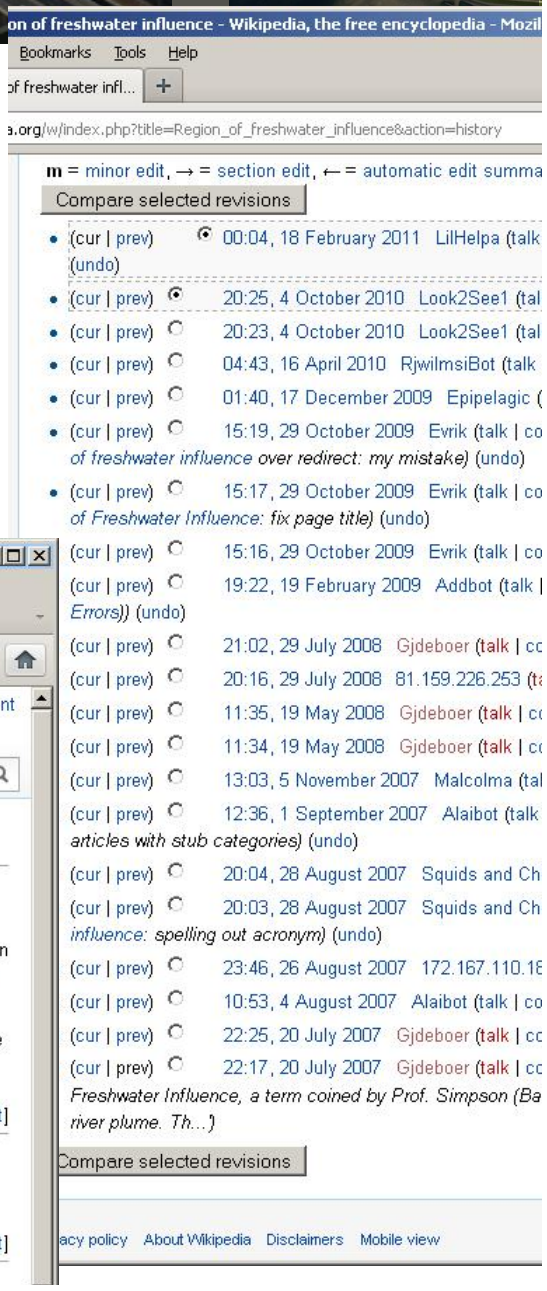
Learn more about this list

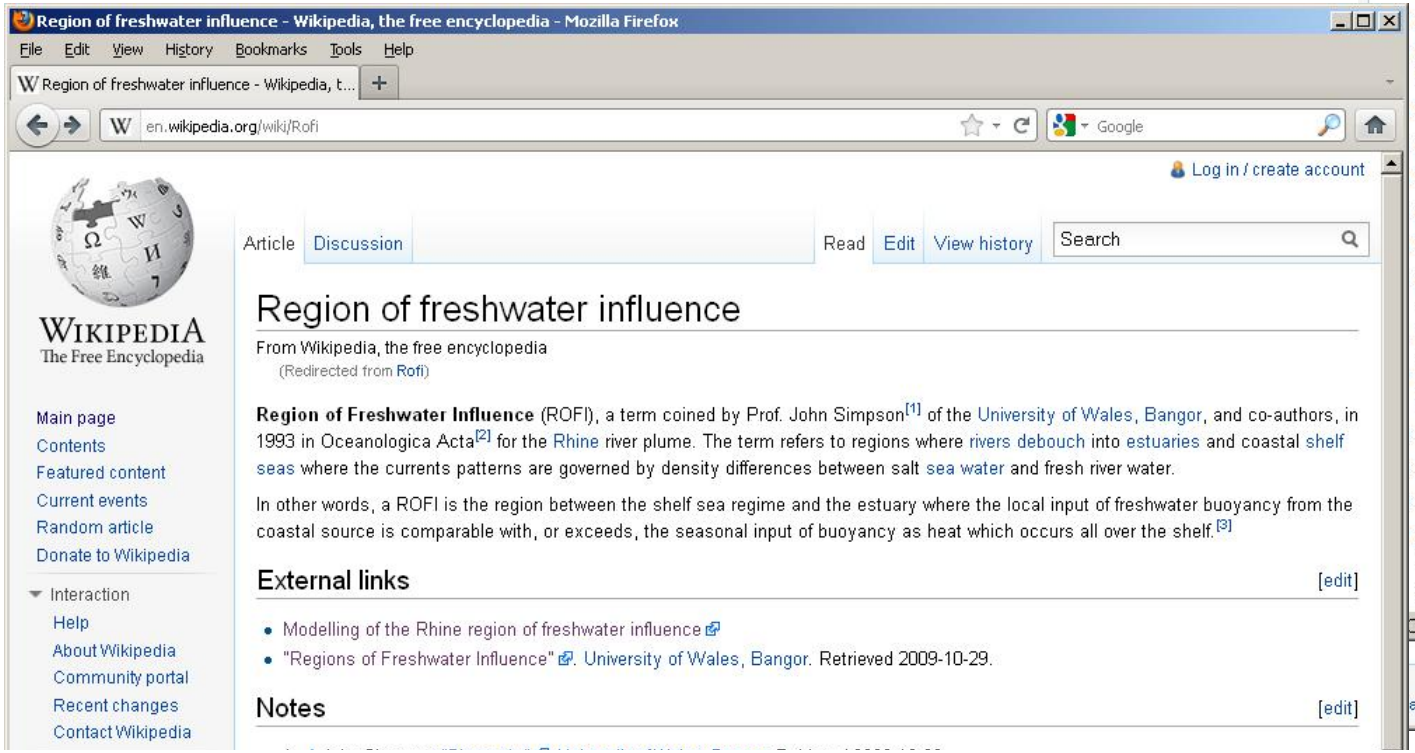
Rank	Site	Category	Unique Visitors (users)	Reach	Page Views	Has Advertising
1	facebook.com	Social Networks	880,000,000	51.3%	1,000,000,000,000	Yes
2	youtube.com	Online Video	800,000,000	46.8%	100,000,000,000	Yes
3	yahoo.com	Web Portals	590,000,000	34.4%	77,000,000,000	Yes
4	live.com	Search Engines	490,000,000	28.7%	84,000,000,000	Yes
5	msn.com	Web Portals	440,000,000	25.8%	20,000,000,000	Yes
6	wikipedia.org	Dictionaries & Encyclopedias	410,000,000	23.7%	6,000,000,000	No
7	blogspot.com	Blogging Resources & Services	340,000,000	19.6%	4,900,000,000	Yes
8	baidu.com	Search Engines	300,000,000	17.5%	110,000,000,000	Yes
9	microsoft.com	Software	250,000,000	14.5%	2,500,000,000	Yes
10	qq.com	Web Portals	250,000,000	14.7%	39,000,000,000	Yes
11	bing.com	Search Engines	230,000,000	13.1%	9,500,000,000	Yes
12	ask.com	Search Engines	190,000,000	11.2%	2,000,000,000	Yes
13	adobe.com	Multimedia Software	160,000,000	9.2%	1,000,000,000	No

WIKIPEDIA
The Free Encyclopedia

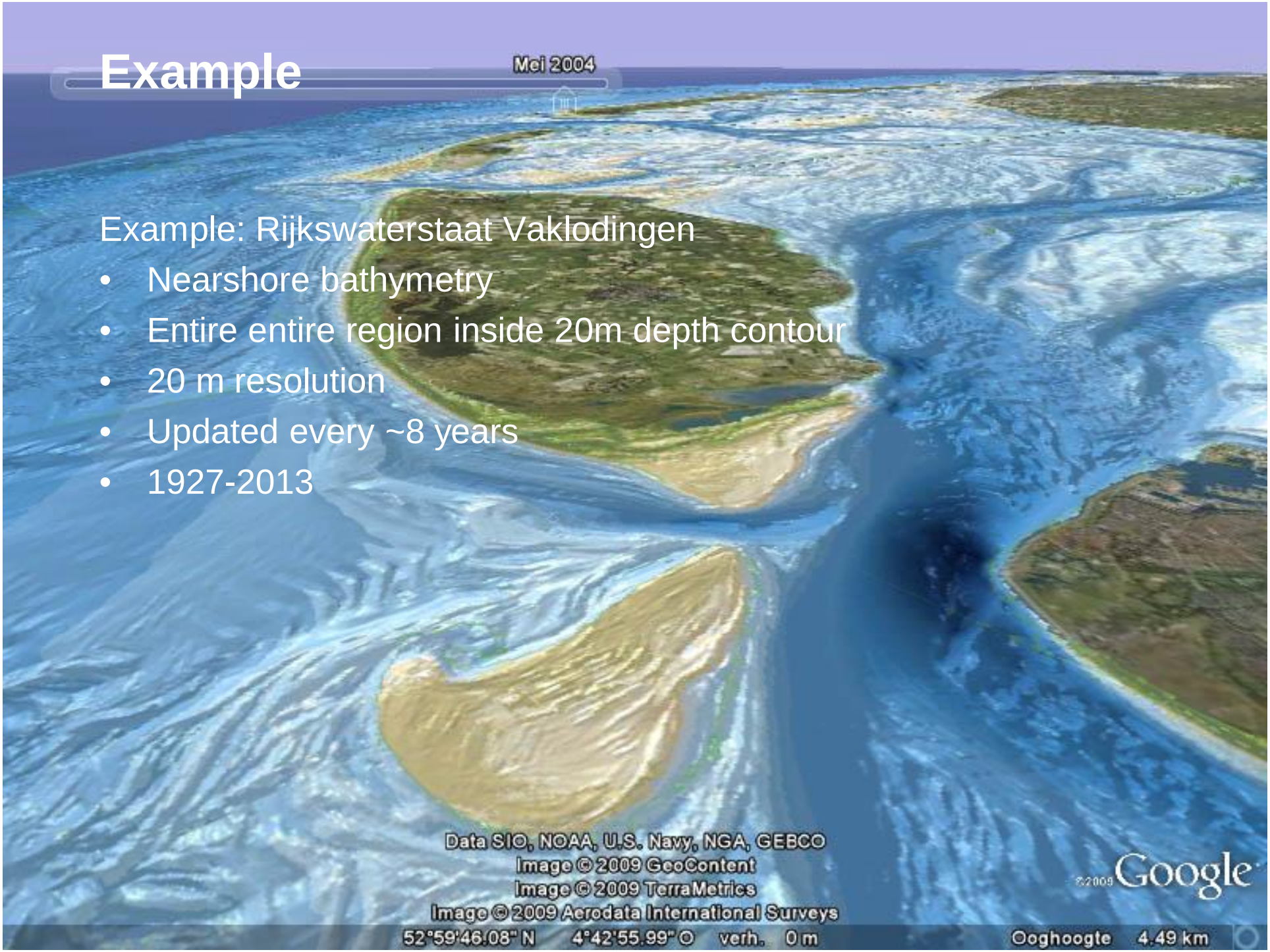
Communities: developers formerly known as users

- Wikipedia's 1 % rule:
  - 100 visitors: only 1 contributes
  - many users = many contributions
  - 1% of 400 million = 4 million
  - provided the workflow is organized smoothly
- Is the marine & coastal community big enough ?



