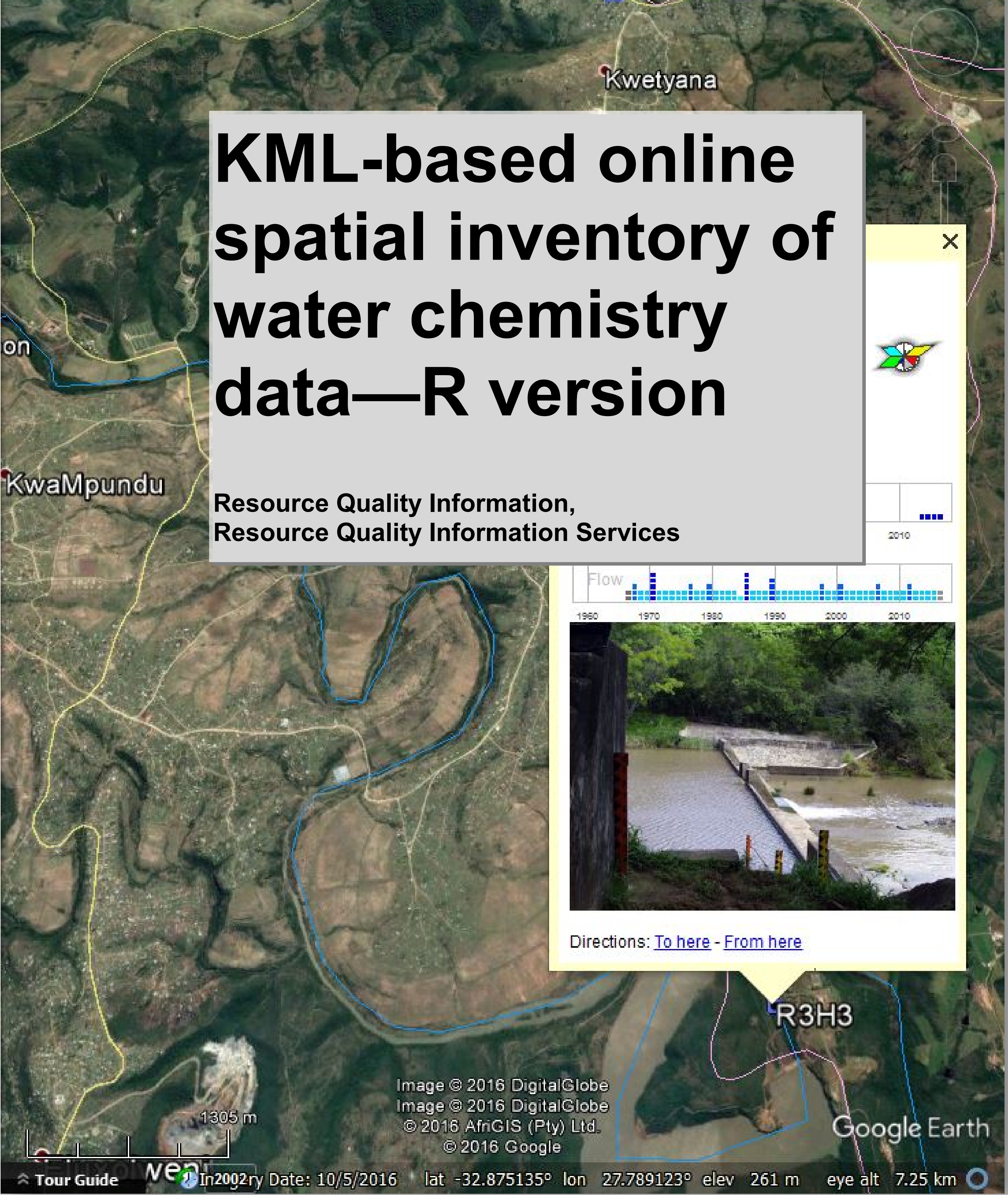
KML-based online spatial inventory of water chemistry data—R version

Resource Quality Information,
Resource Quality Information Services



water & sanitation
Department:
Water and Sanitation
REPUBLIC OF SOUTH AFRICA

DOCUMENT STATUS

WORKING TITLE:	KML-based online spatial inventory of water chemistry data—R version (Conversion to R of the application that creates the online spatial inventory of water chemistry data)
AUTHOR:	Michael Silberbauer
REPORT STATUS:	**Second draft**
RQS REPORT NUMBER:	N/0000/00/RD1/2011
DATE:	January 2017

Executive summary

When Google Earth became publicly available in 2005, it provided an ideal platform for an inventory of the South African national water quality database, Water Management System. The first working prototype of a Google Earth driven spatial inventory comprised a set of ArcInfo-AML, awk and DOS scripts for converting data from the Informix data base to flat text files, then processing these to produce the final set of data plots, data listings and Keyhole Markup Language spatially-referenced XML (KML) files for Google Earth (Silberbauer and Geldenhuys, 2008 and 2009).

In 2011, the Unix-based ArcInfo scripts were approaching obsolescence and the software was ported to R. The R package has a wide range of statistical functions and graphics abilities, including geographical representation of data. R provides many more opportunities for further development on a variety of platforms, including web-based applications. The package is open-source, extensible and licensing is free.

The main components of the Google Earth inventory are the KML files themselves, which contain locality, time and attribute information, the multivariate time-series plots, the data listings and the Maucha multivariate point ionic symbols. This report describes the purpose, structure and operation of the application that creates the KML monitoring site files, kml_WMS_mnpts_html.R. This R script consists of 18 functions that deal with text formatting, and the creation of each KML file with its associated HTML web file. A key component is the popup balloon that displays attribute information when the user clicks on a site, and which provides links to further information, such as time series plots, source data and the Department of Water and Sanitation's hydrological database.

Introduction

In 2005, Google released a free online version of the Keyhole Earth Viewer package under the name Google Earth™. This software uses advanced techniques to speed up the laborious process of combining different sources of spatial data and rendering them in a perspective, 3D view the Earth's surface. Real-time landscape fly-throughs become practical on home computers (Crampton 2008). The most important component of the system for practical applications is the extensible markup language, XML, called Keyhole Markup Language or KML (Google 2011). This allows users to simply and quickly place their own spatial data on the Google Earth landscape without the need for advanced programming. Google has made many enhancements to the software, such as the ability to time tag user data and to splay out overlapping points. Many new datasets have become available and the resolution of the satellite data acquired by Google allows the identification of structures such as weirs and sewage works in many areas. Streetview data became available for South Africa in 2010 and provided an even closer perspective of the situation on the ground. Streetview images from bridges provide an invaluable perspective of the state of rivers and riparian habitat at the time of the photograph. Built-in point-to-point driving directions are of great assistance when planning fieldwork.

