

Freshwater and Citizen Science: Research Hackathon

Data descriptions

25th - 27th May 2017 - Oxford Brooke University, Harcourt Hill Campus

Contacts: Ian Thornhill (ithornhill@earthwatch.org.uk)

Luis Felipe Velasquez (lfvelasquez@earthwatch.org.uk)

Leticia Miguel Chinchilla (lmiguel@earthwatch.org.uk)

Supported by the British Ecological Society, Citizen Science Special Interest Group



Contents

[1 Introduction 1](#_Toc482913442)

[1.1 FreshWater Watch data 1](#_Toc482913443)

[1.1.1 Geographical 1](#_Toc482913444)

[1.1.2 Temporal 1](#_Toc482913445)

[1.1.3 Nutrients 1](#_Toc482913446)

[1.1.4 Descriptive 1](#_Toc482913447)

[1.2 Environmental data 2](#_Toc482913448)

[1.2.1 Hydrobasin - WWF 2](#_Toc482913449)

[1.2.2 Esri – Terrain 2](#_Toc482913450)

[1.2.3 Void Filled Digital Elevation Model – HydroSHEDS - WWF 3](#_Toc482913451)

[1.2.4 Land Use 3](#_Toc482913452)

[1.2.5 Population density 3](#_Toc482913453)

[1.2.6 World Urban Areas 4](#_Toc482913454)

[1.2.7 Biome 4](#_Toc482913455)

[1.2.8 Protected Areas 4](#_Toc482913456)

[1.2.9 NDVI 5](#_Toc482913457)

[1.2.10 Rainfall and Air Temperature 5](#_Toc482913458)

[1.3 Engagement data 6](#_Toc482913459)

[1.3.1 Training 6](#_Toc482913460)

[1.3.2 Sampling protocol 6](#_Toc482913461)

[1.3.3 Engagement markers 6](#_Toc482913462)

[1.3.4 Seasonality 6](#_Toc482913463)

[1.3.5 Environmental quality 6](#_Toc482913464)

# Introduction

The data briefly described here has been attributed to each FreshWater Watch sample and will be provided for the hackathon in .csv format. Unless specified, all data is open source and freely available (though some datasets are subject to open data licensing and require attribution).

FreshWater Watch is a global citizen science dataset which has engaged nearly 10,000 people across 35+ cities to take part in water quality testing. All FreshWater Watch data (~18,000 data) are geolocated and include data regarding nutrients (N-NO3 and P-PO4) and a variety of qualitative observations. In most cases outside of WaterBlitz events, turbidity is also measured. Locally (e.g. in a city), these measures have been coupled with additional measurements and varied sampling designs depending on the local circumstance. For more information on FreshWater Watch data please see the FWW methods document, example data form and here:

<https://freshwaterwatch.thewaterhub.org/>

Two broad datasets have been made available, 1) Environmental and 2) Engagement. These are explained below.

## FreshWater Watch data

### Geographical

* 'City'
* 'Country'
* 'Continent'
* 'Biome': 1 [Categorical]

### Temporal

* 'Sample Date'
* 'Month'
* 'Year'

### Nutrients

* 'NTU' Turbidity (NTU)
* 'Nitrate' (N-NO3, mg/L)
* 'Phosphate' (P-PO4 mg/L)