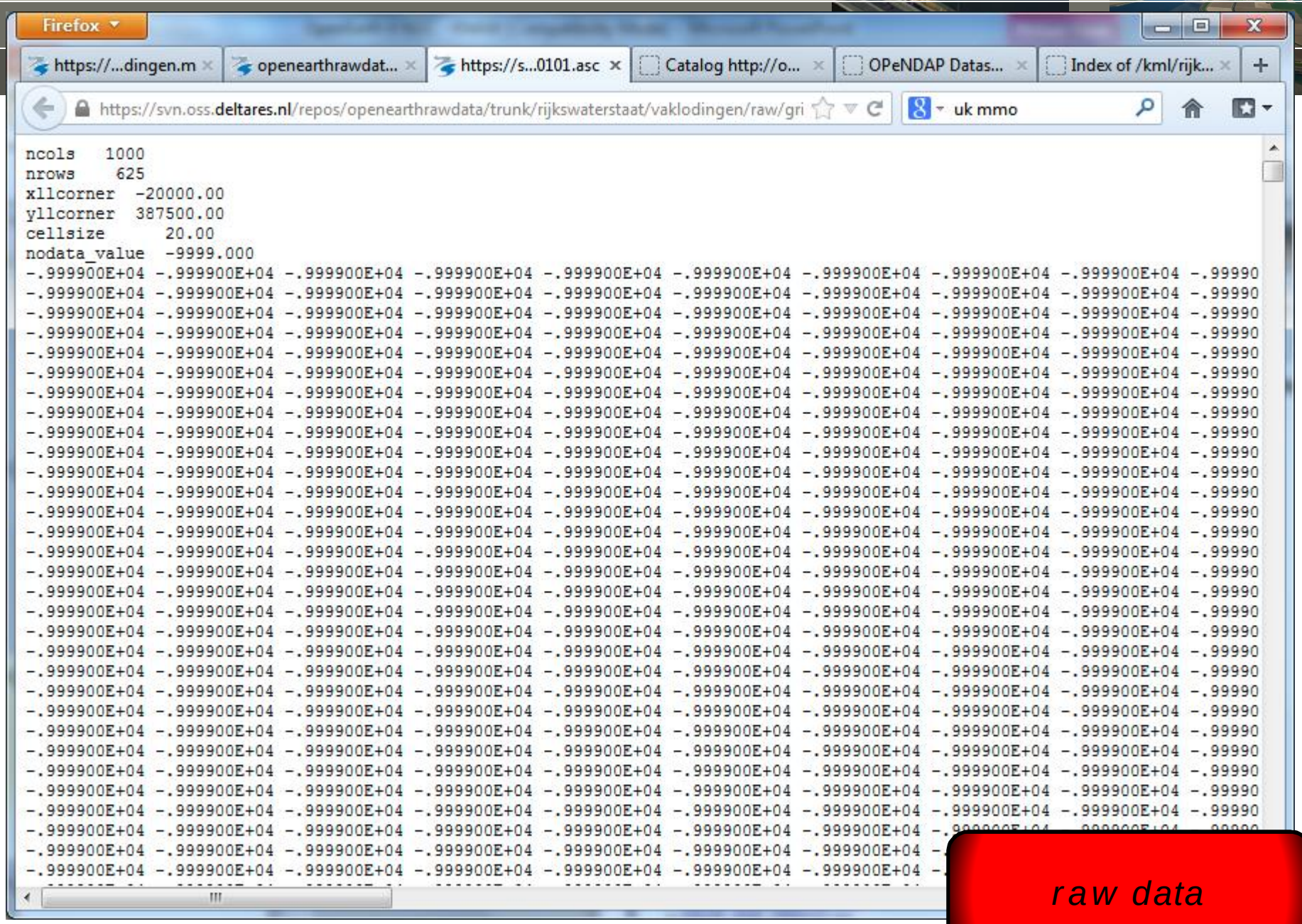


Example



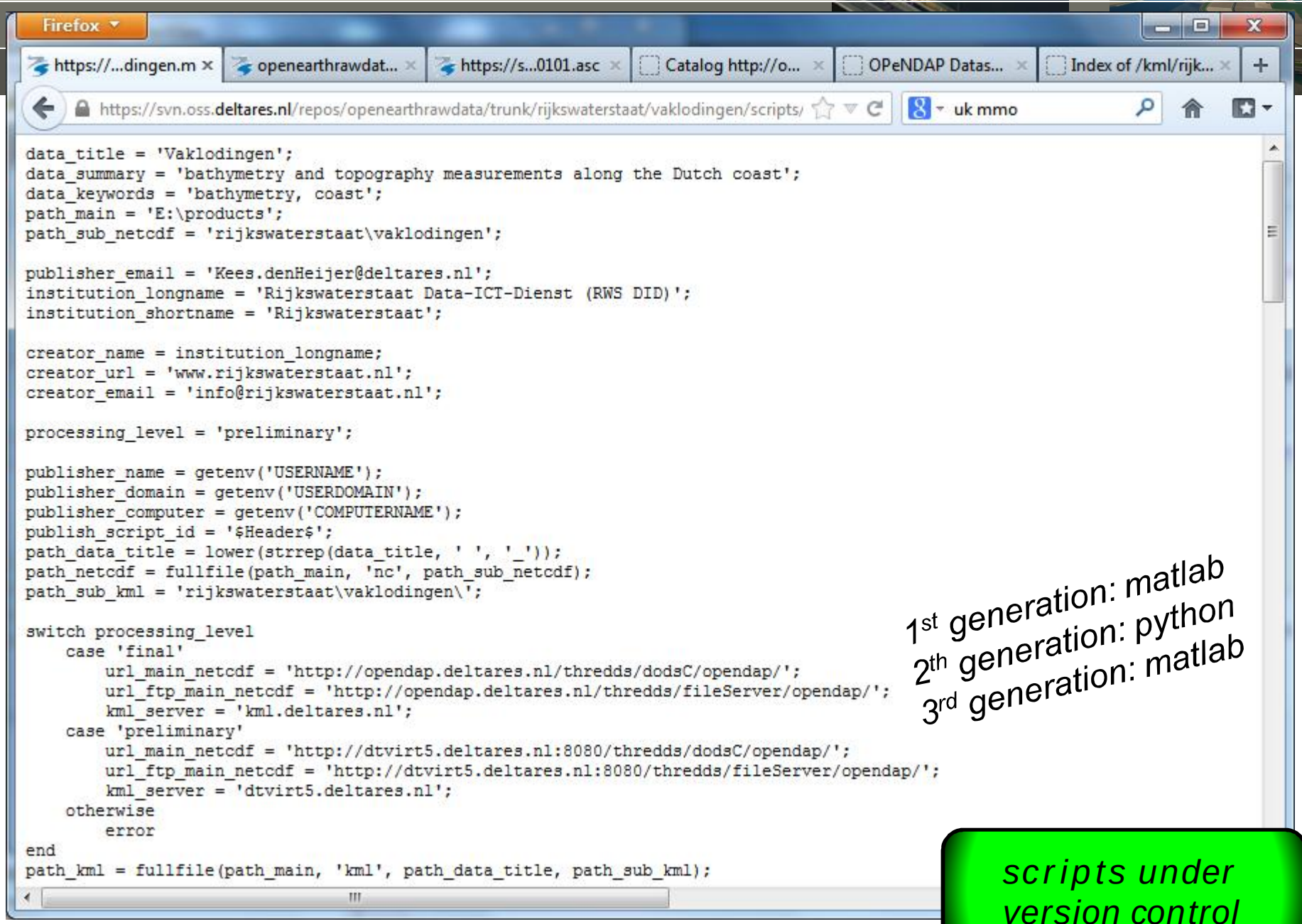
Example: Rijkswaterstaat Vaklodingen

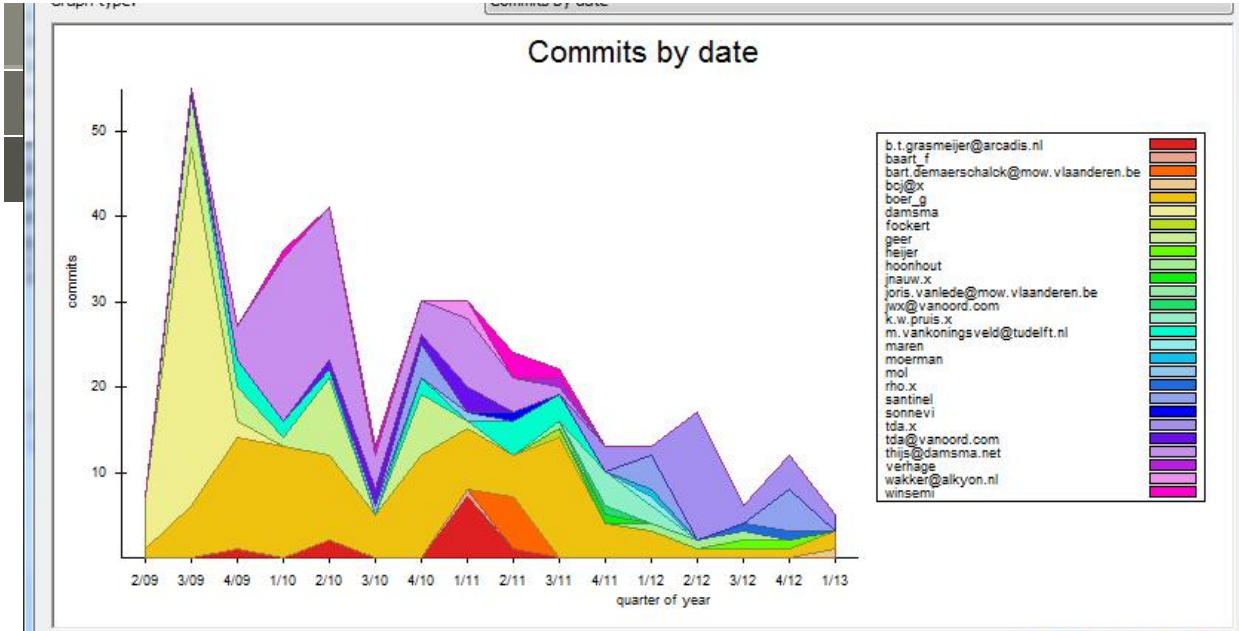
- Nearshore bathymetry
- Entire entire region inside 20m depth contour
- 20 m resolution
- Updated every ~8 years
- 1927-2013