This report describes the `kml_WMS_mnpts_html.R` script in detail. The purpose of the report is twofold: to aid in maintenance of the script and to provide nuts-and-bolts information for anyone wanting to port the process to another platform.

Methods

The software consists of a main routine to generate eight groups of files (Table 1), plus several functions to handle repetitive processes or to make the code simpler to understand and maintain. The version discussed here is 5.9.

Table 1. The eight types of KML inventory file created by the spatial inventory system.

	Surface water sites		Groundwater sites	
	Primary drainage regions	Water management areas	Primary drainage regions	Water management areas
Marker symbols	A to X, + 'O', 'Y', 'Z'	1 to 19	A to X, + 'Y', 'Z'	1 to 19
Maucha symbols	A to X, + 'O', 'Y', 'Z'	1 to 19	A to X, + 'Y', 'Z'	1 to 19

The software environment chosen is primarily the R statistics system (R 2016) with the database package `RODBC`. The advantages of R are that it has no licence fee, has a large and helpful user community, provides a range of statistical analyses and has excellent graphics capabilities. Developers familiar with other programming and scripting languages could achieve the same end results using software that they are more familiar with. Formerly, an ArcInfo AML script converted a dBase inventory file from the Water Management System to a list of sites with information on drainage regions and management areas, but this process was replaced by an R script, `wms2nms.R`, in 2012. This script performs geographical point-in-polygon checks to ensure that each monitoring site is in the correct drainage region and water management area, and outputs a shapefile and a text list of site information. The system still uses the 19 old water management areas because the new areas are larger, with many more sites, which may cause a memory problem on older PCs and portable devices.

The software needs to run on a computer that has an ODBC link to the WMS database. The link should preferably be read-only, to avoid any unfortunate accidents.

Main routine

```
library(foreign)
library(RODBC)
library(RCurl)
odbcCloseAll()
if (!exists ("monitoringSites", mode="function") ) source
("C:/data/program/R/monitoringSites.R")

options(stringsAsFactors = FALSE) # not working with factors so would rather
have strings
time.start <- as.POSIXct(format(Sys.time(), "%Y-%m-%d %H:%M:%S"))
chem.333 <- FALSE # switch to produce output for the top 333 sites only

# file.nms <- "C:/data_large/av/nms_wms_geo.dbf"
# nms <- read.dbf(file.nms, as.is=TRUE) # as.is=FALSE reads characters as
factors
# s <- nms[c("FEAT_ID", "STATION", "QUATERNARY", "QAT_WMS", "WMANUM",
"LOC_TYPE", "FIRSTDATE", "LASTDATE", "N", "DESCRIPTIO", "LATITUDE",
"LONGITUDE")]
# names(s) <- c("PointID", "ID",  "Qat", "QAT_WMS", "WMANUM",
"Located_On_Type", "FirstDate", "LastDate", "No", "Name", "Latitude",
"Longitude") # names used in this script

s <- monitoringSites()
 s <- subset(s, s$No > 0)
s$ID[s$ID == ""] <- NA # is this wise? or even necessary?
s$WMANUM[s$WMANUM == ""] <- 0 # missing WMA number
s$WMANUM <- as.numeric(s$WMANUM)
s$WMANAME[s$WMANAME == ""] <- "WMA_unknown" # missing WMA name
s <- within (s,
             Located_On_Type <- ifelse(!is.na(Located_On_Type),
Located_On_Type, "Class pending")
             # Thanks to Joris Meys,
http://stackoverflow.com/questions/7488068/test-for-na-and-select-values-
based-on-result
)

if(chem.333) {
  # read the sites from the WMS:
  channel<-odbcConnect("wmsdb")
  wmstables<-sqlTables(channel)

  q <- paste("SELECT UNIQUE mon_feature_id ",
             " FROM programme_sample ",
             " WHERE mon_program_id = 146")
  programme.sites <- sqlQuery(channel, q)

  print("*** Only processing \"CHEM 333\" sites ***")
  s <- s[s$PointID %in% programme.sites$mon_feature_id, ]
}
msym <- FALSE

tokml("PDG", s, msym, chem.333) # primary drainage regions, no Maucha

if(!chem.333)
  tokml("WMA", s, msym, chem.333) # water management areas, no Maucha

msym <- TRUE

if(!chem.333) for (regiontype in c("PDG", "WMA")) {
  # primary drainage regions or water management areas, with Maucha option
  tokml(regiontype, s, msym, chem.333)
}
odbcCloseAll()
time.end <- as.POSIXct(format(Sys.time(), "%Y-%m-%d %H:%M:%S"))
print(paste0("Started: ", time.start) )
print(paste0("Finished: ", time.end, " (", time.end - time.start, ")") )
print(paste("chem.333 =", chem.333))
```

The main routine runs the function `msym` once each for primary drainage regions (PDG), water management areas (WMA), first without and then with Maucha ionic diagrams as symbol markers. It runs a separate function for reading monitoring sites prepared by `wms2nms.R`. `s <- monitoringSites();  s <- subset(s, s$No > 0)`
The main routine also has a seldom-used option for updating only the national chemical monitoring programme sites: `chem.333 <- TRUE`.