### Descriptive

* 'Type': River, Stream, Lake, Pond, Wetland, Other [Categorical]
* Land-use (immediate): LU.UrbRes = urban / residential, LU.Agri = Agriculture, LU.UP = urban park, LU.Ind = industrial, LU.Forest = forest, LU.Grass = grassland/shrub, , LU.Other =other [Binary]
* Point discharges: Dis.Res = residential, Dis.Urb = urban/road, Dis.Ind = Industrial, Dis.Oth = other [Binary]
* Dis.Sum = Count of point sources [Integer]
* Water use: Use.Boat = boating, Use.Fish = fishing, Use.Irr = irrigation, Use.Swim = Swimming, Use.Sup = water supply, Use.Dogs = Dog presence, Use.Oth = Other [Binary]
* Bankside vegetation: Bnk.Tress = trees/shrubs, Bnk.Grass = Grass, Bnk.Bare = bare, Bnk.Oth = other [Binary]
* Aquatic vegetation: Veg.Emer = emergent, Veg.Sub = Submerged, Veg.Flt = floating [Binary]
* Wildlife: Fish = Fish presence, Birds = Waterfowl presence [Binary]
* Water flow: Still, Slow, Steady, Surging [Categorical]
* Water level: Low, Average, High [Categorical]
* Water colour: Col.None = Colourless, Col.Yel = yellow, Col.Gre = green, Col.Bro = brown, Col.Oth = other [Binary]
* 'Algae' Algal presence: If water green or algae present = 1 [Binary]

## Environmental data

In addition to FreshWater Watch data

### Hydrobasin - WWF

Name: HydroBASINS

Description: HydroBASINS is a series of polygon layers that depict watershed boundaries and sub-basin delineations at a global scale. The goal of this product is to provide a seamless global coverage of consistently sized and hierarchically nested sub-basins at different scales (from tens to millions of square kilometres), supported by a coding scheme that allows for analysis of watershed topology such as up- and downstream connectivity.

Hackathon use: 'Hybas\_lv#+ID' Each sample is attributed to four levels of HydroBasin from large e.g. Thames (L4) to small (L10), including L6 and L8.

Version: 1.c with and without inserted lakes.

Spatial Resolution: 15 arc-seconds

Source: WWF

Citation: Lehner, B., Verdin, K., Jarvis, A. (2006): HydroSHEDS Technical Documentation. World Wildlife Fund US, Washington, DC. Available at <http://hydrosheds.cr.usgs.gov>.

### Esri – Terrain

Name: Terrain. Slope in Degrees

Description: This layer provides slope values calculated dynamically from the elevation data (within the current extents) using the server-side slope function applied to a [Terrain layer](http://www.arcgis.com/home/item.html?id=58a541efc59545e6b7137f961d7de883). The values are integer and represent the angle of the downward sloping terrain (0 to 90 degrees).  Note slope is a function of the pixel size of the request, so at smaller scales the slope values are smaller as pixel sizes increase.

Hackathon use: 'Slope' Slope of the overlapping pixel

Version: N.A. Data last modified 13/04/2017

Spatial Resolution: pixel resolution of 30 arc-second. The southern strip, falling in the Antarctica below latitude -60 degrees was not considered.

Source: Esri

Citiation: USGS, NGA, NASA, CGIAR, GEBCO,N Robinson, NCEAS, NLS, OS, NMA, Geodatastyrelsen, GSA and the GIS User Community

### Void Filled Digital Elevation Model – HydroSHEDS - WWF

Name: Global Land Cover SHARE (GLC-SHARE)

Description: Elevation layers distributed with HydroSHEDS. Created using a combination of the original SRTM-3 and DTED-1 elevation models of SRTM

Hackathon use: 'Altitude' Each sample is attributed an altitude

Version: Beta-Release Version 1.0-2014

Spatial Resolution: 3 arc-second, approximately 90m at the equator

Source: WWF

Citation: Lehner, B., Verdin, K., Jarvis, A. (2006): HydroSHEDS Technical Documentation. World Wildlife Fund US, Washington, DC. Available at <http://hydrosheds.cr.usgs.gov>.

### Land Use

Name: Global Land Cover SHARE (GLC-SHARE)

Description: The Global Land Cover-SHARE (GLC-SHARE) is a new land cover database at the global level created by Food and Agriculture Organization of the United Nations (FAO), Land and Water Division in partnership and with contribution from various partners and institutions.