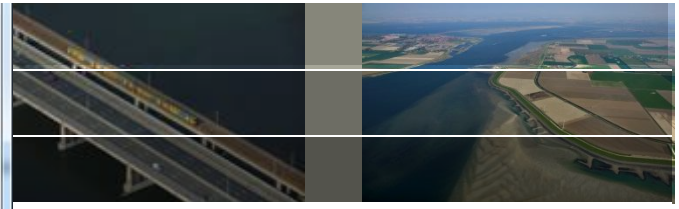

Firefox
https://...dingen.m
openearthrawdat...
https://s...0101.asc
Catalog http://o...
OPeNDAP Datas...
Index of /kml/rijk...
https://svn.oss.deltares.nl/repos/openearthrawdata/trunk/rijkswaterstaat/vaklodingen/raw/gri
uk mmo
ncols 1000
nrows 625
xllcorner -20000.00
yllcorner 387500.00
cellsize 20.00
nodata_value -9999.000
-.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.999900E+04 -.99990
raw data

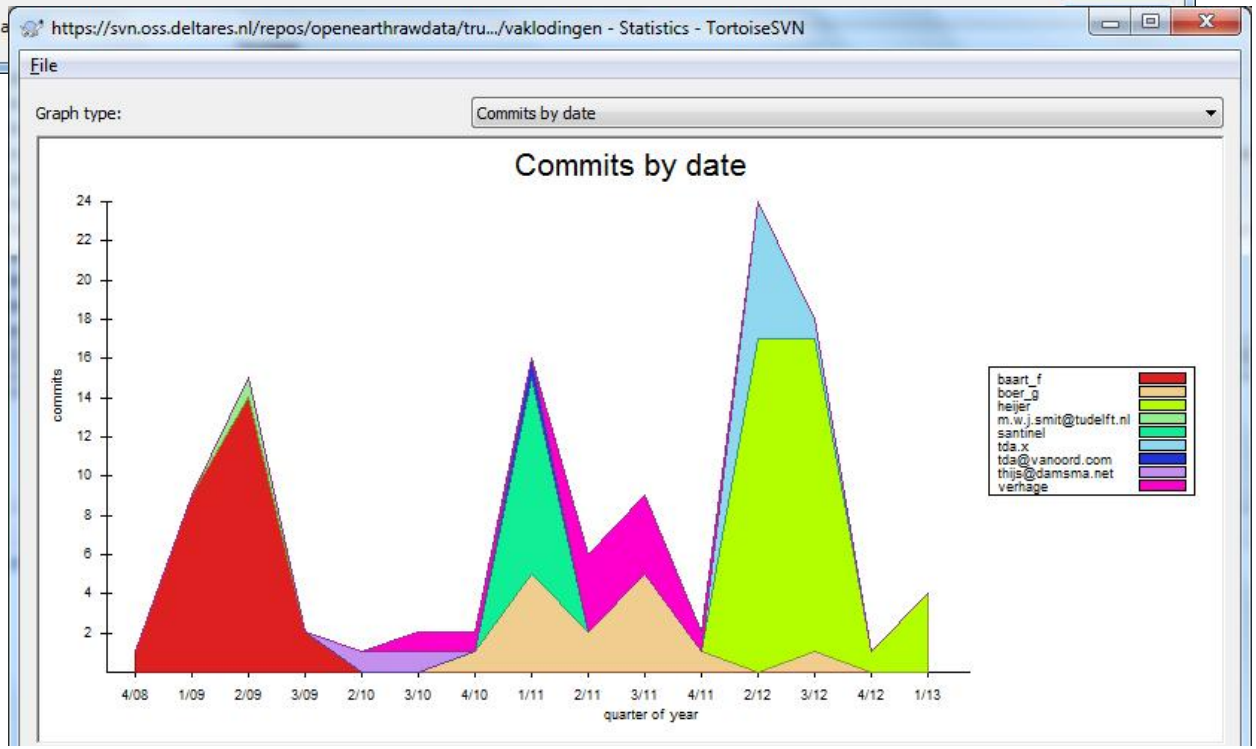
30,000 files

raw data


Firefox
https://...dingen.m
openearthrawdat...
https://s...0101.asc
Catalog http://o...
OPeNDAP Datas...
Index of /kml/rijk...
https://svn.oss.deltares.nl/repos/openearthrawdata/trunk/rijkswaterstaat/vaklodingen/scripts/
uk mmo
data_title = 'Vaklodingen';
data_summary = 'bathymetry and topography measurements along the Dutch coast';
data_keywords = 'bathymetry, coast';
path_main = 'E:\products';
path_sub_netcdf = 'rijkswaterstaat\vaklodingen';
publisher_email = 'Kees.denHeijer@deltares.nl';
institution_longname = 'Rijkswaterstaat Data-ICT-Dienst (RWS DID)';
institution_shortname = 'Rijkswaterstaat';
creator_name = institution_longname;
creator_url = 'www.rijkswaterstaat.nl';
creator_email = 'info@rijkswaterstaat.nl';
processing_level = 'preliminary';
publisher_name = getenv('USERNAME');
publisher_domain = getenv('USERDOMAIN');
publisher_computer = getenv('COMPUTERNAME');
publish_script_id = '$Header$';
path_data_title = lower(strrep(data_title, ' ', '_'));
path_netcdf = fullfile(path_main, 'nc', path_sub_netcdf);
path_sub_kml = 'rijkswaterstaat\vaklodingen\';
switch processing_level
case 'final'
url_main_netcdf = 'http://opendap.deltares.nl/thredds/dodsC/opendap/';
url_ftp_main_netcdf = 'http://opendap.deltares.nl/thredds/fileServer/opendap/';
kml_server = 'kml.deltares.nl';
case 'preliminary'
url_main_netcdf = 'http://dtvirt5.deltares.nl:8080/thredds/dodsC/opendap/';
url_ftp_main_netcdf = 'http://dtvirt5.deltares.nl:8080/thredds/fileServer/opendap/';
kml_server = 'dtvirt5.deltares.nl';
otherwise
error
end
path_kml = fullfile(path_main, 'kml', path_data_title, path_sub_kml);
1st generation: matlab
2th generation: python
3rd generation: matlab
scripts under version control







Revisions:

OpenEarthTools 7143
← GooglePlot 351

OpenEarthRawData 5414
← Vaklodingen 112

scripts under version control

https://svn.oss.deltares.nl/repos/openearthrawdata/t.../vaklodingen - Log Messages - TortoiseSVN

Filter by Messages, Paths, Authors, Revisions, Bug-IDs, Date, Date Range

Revision	Actions	Author	Date	Message
5545		heijer	Thursday, March 07, 2013 7:47:01 PM	acknowledgment added for German Ems-Dollard data to WSA Emden
5523		heijer	Thursday, February 28, 2013 10:29:00 AM	svn keyword added
5522		heijer	Thursday, February 28, 2013 10:26:44 AM	ncgen_vaklodingen.m rewritten in a similar way to ncgen_jarkus_grids
5521		heijer	Thursday, February 28, 2013 10:01:59 AM	Missing data at Zuid-Holland area in KB119_3332/KB119_3130 added - 2 files restored from revision 5326 wl
5388		heijer	Tuesday, October 16, 2012 5:13:23 PM	todo notes added
5345		heijer	Friday, September 14, 2012 11:20:46 AM	description changed
5343		heijer	Tuesday, September 11, 2012 4:11:01 PM	ncgen_vaklodingen.m modified: - nc_cf_gridset_get_list command added - main kml with hard paths added i
5342		heijer	Tuesday, September 11, 2012 1:22:45 PM	a few adjustments to the rename function
5341		heijer	Tuesday, September 11, 2012 1:16:25 PM	file renamed
5340		heijer	Tuesday, September 11, 2012 1:01:00 PM	Friesche Zeegat data 2012 added as received via John
5337		boer_g	Tuesday, September 04, 2012 11:15:53 PM	added baseUrl (web viewing requires absolute links)
5333		heijer	Thursday, August 30, 2012 8:39:21 AM	datatype of x and y changed to int32 and some other modifications to ncgen_vaklodingen
5327		heijer	Tuesday, August 28, 2012 1:25:48 PM	two tiles of 2008 replaced with data as received via John
5326		heijer	Tuesday, August 28, 2012 12:02:59 PM	various global attributes added
5324		tda.x	Friday, August 24, 2012 5:59:04 PM	modified epoch to 1970-01-01
5323		heijer	Tuesday, August 21, 2012 5:33:07 PM	metadata added
5320		heijer	Tuesday, August 14, 2012 5:21:24 PM	datatypes changed, since uint16 is not supported in netcdf-3
5319		heijer	Tuesday, August 14, 2012 5:15:58 PM	changed to netcdf-3 format
5317		heijer	Friday, August 10, 2012 2:46:31 PM	urlPath in catalog.nc corrected
5316		heijer	Thursday, August 09, 2012 5:16:11 PM	path corrected
5315		heijer	Thursday, August 09, 2012 4:44:24 PM	some adjustments to ncgen_vaklodingen.m
5314		heijer	Thursday, August 09, 2012 3:55:56 PM	file renamed
5313		heijer	Thursday, August 09, 2012 3:54:07 PM	raw file zuid-holland 2011 added, as received via John
5307		heijer	Tuesday, June 26, 2012 7:51:09 PM	write.filenameFcn changed conform current vaklodingen names
5306		heijer	Tuesday, June 26, 2012 5:22:40 PM	file locations changed