Main controlling function

```
tokml <- function(regiontype, s, msym, chem.333) { # main controlling function - used to
be the main routine
  #library(RSQLite)
  library(RODBC)
  channel<-odbcConnect("wmsdb")
  wmstables<-sqlTables(channel)

  #q<-paste("SELECT PointID, ID, RefCode, Region, WMANUM, Located_On_Type, No FROM inv2",
sep="") # the WMANUM not available from WMS
  #q<-paste("SELECT PointID, ID, Qat, WMANUM, Located_On_Type, No FROM inv2", sep="") #
the WMANUM not available from WMS
  #s<-dbGetQuery(con,q)
  #s$WMANUM<-as.numeric(s$WMANUM)

  LoGso <- c("river", "dam_lake", "spring", "wetland", "estuary_sea",
             "watersupply", "wastewater", "mine_industry", "agri", "transfer",
             "class_pending", "meteo", "ground")
  if(chem.333) LoGso <- LoGso[LoGso != "ground"]

  # seems to be a mis-reference of ID and RefCode - watch out when linking to WMS via
ODBC:
  if (regiontype=="PDG") s <- s[order(s$Qat, s$ID, s$PointID), ]

  #  Note that WMS Region is the WMS quaternary region - not right
  if (regiontype=="WMA") s <- s[order(s$WMANUM, s$Qat, s$ID, s$PointID), ]

  s$pri <- substr(s$Qat, 1, 1) # Define primary drainage regions from quaternary
  s$sec <- substr(s$Qat, 1, 2) # Define secondary drainage regions from quaternary
  LoTs <- subset(s$Located_On_Type, !duplicated(s$Located_On_Type)) # List of site types
  s <- Located_On_Group(s)
  LoGs <- subset(s$Located_On_Group, !duplicated(s$Located_On_Group)) # List of site
groups
  if (regiontype=="PDG") {
    pris <- s$pri[!duplicated(s$pri)]
  }
  if (regiontype=="WMA") {
    pris <- s$WMANUM[!duplicated(s$WMANUM)]
    pris <- pris[!is.na(pris)]
  }

  for (pr in pris) { # for each primary or management region, create a new KML file,
including a Maucha symbol version
    # for (pr in c("A")) { # test line for limited data set - crashes though:
    # Error in pr + 1 : non-numeric argument to binary operator
    if (regiontype=="PDG") spr <- subset(s, s$pri==pr)
    if (regiontype=="WMA") spr <- subset(s, s$WMANUM==pr)
    KMLhead (pr, spr, regiontype, msym, chem.333)
    if(!msym) HTMhead (pr, spr, regiontype, chem.333)
    ltys<-subset(spr$Located_On_Type, !duplicated(spr$Located_On_Type)) # List of site
types
    #lgps<-subset(spr$Located_On_Group, !duplicated(spr$Located_On_Group)) # List of site
groups
    for (lgp in LoGso) { # for each group in predetermined sort order
      secs <- 0
      lgps <- subset(spr$Located_On_Group, !duplicated(spr$Located_On_Group)) # List of
site groups
      if(lgp %in% lgps) {
        secs <- subset(spr$sec, !duplicated(spr$sec))
        GRPhead(lgp, pr, msym, regiontype, chem.333)
        for (se in secs) {
          sse <- subset(spr, spr$Located_On_Group==lgp & spr$sec==se)
          #sse<-sse[order(sse$RefCode,sse$PointID,na.last=TRUE),] # put those with a
hydro number first
          # put those with a hydro number first: (no use, messes up catchment order)
          sse <- sse[order(sse$sec, sse$ID, sse$PointID, na.last=TRUE),]
          if(nrow(sse) > 0) {
            SEChead(pr, lgp, se, sse, msym, chem.333)
            # for each item, create an entry in the KML file and HTML file
            SECbody(pr, lgp, sse, channel, msym, chem.333)
            SECtail(pr, se, lgp, msym, chem.333)
          }
        }
        GRPtail(lgp, pr , msym, chem.333)
      }
    }
    KMLtail(pr, msym, chem.333)
    if(!msym) HTMtail(pr, chem.333)
  }
}
```

The main controlling function opens the Water Management System (WMS) database using open database connectivity (ODBC). The WMS has about 40 types of monitoring site, and in order to simplify the map legend, these are grouped into 12 categories (Table 2).

Table 2. Grouping of Located_On_Type into classes.

Group	Classes of Located_On_Type
class_pending	Class pending, Formal, unknown
meteo	Meteorology
ground	Borehole, Other Ground Fractures, Sinkhole, Dug Well, Well Point, Excavation - Quarry
transfer	Pipeline, Canal, Pump Station, Unknown Transfer Feature Type, Storm Water System, Tunnel, Lateral Collector
watersupply	Potable Water Treatment Works
mine_industry	Mine Property, Mineral Process Plant/Area, Industrial Property, Slime/Slurry Dam, Evaporation Dam, Containment/Emergency/Return Water Dam, Mine Shaft, Intensive Livestock/Irrigated Cropping
river	Rivers
spring	Spring/Eye
dam_lake	Dam / Barrage, Reservoir, Lake
estuary_sea	Estuary/Lagoon, Ocean / Bay
wastewater	Waste Water Treatment Works, Solid Waste Transfer Site, Water/Effluent Treatment Plant, Oxidation Pond
wetland	Pan, Wetland, Nature Site

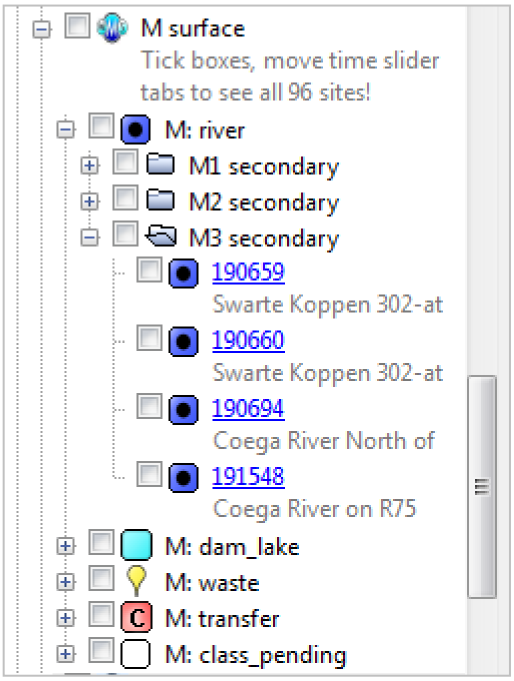


Figure 1. The hierarchical table of contents in Google Earth.

The controlling function processes each region in turn, i.e. primary drainage regions A to X and water management areas 1 to 19. Groundwater is dealt with separately because of the large number of sites, each with only one or a few records. The first step in the procedure is to call the function that creates the headers of the output files in KML and HTML format. Within each region, the function processes the 12 groups of sites in turn, further subdividing them by secondary drainage region in order to create a navigable hierarchy in the Google Earth table of contents (Figure 1). Within each secondary drainage region, the controlling function calls the `SECbody` function to locate, format and write the information for each site to the KML and HTML files.

Creating KML files: KMLhead, GRPhead, SEChead, SECbody, SECtail, GRP tail

Keyhole Markup Language, KML, is an extensible markup language (XML) for presenting spatial data with an optional time component. The coordinate system allows for portrayal of objects on or above the Earth's surface.