Hackathon use: 10 categories; Each sample is attributed the percentage coverage of various land-uses from the nearest pixel. 'Dominant' refers to dominant land-class within pixel where 1= Artificial Surfaces, 2 = Cropland, 3 = Grassland, 4 = Trees, 5 = Shrubs, 6 = Aquatic herbs and vegetation, 7 = Sparse vegetation, 8 = Bare soil, 9 - SnowGlacier, 10 - Water bodies.

Version: Beta-Release Version 1.0-2014

Spatial Resolution: pixel resolution of 30 arc-second. The southern strip, falling in the Antarctica below latitude -60 degrees was not considered.

Source: Food and Agriculture Organization of the United Nations (FAO)

Citation: FAO Global Land Cover (GLC-SHARE) Beta-Release 1.0 Database, Land and Water Division, John Latham, Renato Cumani, Ilaria Rosati and Mario Bloise, 2014. <http://www.glcn.org/databases/lc_glcshare_en.jsp>

### Population density

Name: Gridded Population of the World – GPWv4

Description: The fourth version of GPW (GPWv4) is a gridded data product of globally-integrated national population data from the 2010 round of Population and Housing Censuses. The gridded data sets are constructed from national or subnational input administrative units

Hackathon use: 'popGPWv4\_dnt' Each sample is attributed population density from the nearest pixel

Version: Version 4

Spatial Resolution: pixel resolution of 30 arc-second (approx. 1Km at the equator)

Source: Center for International Earth Science Information Network - CIESIN - Columbia University

Citation: Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Documentation for the Gridded Population of the World, Version 4 (GPWv4). Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4D50JX4 Accessed XX/XX/XXXX.

### World Urban Areas

Name: World Urban Areas

Description: World Urban Areas represents the urban areas of the world with populations of more than 10,000.

Hackathon use: 'UrbArea' Each sample is classified into urban or rural (also used to identify nearest city see 1.1.1)

Version: N.A. Data updated every year

Spatial Resolution: < 100 metres

Source: UCLA Geoportal

Citation: World Urban Areas. Data and Maps for ArcGIS® 2014-07-01

### Biome

Name: Terrestrial Ecoregions of the World

Description: 825 terrestrial ecoregions of the globe. Ecoregions are relatively large units of land containing distinct assemblages of natural communities and species, with boundaries that approximate the original extent of natural communities prior to major land-use change.

Hackathon use: 'Biome' Each sample is classed into one of ten broad biomes

Version: 2.0

Spatial Resolution: approximately 0.1 meters at the equator

Source: WWF - US

Citation: Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'Amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y. Kura, J.F. Lamoreux, W.W. Wettengel, P. Hedao, and K.R. Kassem. Terrestrial Ecoregions of the World: A New Map of Life on Earth (PDF, 1.1M) BioScience 51:933-938.

### Protected Areas

Name: World Database on Protected Areas - WDPA

Description: The World Database on Protected Areas (WDPA) is the only global database of protected areas. It is a joint effort between IUCN and UNEP, managed by UNEP-WCMC, to compile protected area information for all countries in the world from governments and other authoritative organizations which are referred to as data providers.

Hackathon use: 'PrctArea' Each sample is attributed to either within or outside a protected area

Version: N.A.

Spatial Resolution: < 100 metres

Source: United Nations Environment Programme (UNEP) and the International Union for Conservation of Nature (IUCN).

Citation: UNEP-WCMC (2016). World Database on Protected Areas User Manual Cambridge, UK. Available at: <http://wcmc.io/WDPA_Manual>

### NDVI

Name: NDVI 1Km

Description: 10-daily Normalized Difference Vegetation Index. The product was generated by the land service of Copernicus, the Earth Observation program of the European Commission.

Hackathon use: Each sample is attributed the NDVI value for the pixel it overlaps. Available on request.