acknowledgment added for German Ems-Dollard data to WSA Emden

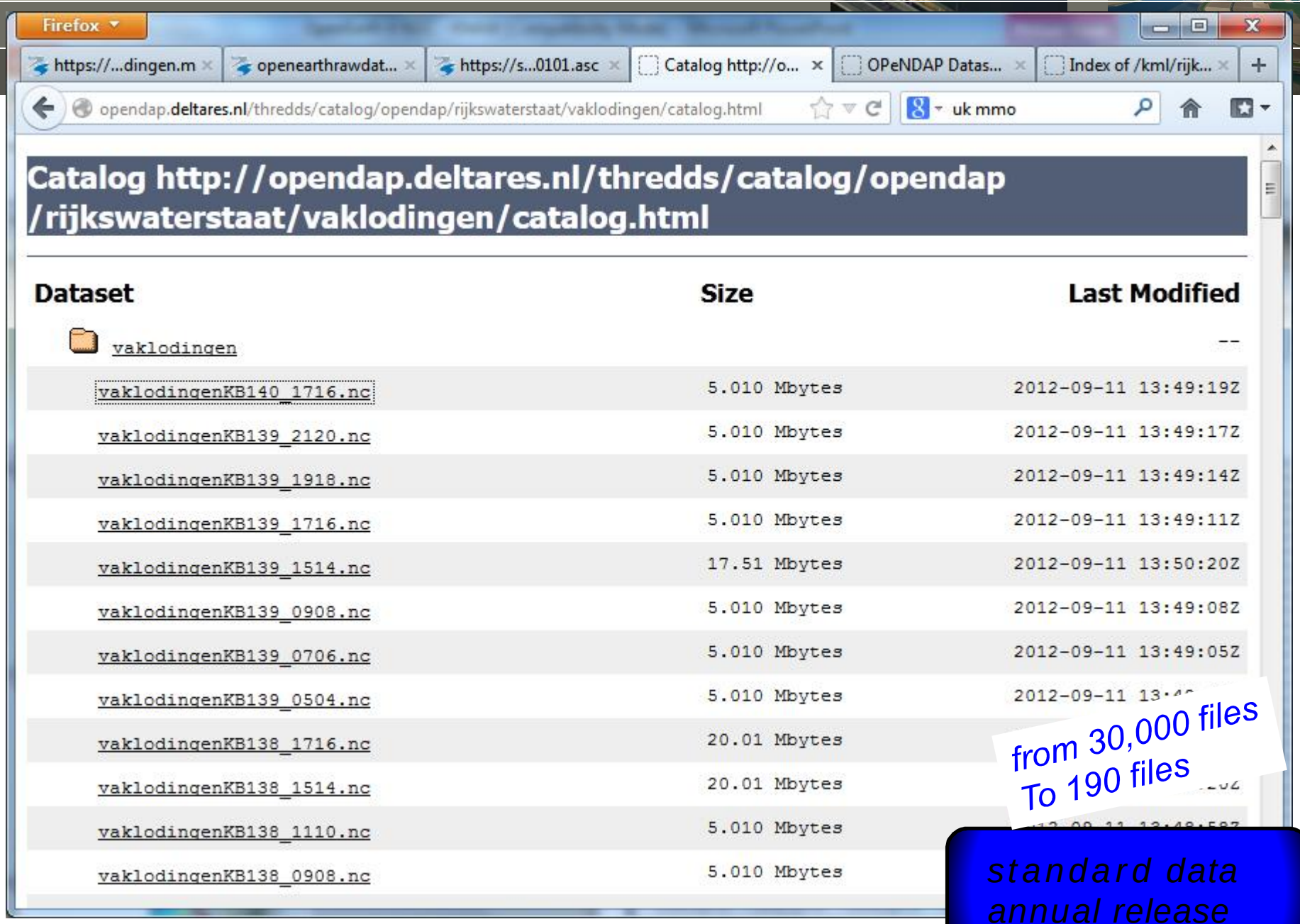
Path	Action	Copy from path	Revision

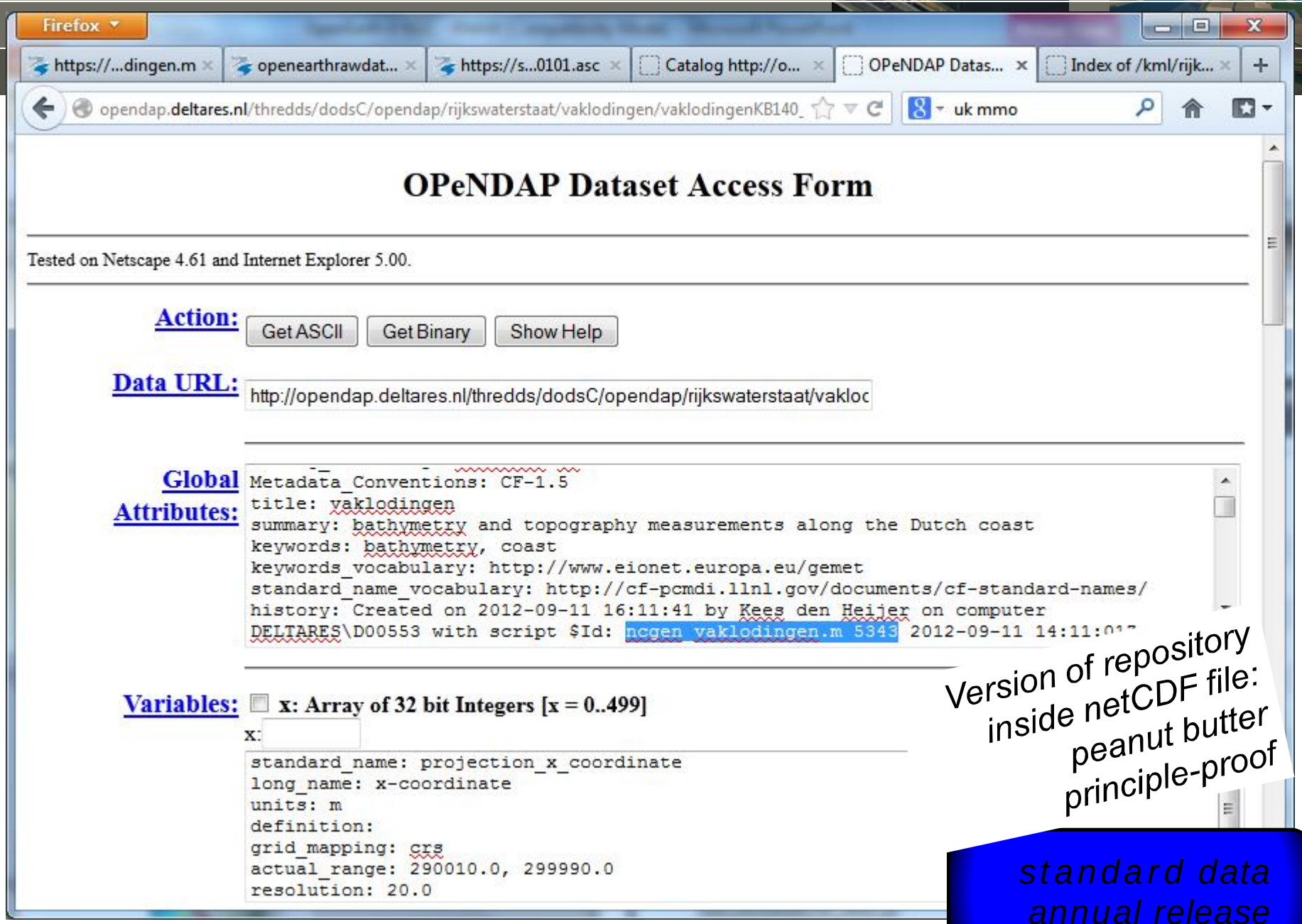
Showing 100 revision(s), from revision 296 to revision 5545 - 1 revision(s) selected, showing 1 changed paths

- [ ] Show only affected paths
- [ ] Stop on copy/rename
- [ ] Include merged revisions

Next 100 | Refresh

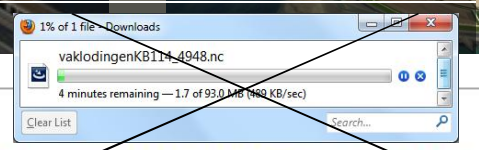
scripts under version control


Firefox
https://...dingen.m
openearthrawdat...
https://s...0101.asc
Catalog http://o...
OPeNDAP Datas...
Index of /kml/rijk...
opendap.deltares.nl/thredds/catalog/opendap/rijkswaterstaat/vaklodingen/catalog.html
uk mmo
Catalog http://opendap.deltares.nl/thredds/catalog/opendap/rijkswaterstaat/vaklodingen/catalog.html
Dataset
Size
Last Modified
vaklodingen --
vaklodingenKB140_1716.nc 5.010 Mbytes 2012-09-11 13:49:19Z
vaklodingenKB139_2120.nc 5.010 Mbytes 2012-09-11 13:49:17Z
vaklodingenKB139_1918.nc 5.010 Mbytes 2012-09-11 13:49:14Z
vaklodingenKB139_1716.nc 5.010 Mbytes 2012-09-11 13:49:11Z
vaklodingenKB139_1514.nc 17.51 Mbytes 2012-09-11 13:50:20Z
vaklodingenKB139_0908.nc 5.010 Mbytes 2012-09-11 13:49:08Z
vaklodingenKB139_0706.nc 5.010 Mbytes 2012-09-11 13:49:05Z
vaklodingenKB139_0504.nc 5.010 Mbytes
vaklodingenKB138_1716.nc 20.01 Mbytes
vaklodingenKB138_1514.nc 20.01 Mbytes
vaklodingenKB138_1110.nc 5.010 Mbytes
vaklodingenKB138_0908.nc 5.010 Mbytes
from 30,000 files
To 190 files
standard data
annual release