KMLhead

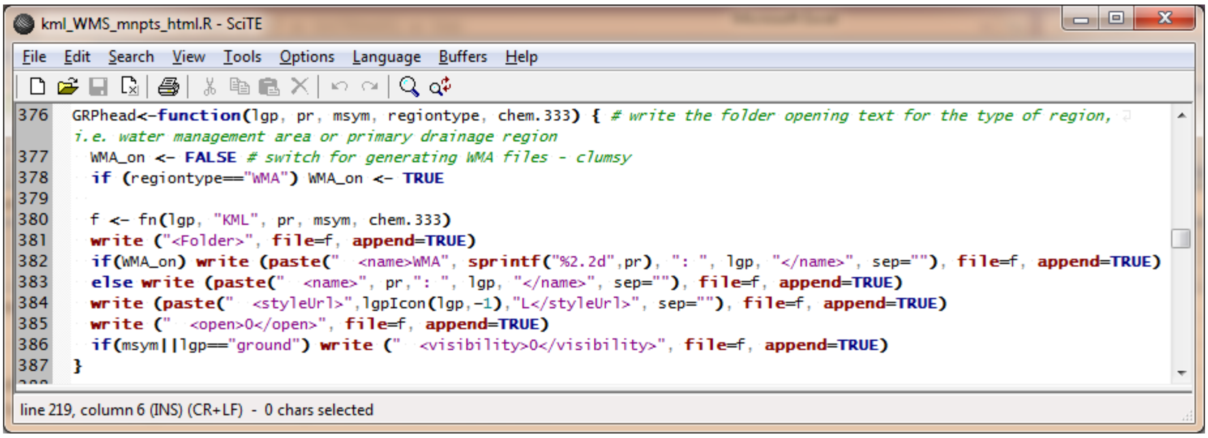
kml_WMS_mnpts_html.R - SciTE

```
KMLhead <- function(pr, spr, regiontype, msym, chem.333) { # write the header of the extensible markup language (XML) keyhole markup language (KML) file
  icons <- ico()
  dwa <- "Department of Water and Sanitation"
  WMA_on <- FALSE # switch for generating WMA files - clumsy but it works
  if (regiontype == "WMA") WMA_on <- TRUE
  if (WMA_on) reg <- wmaname()[pr+1] else reg <- pr
  print(paste("***", format(Sys.time(), "%Y-%m-%d %H:%M:%S"), "KML ",reg, "region", nrow(spr), "monitoring sites", mtxt(msym) , "***"))
  for(lgp in c("ground","surface")) {
    f <-fn (lgp,"KML", pr, msym, chem.333)
    if(lgp == "ground") rows <- nrow(subset(spr, spr$Located_On_Group == "ground"))
    else rows <- nrow(subset(spr, spr$Located_On_Group != "ground"))
    write ("<?xml version=\"1.0\" encoding=\"UTF-8\"?>", file=f)
    write ("<kml xmlns=\"http://earth.google.com/kml/2.2\"\n  xmlns:atom=\"http://www.w3.org/2005/Atom\">", file=f, append=TRUE)
    write (paste("<Document>\n<atom:author>\n  <atom:name>", dwa, "</atom:name>\n</atom:author>", sep=""), file=f, append=TRUE)
    write ("\n<atom:link href=\"https://www.dwa.gov.za/iwqs/wms/data/000key.asp\" />\n", file=f, append=TRUE)
    if(WMA_on) write (paste("<name>", wmaname()[pr+1], lgp, "water", dwa, "</name>"), file=f, append=TRUE)
    else write (paste("<name>", pr, lgp, "water", dwa, "</name>"), file=f, append=TRUE)
    write (paste("<Snippet maxLines=\"2\">Tick boxes, move time slider tabs to see all", rows, "sites!</Snippet>"), file=f, append=TRUE)
    write ("<Style id=\"wmsPlacemark\">", file=f, append=TRUE)
    write ("<ListStyle>", file=f, append=TRUE)
    write ("<ItemIcon>", file=f, append=TRUE)
    write ("    <href>https://www.dwa.gov.za/iwqs/wms/data/wms_m.jpg</href>", file=f, append=TRUE)
    write ("</ItemIcon>", file=f, append=TRUE)
    write ("</ListStyle>", file=f, append=TRUE)
    write ("</Style>", file=f, append=TRUE)
    write ("<styleUrl>#wmsPlacemark</styleUrl>", file=f, append=TRUE)
    if(msym) mkey(f)
    for (ic in 1:nrow(icons))   {
      write (paste("<Style id=\"", icons[ic,1], "\">", sep=""), file=f, append=TRUE)
      write ("  <IconStyle><scale>0.4</scale>", file=f, append=TRUE)
      write ("    <Icon>", file=f, append=TRUE)
      write (paste("      <href>", icons[ic,2], "</href>", sep=""), file=f, append=TRUE)
      write ("    </Icon>", file=f, append=TRUE)
      write ("  </IconStyle>", file=f, append=TRUE)
      write ("  <BalloonStyle>", file=f, append=TRUE)
      write ("  <bgColor>ccffff</bgColor>", file=f, append=TRUE)
      write ("    <text><![CDATA[", file=f, append=TRUE)
      write ("    <font  face=\"Arial\" color=\"#666666\" size=\"+3\"><b>$[name]</b></font><br />",
             file=f, append=TRUE)
      write ("    <font  face=\"Verdana\">$[description]</font><br />", file=f, append=TRUE)
      #write ("    <small>(Links may be slow to open or offline.)<br />", file=f, append=TRUE)
      write ("    $[geDirections]</small>", file=f, append=TRUE)
      write ("    ]]></text>", file=f, append=TRUE)
      write ("  </BalloonStyle>", file=f, append=TRUE)
      write ("</Style>", file=f, append=TRUE)
    }
    for (ic in 1:nrow(icons)) {
      write (paste("<Style id=\"", icons[ic,1], "L\">", sep=""), file=f, append=TRUE)
      write ("  <ListStyle>", file=f, append=TRUE)
      write ("    <ItemIcon>", file=f, append=TRUE)
      write (paste("      <href>", icons[ic,2], "</href>", sep=""), file=f, append=TRUE)
      write ("    </ItemIcon>", file=f, append=TRUE)
      write ("  </ListStyle>", file=f, append=TRUE)
      write ("</Style>", file=f, append=TRUE)
    }
    write ("<open>1</open>", file=f, append=TRUE)
  }
}
```

line 174, column 1 (INS) (CR+LF) - 0 chars selected

A KML file begins with a standard header that declares the file type and version, and defines styles for placemarks (Google 2011). Function `KMLhead` writes these lines, with loops to set up definitions of the icons used on maps and in the table of contents. The script creates output files line by line using write statements, harking back to the earlier awk script on which it is based (Silberbauer 2009 Ch4).