Version: V2.2

Spatial Resolution: 30 arc sec, around 1Km at the equator

Citation: The product was generated by the land service of Copernicus, the Earth Observation program of the European Commission. The research leading to the current version of the product has received funding from various European Commission Research and Technical Development programs. The product is based on PROBA-V 333m data ((c) ESA and distributed by VITO). <http://land.copernicus.vgt.vito.be>

### Rainfall and Air Temperature

Name: RNCEP-package

Description: This package of functions retrieves, organizes, and visualizes weather data from either the NCEP/NCAR Reanalysis or NCEP/DOE Reanalysis II datasets

Hackathon use: Each sample has a corresponding value for average rainfall and air temperature for the timeframes listed below

Version: 1.0.8 – 21/01/2017

Spatial Resolution: Surface

Temporal resolution: (Immediate, 24hr, 72hr) – RNCEP And Surface temp (24hr, 72hr) – RNCEP

Source: <https://sites.google.com/site/michaelukemp/rncep>

Citation: Kemp, M. U., Emiel van Loon, E., Shamoun-Baranes, J., Bouten, W., 2012. RNCEP: global weather and climate data at your fingertips. Methods in Ecology and Evolution (3), 65-70., DOI: 10.1111/j.2041-210X.2011.00138.x

## Engagement data

Two datasets are available. User-points related to individual events by any user that has generated points. Users is a summary of all data related to a specific user.

### General

* 'Staff' User is a member of staff
* 'HWP' Were the users trained as part of the HSNC Water Programme? [Binary]
* 'created' Date user created on platform
* 'Latest' Last active log in to profile page

### Training

* 'Attendees' Average number of people attending training days
* 'Paid' On a paid working day [Binary]
* 'TranTrai' Transport arranged to training location [Binary]
* 'TimeTrai' Length of travel to training location: 1 (short) to 3 (long) [Categorical]
* 'Upload' User uploaded results during training [Binary]
* 'Access' Ease of access to training venue via public transport: 1 (easy) to 3 (hard)

### Sampling protocol

* 'Time' Time needed to take sample
* 'Bulk' Bulkiness of equipment: 1 (light) to 5 (bulky)
* 'Complex' Complexity of equipment: 1 (straight forward) to 5 (complex)
* 'Task' Complexity of task: 1 (simple) to 5 (complex)
* 'Online' Earthwatch training delivered online only [Binary]
* 'Team' Worked as part of team [Binary]
* 'Self' Team was self-selecting [Binary]
* 'People' Number of people per team [Integer]
* 'Assign' Sites were assigned [Binary]
* 'Home' Proximity of sample sites to home: 1 (close) to 3 (far) [Categorical]
* 'Coord' Testing days coordinated by the PI [Binary]
* 'TranSamp' Availability of public transport to sampling venue: 1 (plentiful) to 3 (difficult)
* 'TimeSamp' Time taken to get to sampling venue

### Engagement markers

* 'EventDate' Date of engagement marker [Date]

#### (Column 'Type'; category)

* 'Sample' Upload sample
  + 'Upload' Method of upload: 1 = Website, 2 = App. [Categorical] - only applies to Type 'Sample'
* 'Blog' Blog writing on FWW platform
* 'Comment' Blog comment on FWW platform [Date and time]
* 'Pres' Gave a presentation
* 'Forum' Used the FWW platform forums
* 'Quiz' Completed the FWW quiz
* 'Share' Shared via social media
* 'Invite' Invited someone else to get involved
* 'InviteA' Invited a user who then accepted
* 'Foot' Completed the water footprint calculator
* 'Del' Deleted a sample

### Temporal

* 'Sample Date'
* 'Month'
* 'Year'
* Season
  + Temperate: 1 = spring, 2 = summer, 3 = autumn, 4 = winter
  + Tropical: 1 = Premonsoon, 2 = monsoon, 3 = postmonsoon, 4 = winter

### Weather conditions on sampling day

* Rainfall and air temperature: See

### Environmental quality

* 'WQO' Observed water quality (visual cues e.g. algal blooms)
* 'WQC' Catchment quality (e.g. urban land-use)
* 'WQM' Measured quality (e.g. chemical)
* 'WQS' Observed surface pollution (e.g. litter)