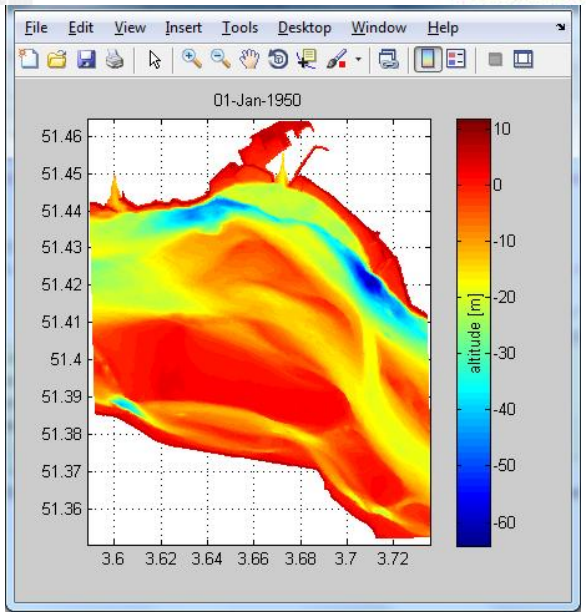

Firefox
https://...dingen.m
openearthrawdat...
https://s...0101.asc
Catalog http://o...
OPeNDAP Datas...
Index of /kml/rijk...
opendap.deltares.nl/thredds/dodsC/opendap/rijkswaterstaat/vaklodingen/vaklodingenKB140_
uk mmo
OPeNDAP Dataset Access Form
Tested on Netscape 4.61 and Internet Explorer 5.00.
Action:
Get ASCII
Get Binary
Show Help
Data URL:
http://opendap.deltares.nl/thredds/dodsC/opendap/rijkswaterstaat/vakloc
Global Attributes:
Metadata_Conventions: CF-1.5
title: vaklodingen
summary: bathymetry and topography measurements along the Dutch coast
keywords: bathymetry, coast
keywords_vocabulary: http://www.eionet.europa.eu/gemet
standard_name_vocabulary: http://cf-pcmdi.llnl.gov/documents/cf-standard-names/
history: Created on 2012-09-11 16:11:41 by Kees den Heijer on computer
DELTARES\D00553 with script $Id: ncgen vaklodingen.m 5343 2012-09-11 14:11:
Variables:
x: Array of 32 bit Integers [x = 0..499]
x:
standard_name: projection_x_coordinate
long_name: x-coordinate
units: m
definition:
grid_mapping: crs
actual_range: 290010.0, 299990.0
resolution: 20.0
Version of repository inside netCDF file: peanut butter principle-proof
standard data annual release

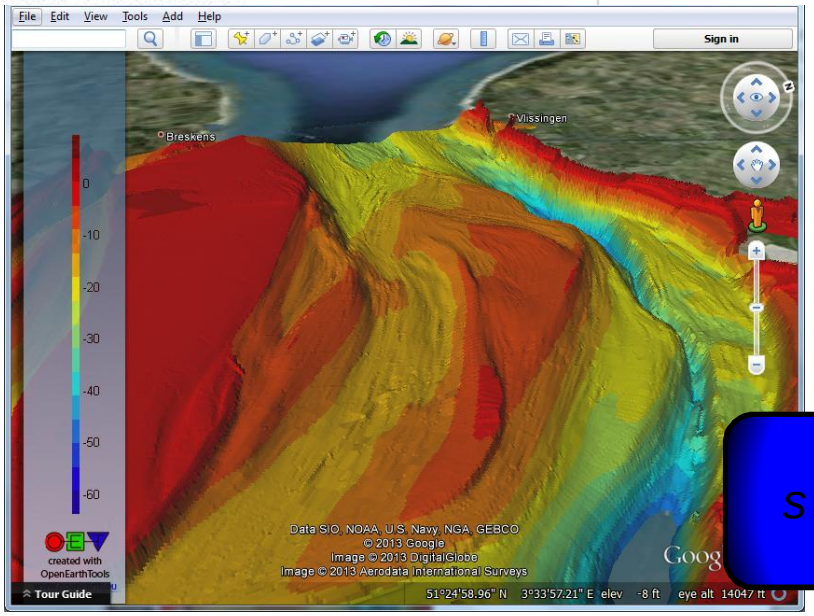
② How we switched to web-data changing *one line* of code

```
1   % oetsettings
2   %% load web-data
3 - url = 'D:/download/vaklodingenKB114_4948.nc'; % offline cache
4 - url = 'http://opendap.deltares.nl/thredds/dodsC/opendap/rijkswaterstaat/vaklodingen/vaklodingenKB114_4948.nc'; % online
5 - D.x           = ncread(url,'x');
6 - D.y           = ncread(url,'y');
7 - [D.x,D.y]     = meshgrid(D.x,D.y);
8 - [D.lon,D.lat] = convertCoordinates(D.x,D.y,'CS1.code',28992,'CS2.code',4326); % RD to WGS84
9 - D.time        = nc_cf_time(url); % 38 times
10- D.z           =  ncread(url,'z',[1 1 length(D.time)],[Inf Inf 1])'; % get last z
11- D.label       = [nc_attget(url,'z','long_name'),' [',nc_attget(url,'z','units'),']']
12  %% plot data
13- pcolorcorcen(D.lon,D.lat,D.z);
14- title(datestr(D.time(1)))
15- colorbarwithvtext(D.label)
16- axislat; grid on; tickmap('ll','fmt','%g')
17  %% plot in Google Earth
18- KMLsurf(D.lat,D.lon,D.z, 'fileName',[vaklodingenKB114_4948,'_',datestr(D.time(end)),'.kmz'],...
19                         'zScaleFun',@(z)(z+50).*5);
```



CERTIFIED OGC® COMPLIANT





How scientist work with standard data

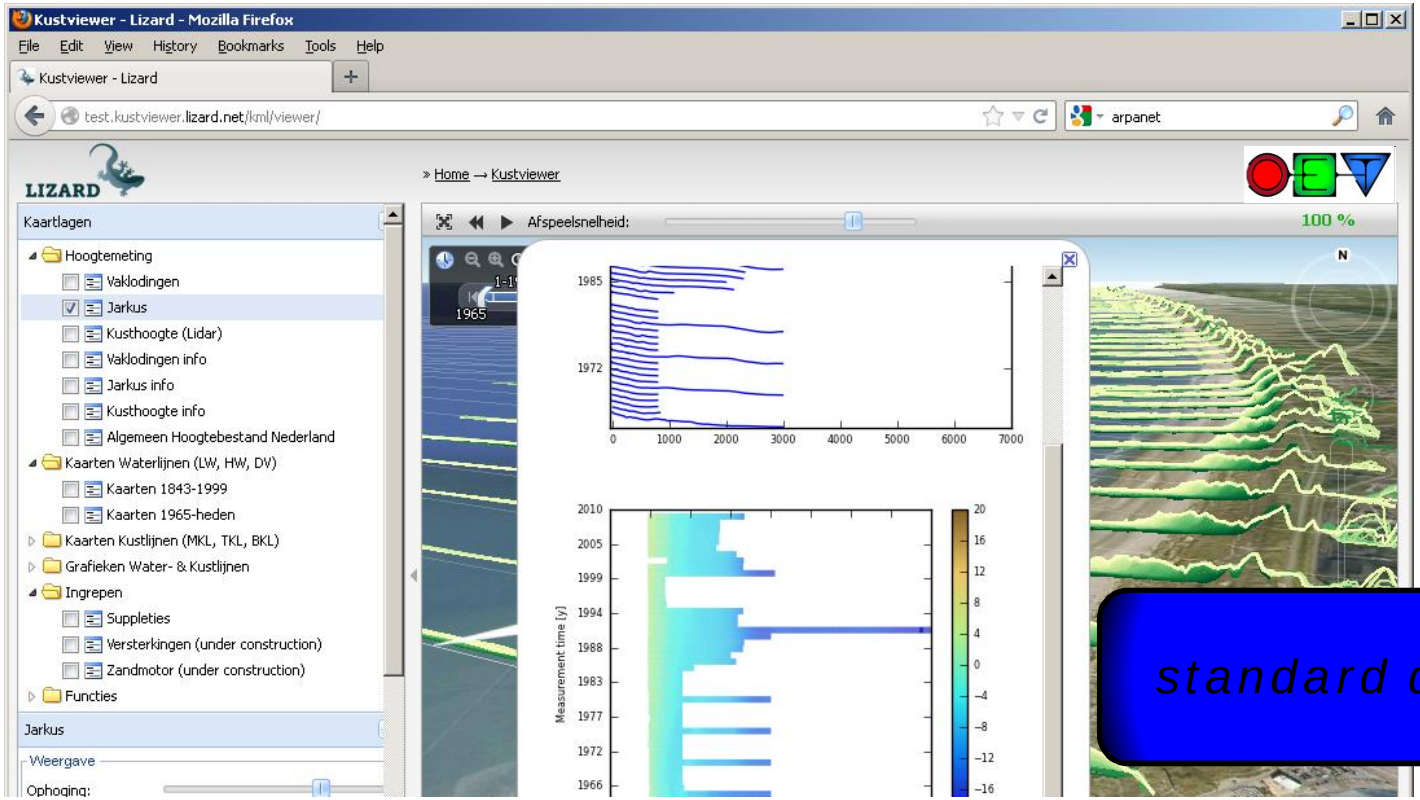
standard data

4th paradigm: interactive web viewer:

http://kustviewer.lizard.net/

Data scarcity > data abundance: data scientist most sexy job 21st century

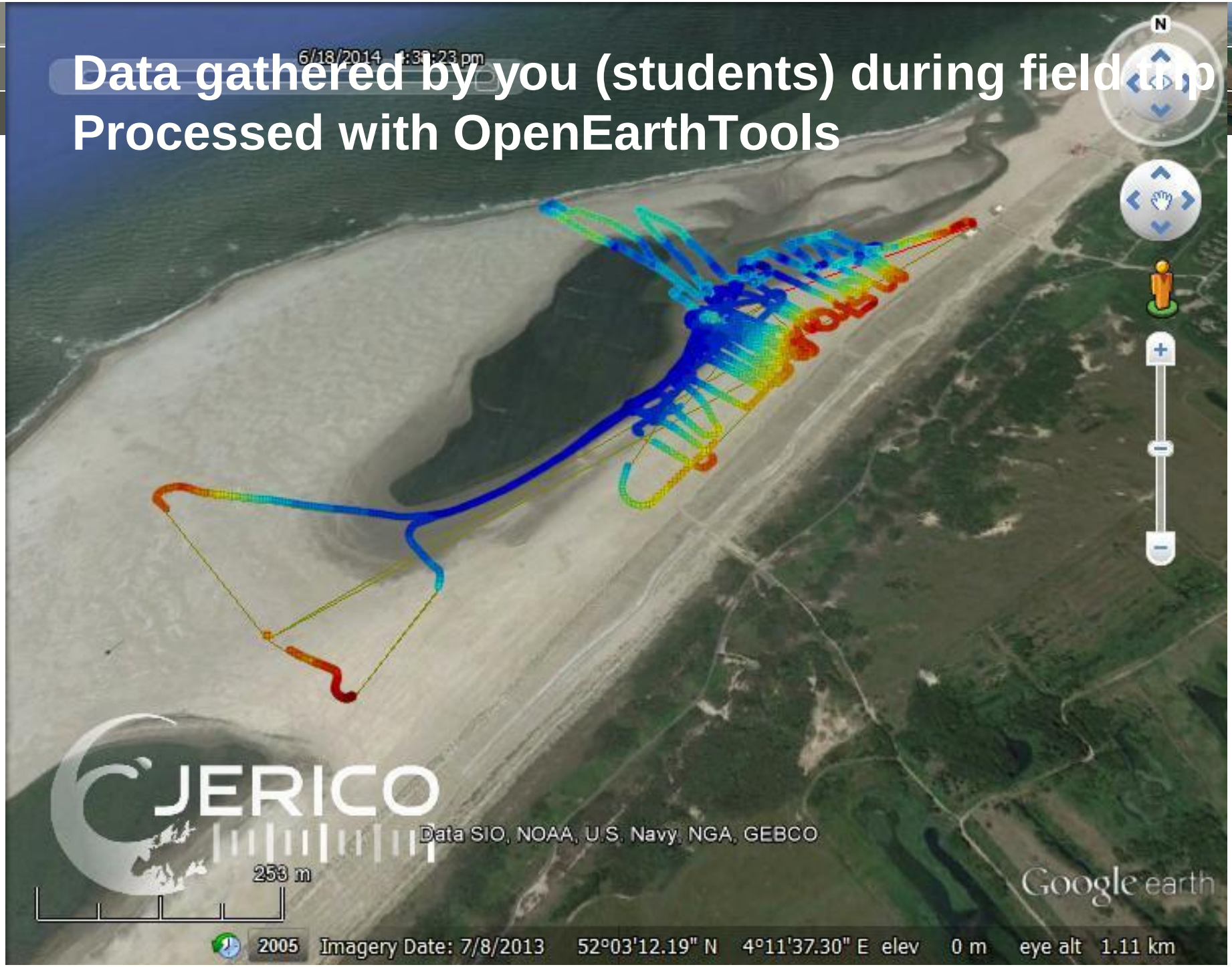
1. 1000's yrs: Empirical Archimedes ea.
2. 100's yrs: Theoretical Newton ea.
3. 10's yrs: Computational Neumann ea.
4. Now 4th paradigm by Microsoft Research



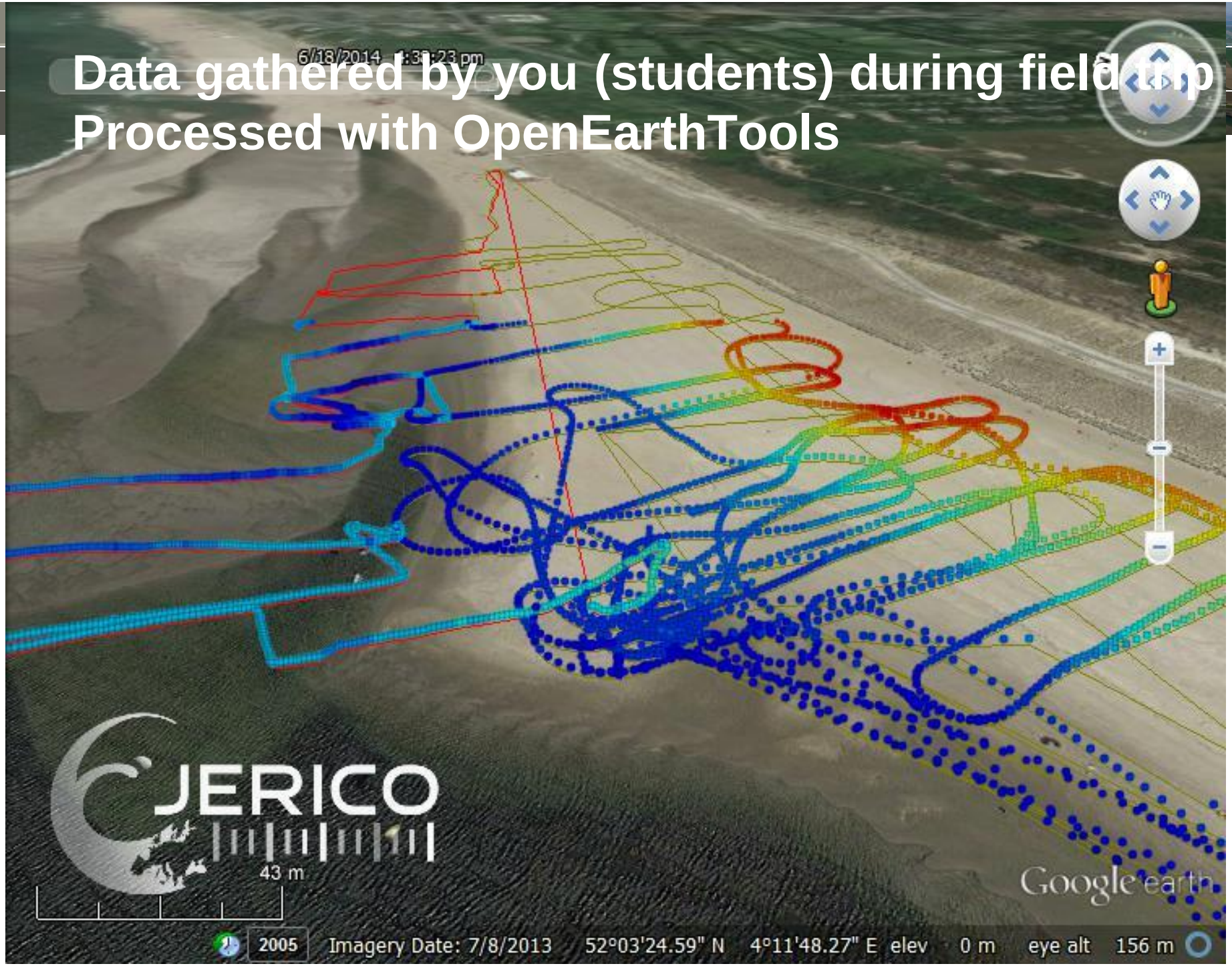
4th paradigm: client-server in this summerschool

- This summer school featured 1 day courses
- 1 day is too short to get hands-on with a complex command line executable
- Many difficult analyses this week deployed an interactive web tool:
  - 15th interactive EMECO web tool
  - 16th interactive DIVA web tool
  - 16th interactive DINEOF web tool
  - 17th interactive ICES web tool
  - 17th interactive EMODnet-physics web tool
  - 20th OpenEarth web tool on dune profiles
  - (20th instead of: EMODnet chemistry web tool)
- Your generation will be the ones making/using such tools
- Exercise: analysis the elevation “transects” gathered by JERICO summer school using OpenEarthTools or any other tool

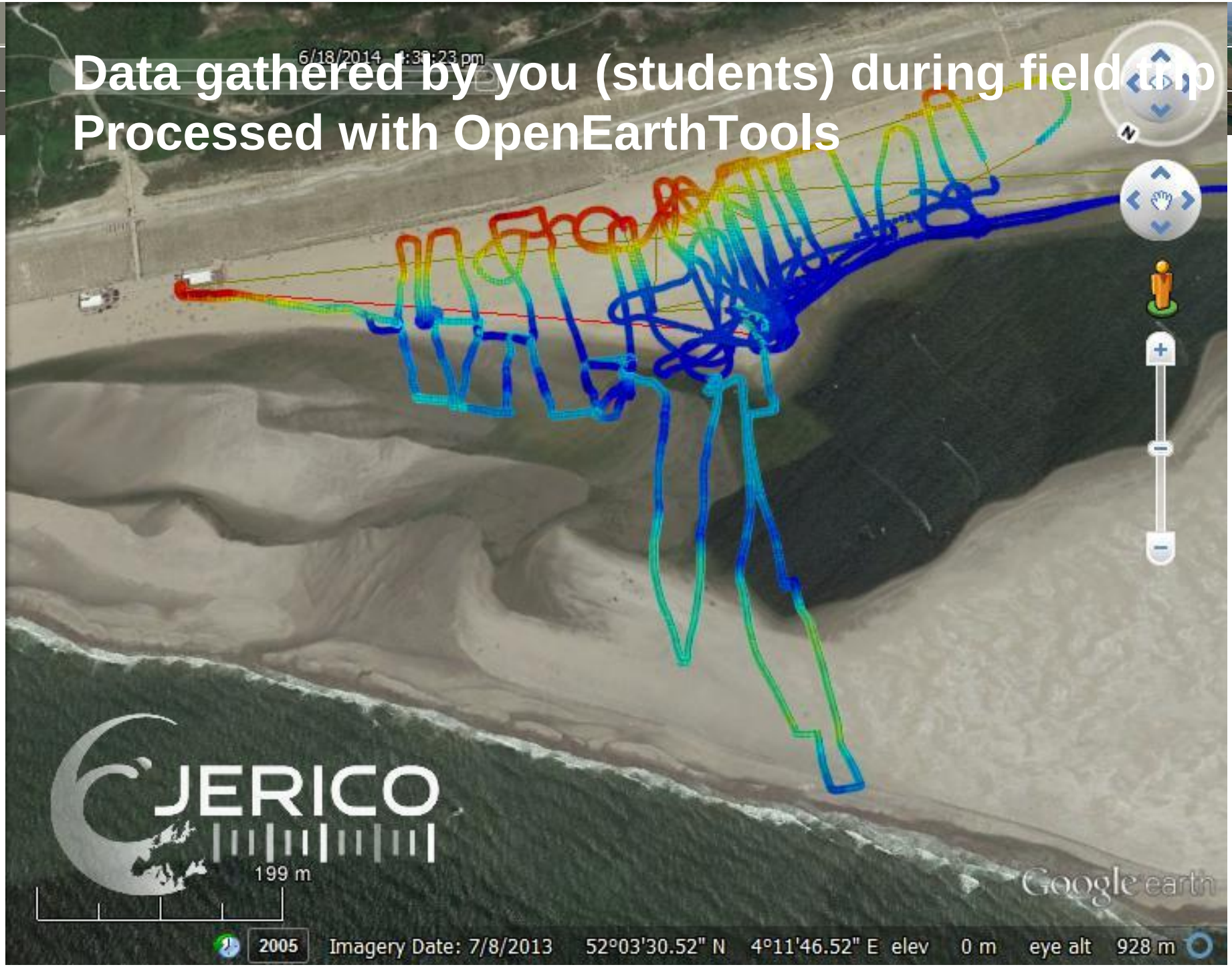
Data gathered by you (students) during field trip
Processed with OpenEarthTools