GRPhead

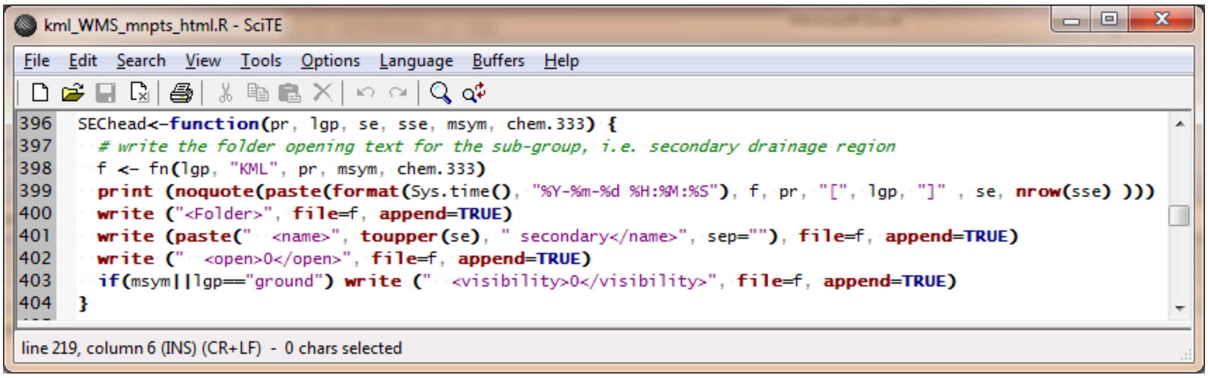


```
GRPhead<-function(lgp, pr, msym, regiontype, chem.333) { # write the folder opening text for the type of region, i.e. water management area or primary drainage region
  WMA_on <- FALSE # switch for generating WMA files - clumsy
  if (regiontype=="WMA") WMA_on <- TRUE

  f <- fn(lgp, "KML", pr, msym, chem.333)
  write ("<Folder>", file=f, append=TRUE)
  if(WMA_on) write (paste("  <name>WMA", sprintf("%2.2d",pr), ": ", lgp, "</name>", sep=""), file=f, append=TRUE)
  else write (paste("  <name>", pr,": ", lgp, "</name>", sep=""), file=f, append=TRUE)
  write (paste("  <styleUrl>",lgpIcon(lgp,-1),"L</styleUrl>", sep=""), file=f, append=TRUE)
  write ("  <open>0</open>", file=f, append=TRUE)
  if(msym||lgp=="ground") write ("  <visibility>0</visibility>", file=f, append=TRUE)
}
```

Within the KML file, `<Folder>` statements define the hierarchy in the table of contents (Figure 1) and the `GRPhead` function writes the appropriate heading for a site group class (Table 2). The folder heading includes the style instruction for displaying the appropriate group icon and `<open>0</open>` combination that specifies that the hierarchy will be in a "closed" or compact form when first displayed. If groundwater sites are to be displayed, the tick boxes are left open so that the user can choose which sites to switch on. The groundwater sites are numerous and can overload a computer with insufficient memory.

SEChead



```
SEChead<-function(pr, lgp, se, sse, msym, chem.333) {
  # write the folder opening text for the sub-group, i.e. secondary drainage region
  f <- fn(lgp, "KML", pr, msym, chem.333)
  print (noquote(paste(format(Sys.time(), "%Y-%m-%d %H:%M:%S"), f, pr, "[", lgp, "]" , se, nrow(sse) )))
  write ("<Folder>", file=f, append=TRUE)
  write (paste("  <name>", toupper(se), " secondary</name>", sep=""), file=f, append=TRUE)
  write ("  <open>0</open>", file=f, append=TRUE)
  if(msym||lgp=="ground") write ("  <visibility>0</visibility>", file=f, append=TRUE)
}
```

The `SEChead` function writes the `<Folder>` statements for a set of sites in a single secondary drainage region. As in the group header, the hierarchy is "closed" initially, and the user can open it up as in Figure 1. Again, the groundwater sites are not "on" by default.

SECbody

In `SECbody`, the processing of the individual sites within a secondary drainage region takes place. For each row in the site list passed from the main controlling function, the `SECbody` function outputs descriptive information and coordinates. `SECbody` checks the WMS for electrical conductivity data for the site and takes the median as a gross indicator of salinity and thus water quality.

The function works out a compact tag for labelling each site. If no site ID is present, the function uses the feature code, with anything more than 4 sequential zoeroes ("`0000`") replaced by "-" with `gsub`. If the site ID is a hydrological code like `A2H027Q01` (McDonald 1989), the function checks whether it is a river site, a dam site or a treatment works. In the case of a borehole code like `2918AC00056`, the 1:50 000 topographical map portion is deleted, leaving `56` in this example (Table 3).

Table 3. Examples of map labels generated in SECbody.

Site ID type	Example of full site ID	Truncated map label
Rivers	A2H027Q01	A2H27
Dam	A2R009Q02	A2R9.2
Water treatment works	R3H002S01	R2H2s1
Rivers – no hydro code	1000010300	1-10300
Borehole	2918AC00056	56

Each site entry in the KML file contains an extended label which the user accesses by clicking on the site icon. The label or pop-up balloon consists of 19 or more items, where applicable for the type of site (Figure 2).