Data gathered by you (students) during field trip
Processed with OpenEarthTools



Data gathered by you (students) during field trip
Processed with OpenEarthTools



Raw data (ascii) + script (m) = processed data (kmz)

raw data

- Raw data (ascii + mat) for download from JERICO wiki: frozen
- https://publicwiki.deltares.nl/display/OET/jerico

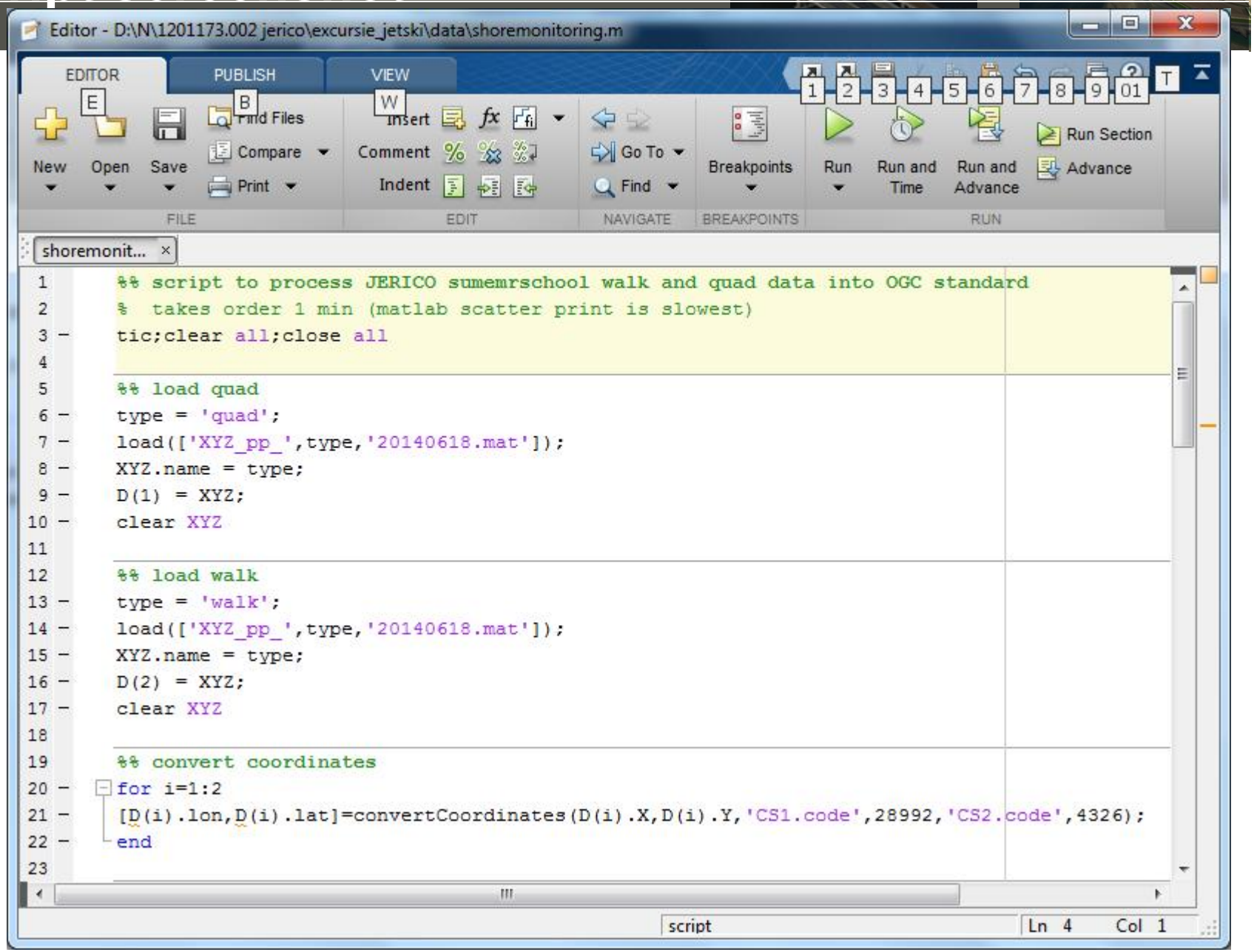
scripts under version control

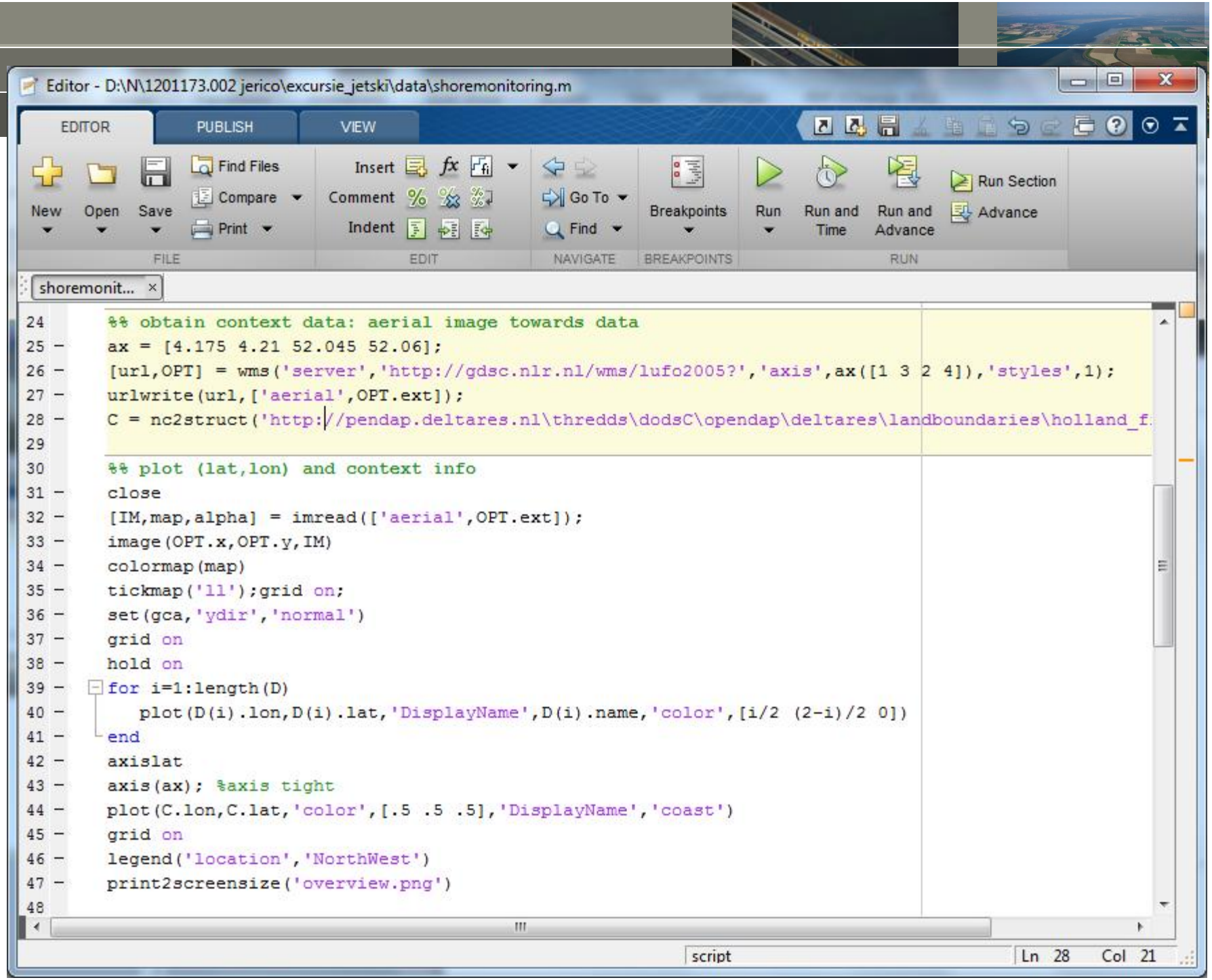
- Matlab scripts: not for download, added to SubVersion repository:
- Get OpenEarth account: oss.deltares.nl > sign-up (upper right)
- Install TortoiseSvn, make checkout of (only jerico folder due to size!!)
- http://svn.oss.deltares.nl/repos/openearthrawdata/trunk/jerico

standard data

- Processed data product (kmz) for download from JERICO wiki: https://publicwiki.deltares.nl/display/OET/jerico

Script screenshot




Editor - D:\N\1201173.002 jerico\excursie_jetski\data\shoremonitoring.m
EDITOR
PUBLISH
VIEW
New
Open
Save
Find Files
Compare
Print
FILE
Insert
Comment
Indent
EDIT
Go To
Find
NAVIGATE
Breakpoints
BREAKPOINTS
Run
Run and Time
Run and Advance
Run Section
Advance
RUN
shoremonit...
24 %% obtain context data: aerial image towards data
25 – ax = [4.175 4.21 52.045 52.06];
26 – [url,OPT] = wms('server','http://gdsc.nlr.nl/wms/lufo2005?','axis',ax([1 3 2 4]),'styles',1);
27 – urlwrite(url,['aerial',OPT.ext]);
28 – C = nc2struct('http://pendap.deltares.nl\thredds\dodsC\opendap\deltares\landboundaries\holland_f
29
30 %% plot (lat,lon) and context info
31 – close
32 – [IM,map,alpha] = imread(['aerial',OPT.ext]);
33 – image(OPT.x,OPT.y,IM)
34 – colormap(map)
35 – tickmap('ll');grid on;
36 – set(gca,'ydir','normal')
37 – grid on
38 – hold on
39 – for i=1:length(D)
40 – plot(D(i).lon,D(i).lat,'DisplayName',D(i).name,'color',[i/2 (2-i)/2 0])
41 – end
42 – axislat
43 – axis(ax); %axis tight
44 – plot(C.lon,C.lat,'color',[.5 .5 .5],'DisplayName','coast')
45 – grid on
46 – legend('location','NorthWest')
47 – print2screensize('overview.png')
48
script
Ln 28 Col 21