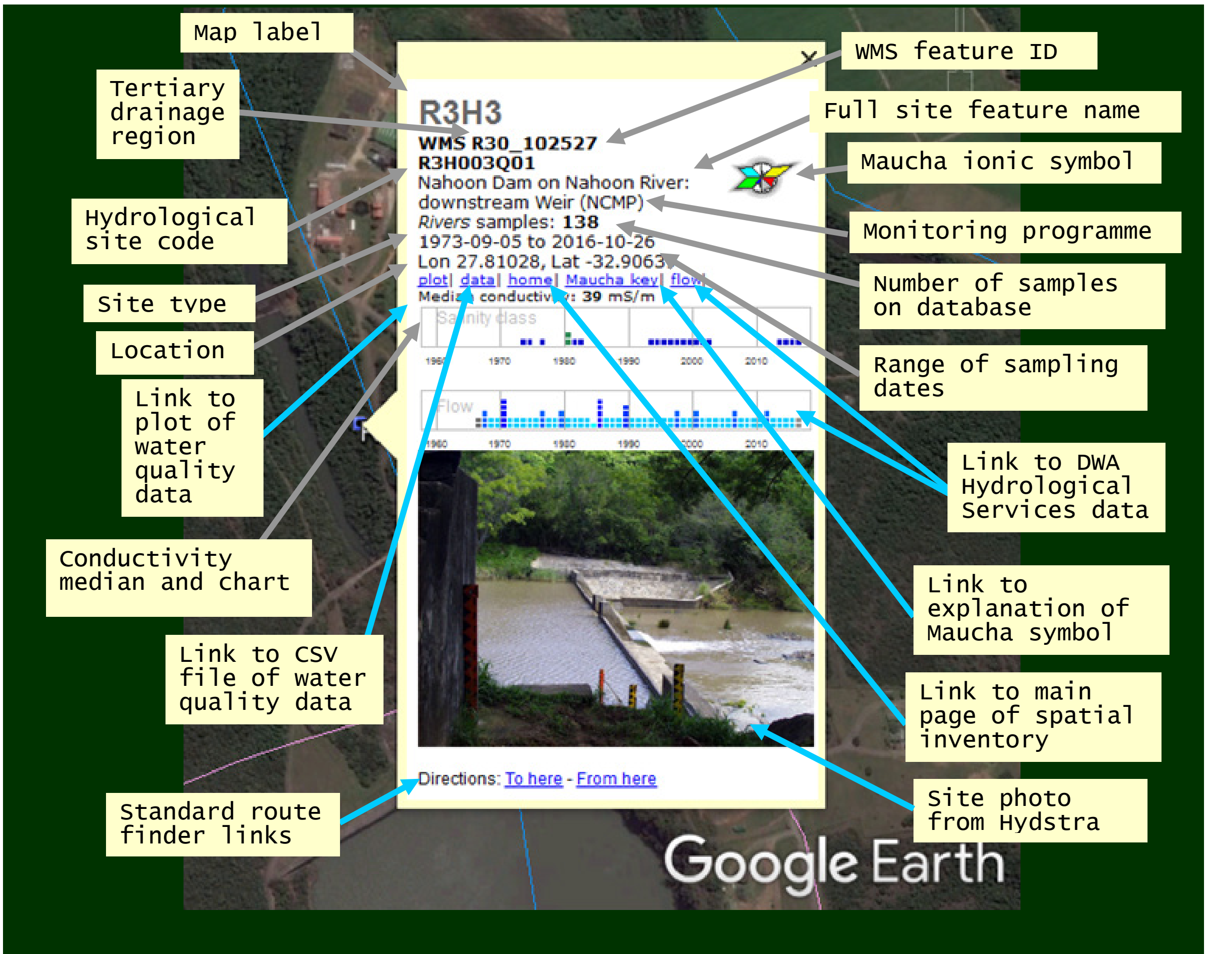


Figure 2. A typical pop-up information balloon.

The following sections explain the `SECbody` code for each item within a pop-up balloon. The sequence is that of the balloon text, not necessarily that of the R code. The body of the balloon text is in HTML format, enclosed in the KML description element and the standard XML CDATA element:

```
<description><![CDATA[ HTML text here ]]></description>
```

Map label

`SECbody` reformats the site label based on its type and writes the name at the start of the pop-up balloon:

```
write ("<Placemark>", file=f, append=TRUE)
write (paste("<name>", st, "</name>", sep=""), file=f, append=TRUE)
```

Tertiary drainage region

The tertiary drainage region consists of the first three characters of the quaternary drainage region:

```
Ter<-substr(s$Qat,1,3)
```

Embedded graphics

This code in `SECbody` compiles the filenames of the Maucha diagram, and miniplots of electrical conductivity, hydrology and chlorophyll *a* on the local disk, and the corresponding URL on the Internet:

```
 png.maucha <- paste0("C:/tmp/wms/", Ter, "/", Ter, "_", PointID, "_m.png")
 png.ecplot <- paste0("C:/tmp/wms/", Ter, "/", Ter, "_", PointID, "_p.png")
 png.fvplot <- paste0("C:/tmp/wms/", Ter, "/", Ter, "_", PointID, "_f.png")
 png.chplot <- paste0("C:/tmp/wms/", Ter, "/", Ter, "_", PointID, "_e.png")
 imglink.maucha <- paste0("https://www.dwa.gov.za/iwqs/wms/data/", Ter, "/", Ter,
"_", PointID, "_m.png")
 imglink.ecplot <- paste0("https://www.dwa.gov.za/iwqs/wms/data/", Ter, "/", Ter,
"_", PointID, "_p.png")
 imglink.fvplot <- paste0("https://www.dwa.gov.za/iwqs/wms/data/", Ter, "/", Ter,
"_", PointID, "_f.png")
 imglink.chplot <- paste0("https://www.dwa.gov.za/iwqs/wms/data/", Ter, "/", Ter,
"_", PointID, "_e.png")
```

If any file is present on the local disk, the code assumes that the operator has also uploaded it to the Internet:

```
if(file.exists(imgFile)) write(paste("<img src=\"", imgLink, "\" align=\"right\"
/>", sep=""), file=f, append=TRUE)
```

It also sets the URL of the Hydstra photo, for checking further on:

```
 img.hispic <- paste0("https://www.dwa.gov.za/hydrology/CGI-BIN/HIS/Photos/", sta,
".JPG")
```

The following scripts need to run first in order for the embedded graphics check to work: `NCMP_miniplots.R`, `Maucha_per_site_chubby.R`, `NEMP_miniplots.R`.

WMS feature_id

`SECbody` outputs the tertiary drainage region and the `feature_id` linked by an underscore:

```
write (paste("<b>WMS ", toupper(Ter), "_", PointID, "</b><br />", sep=""), file=f,
append=TRUE)
```

Site code using numbering convention

If the site has a code of the type used in column two of Table 3, `SECbody` outputs it:

```
write (paste(ref, "<br />", sep=""), file=f, append=TRUE)
```

WMS feature_name

The script writes the mixed case site name to the table of contents entry:

```
write (paste("<Snippet maxLines=\"1\">",s$Name, "</Snippet>", sep=""), file=f,
append=TRUE)
```

and in the balloon text:

```
write (paste(Name,"<br />"), sep="", file=f, append=TRUE)
```

Site type and number of samples

`SECbody` combines the site type and number of samples on a single line:

```
write (paste("<i>", s$Located_On_Type[1], "</i> samples: <b>", s$No[1], "</b><br
/>", sep=""), file=f, append=TRUE)
```

Electrical conductivity

`SECbody` queries the WMS and calculates the median conductivity:

```
vars <- sqlQuery(channel, "select * from monitorng_variable")
    varEC <- subset(vars$mon_variable_id, grepl("EC-Phys-Water",
                    vars$mon_variable_abbr))
    q <- paste("SELECT result_num_value FROM released_result WHERE
                    mon_variable_id = ", varEC,
                   " AND mon_feature_id = ", PointID, sep="")
c <- sqlQuery(channel, q)
    c <- na.omit(c)
if(nrow(c) > 0) medEC <- round(median(c$result_num_value))
    else medEC <- NA
```