Editor - D:\N\1201173.002 jerico\excursie_jetski\data\shoremonitoring.m

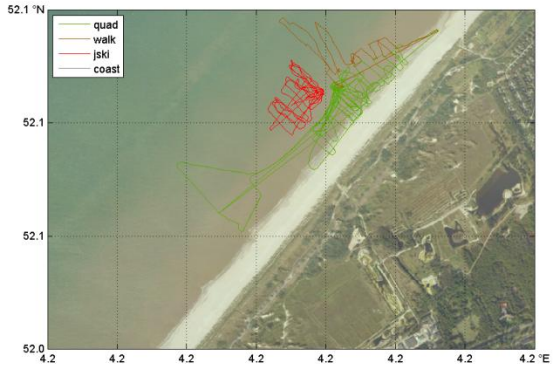
EDITOR | PUBLISH | VIEW

New | Open | Save | Find Files | Compare | Print — FILE
Insert | Comment | Indent — EDIT
Go To | Find — NAVIGATE
Breakpoints — BREAKPOINTS
Run | Run and Time | Run and Advance | Run Section | Advance — RUN

shoremonit...

```
24      %% obtain context data: aerial image towards data
25 -    ax = [4.175 4.21 52.045 52.06];
26 -    [url,OPT] = wms('server','http://gdsc.nlr.nl/wms/lufo2005?','axis',ax([1 3 2 4]),'styles',1);
27 -    urlwrite(url,['aerial',OPT.ext]);
28 -    C = nc2struct('http://pendap.deltares.nl\thredds\dodsC\opendap\deltares\landboundaries\holland_f
29
30      %% plot (lat,lon) and context info
31 -    close
32 -    [IM,map,alpha] = imread(['aerial',OPT.ext]);
33 -    image(OPT.x,OPT.y,IM)
34 -    colormap(map)
35 -    tickmap('ll');grid on;
36 -    set(gca,'ydir','normal')
37 -    grid on
38 -    hold on
39 -    for i=1:length(D)
40 -       plot(D(i).lon,D(i).lat,'DisplayName',D(i).name,'color',[i/2 (2-i)/2 0])
41 -    end
42 -    axislat
43 -    axis(ax); %axis tight
44 -    plot(C.lon,C.lat,'color',[.5 .5 .5],'DisplayName','coast')
45 -    grid on
46 -    legend('location','NorthWest')
47 -    print2screensize('overview.png')
48
```

script



Editor - D:\N\1201173.002 jerico\excursie_jetski\data\shoremonitoring.m

EDITOR | PUBLISH | VIEW

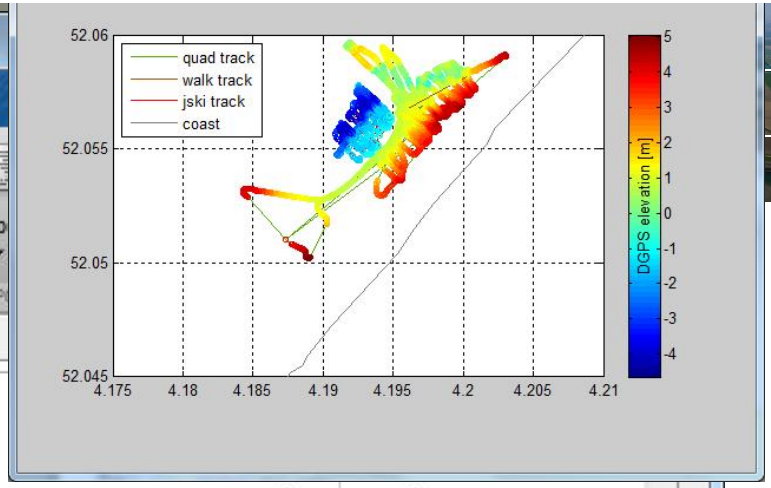
New | Open | Save | Find Files | Compare | Print | Insert | Comment | Indent | Go To | Find | Breakp

FILE | EDIT | NAVIGATE | BREAKP

shoremonit... ×

```
49    %% scatter
50 -  close
51 -  for i=1:length(D)
52 -     plot(D(i).lon,D(i).lat,'DisplayName',[D(i).name,' track'],'color',[i/2 (2-i)/2 0])
53 -     hold on
54 -     scatter(D(i).lon,D(i).lat,20,D(i).Z,'HandleVisibility','off')
55 -  end
56 -  axislat
57 -  axis(ax); %axis tight
58 -  plot(C.lon,C.lat,'color',[.5 .5 .5],'DisplayName','coast')
59 -  grid on
60 -  colorbarwithvtext('DGPS elevation [m]')
61 -  legend('location','NorthWest')
62    %print2screensize('overview_scatter.png')% too slow, Matlab is rubbish
63
64    %% plot Google Earth: data towards aerial image
65 -  for i=1:2
66 -  KMLline(D(i).lat,D(i).lon,'fileName',[D(i).name,'_line.kml'],...
67        'lineColor',[i/2 (2-i)/2 0],...
68        'kmlName',D(i).name)
69 -  KMLscatter(D(i).lat,D(i).lon,D(i).Z,'fileName',[D(i).name,'_time.kml'],...
70        'description','JERICO summerschool Zandmotor practical',...
71        'name',cellstr((num2str(D(i).Z'))),...
72        'timeIn',D(i).T,...
73        'CBcolorTitle',['DGPS elevation [m] by ',type],...
74        'timeOut',repmat(D(i).T(end),size(D(i).T)))
75 -  end
```

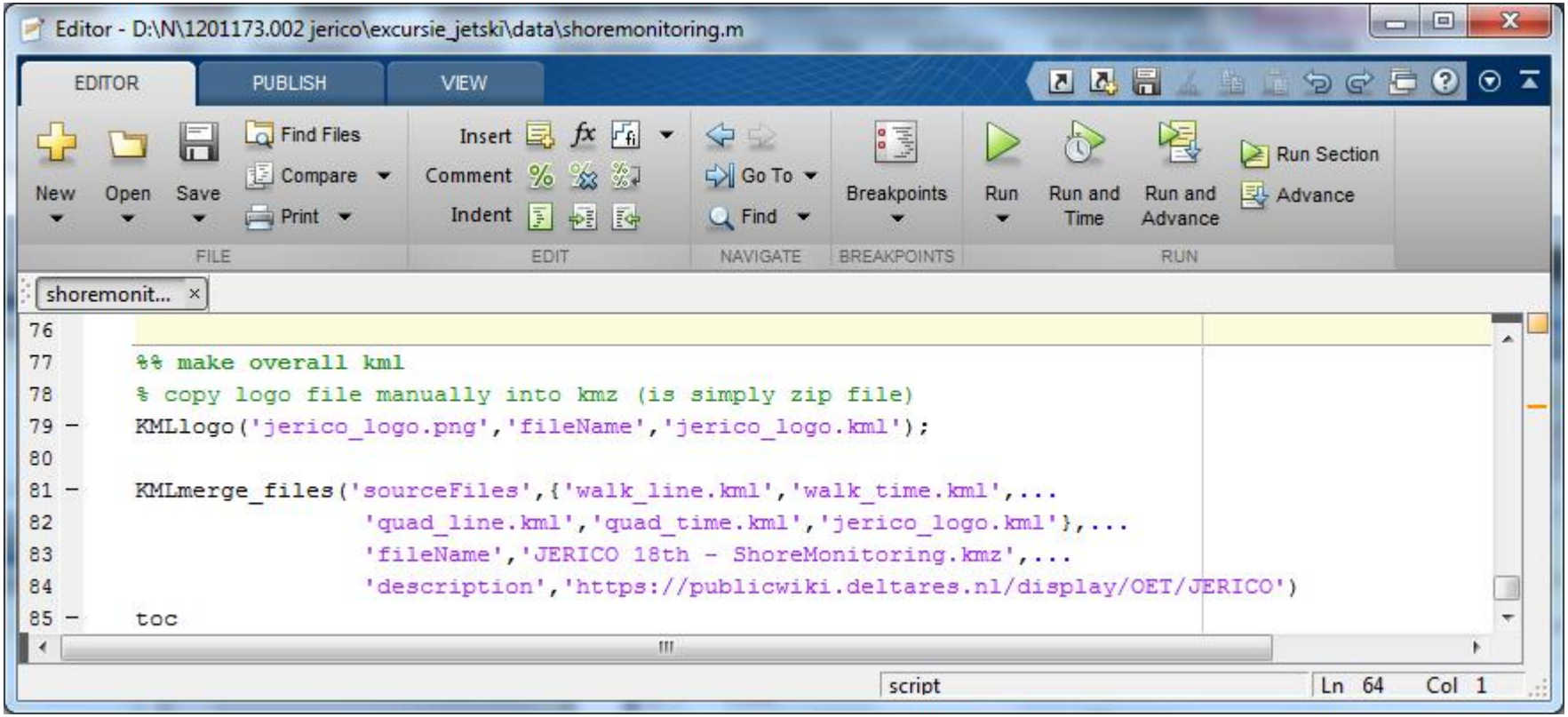
script | Ln 64 | Col 1



Running this script requires OpenEarthTools

https://publicwiki.deltares.nl/display/OET/Join+OpenEarth

http://oss.deltares.nl/web/delft3d/webinars, see webinar no 15



6/18/2014 4:38:23 pm
DGPS elevation [m] by jski
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
JERICO
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
327 m
created with OpenEarthTools
Google earth
Tour Guide
2005
Imagery Date: 7/8/2013 52°03'11.70" N 4°11'34.38" E elev 0 m eye alt 1.41 km