If a valid median conductivity value is available, the script outputs the data, colouring the text red if the value is greater than or equal to 350 mS/m:

```
if (is.na(medEC)) print (paste(stas,"- no EC data")) else
        {
         if(s$No[1]==1)  write ("<small>Conductivity: ", file=f, append=TRUE)
         if(s$No[1]>1) write("<small>Median conductivity: ", file=f, append=TRUE)
         if (medEC<350) write (paste("<b>", medEC, "</b> mS/m</small><br />",
sep=""), file=f, append=TRUE)
            else write (paste("<b><font color=\"#FF0000\">", medEC, "</font></b>
mS/m</small><br />", sep=""), file=f, append=TRUE)
        }
```

Range of sampling dates

The first and last sampling dates come from `sse$FirstDate` and `sse$FirstDate`.

```
if (No == 1) write (paste(FirstDate, "<br />", sep=""), file=f, append=TRUE)
    else write (paste(FirstDate, " to ", LastDate, "<br />", sep=""),
                file=f, append=TRUE)
```

The first and last sampling dates timestamp the point as well:

```
if(No==1) {
      write ("<TimeStamp id=\"ID\">", file=f, append=TRUE)
      write (paste("<when>", FirstDate, "</when>", sep=""), file=f, append=TRUE)
      write ("</TimeStamp>", file=f, append=TRUE)
    }
    else {
      write ("<TimeSpan id=\"ID\">", file=f, append=TRUE)
      write (paste("  <begin>", FirstDate, "</begin>", sep=""), file=f,
append=TRUE)
      write (paste("  <end>", LastDate, "</end>", sep=""), file=f, append=TRUE)
      write ("</TimeSpan>", file=f, append=TRUE)
    }
```

Coordinates

The site coordinates come from `sse$Latitude` and `sse$Longitude`.

```
sw<-sqlQuery(channel, qw)
Latitude<-sw$point_latitude[1]
Longitude<-sw$point_longitude[1]
[…]
write (paste("Lon ", signif(Longitude,7), ", Lat ", signif(Latitude,7), "<br />",
sep=""), file=f, append=TRUE)
```

Link to time-series plot of water quality

A separate script, barcode.R, pre-generates a set of time-series plots of inorganic solutes, conductivity and pH from the WMS database (Silberbauer 2017). The files are currently in PNG format, because the more crisp PDF files are about ten times larger than the PNG files. SECbody checks for the existence of a plot file on the local disk and, if it is available, assumes that the operator has already uploaded it to the Internet, as described under "Embedded graphics".

Link to time-series water quality data

At the same time as it generates the time-series plot for a site, the barcode.R script creates a compressed comma-separated value (CSV) file with the data used in the plot. SECbody checks for the existence of the compressed ZIP file on the local disk and, if it is available, assumes that the file is also available on the Internet:

```
zipFile<-paste("C:/tmp/wms/",Ter,"/",Ter,"_",PointID,".zip", sep="")
zFile<-unlist(strsplit(fn(lgp,"HTM",pr, msym),"/"))
zLink<-paste("./",zFile[length(zFile)], sep="")  # take HTML file from path
zipLink<-paste("http://www.dwa.gov.za/iwqs/wms/data/",Ter,"/",Ter,"_",PointID,
".zip", sep="")
if(file.exists(zipFile)) write(paste("<a href=\"", zipLink, "\">data</a>|",
sep=""), file=f, append=TRUE)
```

Link to hydrological data

If the site code has the characteristics of a hydrological station as shown in Table 3, SECbody generates a link to the Department of Water and Sanitation Hydrology server. Note that from time to time the Hydstra managers may update their site and the link format may change or become inaccessible. This is the code to update if that happens.

```
    if(his) {
      fLink <- paste(
        "\"https://www.dwa.gov.za/hydrology/Verified/HyDataSets.aspx?Station=",
        sta, "\"", sep="")
      write(paste("<a href=", fLink, ">flow</a>|", sep=""), file=f, append=TRUE)
    }
```

Link to home page of spatial inventory

All pop-up balloons have a link back to the home page of the inventory site. Bear this in mind if any change to the structure of the web site occurs.

```
write( "<a href=\"http://www.dwa.gov.za/iwqs/wms/data/000key.htm\">home</a>|",
file=f, append=TRUE)
```

Link to explanation of Maucha diagram

If the pop-up balloon contains a Maucha diagram, this link takes the user to an explanation of the origins and meaning of the diagram:

```
MauchaLink<-"https://www.dwa.gov.za/iwqs/gis_apps/maucha.pdf"
    if(file.exists(png.maucha)) {
      write(paste("<a href=\"", MauchaLink, "\">Maucha key</a>|", sep=""),
            file=f, append=TRUE)}
```

If the script is generating a KML with Maucha symbols at the monitoring points, the following lines produce the required style code:

```
if(msym && file.exists(imgFile)) {
    write (paste("<Style id=\"",st,"\"><IconStyle><scale>0.67</scale>", sep=""),
file=f, append=TRUE)
    write (paste("
<Icon>http://www.dwa.gov.za/iwqs/wms/data/",Ter,"/",Ter,"_",PointID,"_m.png</Icon>"
, sep=""), file=f, append=TRUE)
```

```
    write ("</IconStyle></Style>", file=f, append=TRUE)
    write (paste("<styleUrl>#",st,"</styleUrl>", sep=""), file=f, append=TRUE)
   }
    else
      write (paste("<styleUrl>", lgpIcon(lgp,s$No), "</styleUrl>", sep=""), file=f,
append=TRUE)
```

KMLhead

`KMLhead` writes the definition of the balloon style at the top of the KML file, within each icon style definition:

```
<styleUrl>#wmsPlacemark</styleUrl>
<Style id="normalPlacemark">
  <IconStyle><scale>0.4</scale>
    <Icon>
      <href>http://maps.google.com/mapfiles/kml/paddle/wht-circle-lv.png</href>
    </Icon>
  </IconStyle>
  <BalloonStyle>
  <bgColor>ccffff</bgColor>
    <text><![CDATA[
    <font  face="Arial" color="#666666" size="+3"><b>$[name]</b></font><br />
    <font  face="Verdana">$[description]</font><br />
    <small>(Links may be slow to open.)<br />
    $[geDirections]</small>
    ]]></text>
  </BalloonStyle>
</Style>
```

SECtail and GRPtail

`SECtail` closes off the secondary folder in the KML file:

```
SECtail<-function (pr,se,lgp, msym, chem.333) { # secondary catchment tail text
  write(paste("  </Folder><!-- end of",se,"-->"),
   file=fn(lgp,"KML",pr, msym, chem.333), append=TRUE)
}
```

and `GRPtail` closes off the group

```
GRPtail<-function(lgp, pr, msym, chem.333) { # write the folder closing text
  f <- fn(lgp, "KML", pr, msym, chem.333)
  write("</Folder>", file=f, append=TRUE)
}
```

This brings us to the end of the KML creation code. The script also produces a set of HTML tables to permit access to the plots and data without using Google Earth.

Creating HTML tables: HTMhead, SECbody, HTMtail

The HTML functions in `kml_WMS_mnpts_html.R` produce standard HTML text and tables with information about the dataset and links to the stations. As with the KML files, the HTML files are divided by primary drainage region or water management area, and surface water or groundwater. The HTML files do not include Maucha diagrams, to avoid reducing the download speed.

HTMhead

The `HTMhead` function writes the document head block, CSS style code, corporate labelling, standard links and explanatory text (Figure 3).

```
    write ("<!DOCTYPE html PUBLIC \"-//W3C//DTD XHTML 1.0 Transitional//EN\"
\"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd\">", file=f)
    write ("<html>" , file=f, append=TRUE)
    write ("<head>" , file=f, append=TRUE)
    write("<style type=\"text/css\">" , file=f, append=TRUE)
    write("<!-- " , file=f, append=TRUE)
    write("  @import url(\"style_wms_ge.css\"); " , file=f, append=TRUE)
    write(" -->" , file=f, append=TRUE)
    write("</style>" , file=f, append=TRUE)
    write (paste("<title>"), file=f, append=TRUE)
    if(WMA_on) write (paste(wmaname()[pr+1]," WMA ", format(Sys.time(), "%Y-%m-
%d %H:%M:%S"),
                                bfold, sep=" "), file=f, append=TRUE)
    else write (paste(toupper(pr)," region ", format(Sys.time(), "%Y-%m-%d
%H:%M:%S"),
                          bfold, sep=" "), file=f, append=TRUE)
    write (paste(", RQIS WMS water quality data Department of Water and
Sanitation, ",ver(),"</title>") , file=f, append=TRUE)
[…]
```

SECbody

The `SECbody` function described above (page 8) deals mainly with the data for a single point in the KML file but also includes the following code to write out a one-line site entry in the HTML table.

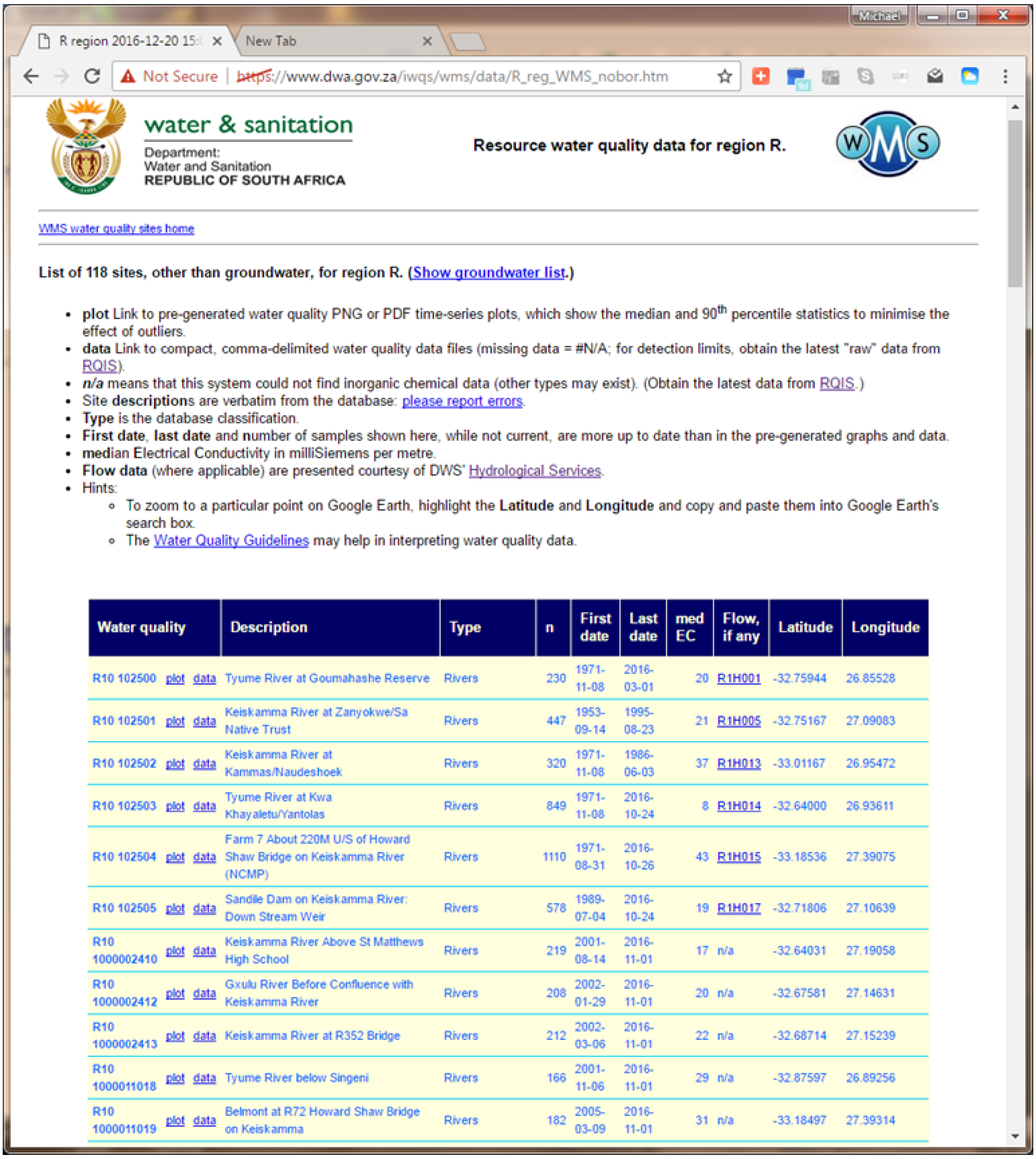
```
kml_WMS_mnpts_html.R - SciTE
File Edit Search View Tools Options Language Buffers Help

      if(!msym) { # no sense in creating a separate HTML for the Maucha process, as the HTML already exists
        # Write the point data to the HTML file
        f <- fn(lgp, "HTM", pr, msym, chem.333)
        write ("<tr><td>", file=f, append=TRUE) # start of table row
        write (paste("<b>", toupper(Ter), " ", PointID, "</b></td>", sep="") , file=f, append=TRUE)
        if(file.exists(plotFile)) write (paste("<td><a href=\"", plotLink, "\">plot</a></td>", sep=""), file=f, append=TRUE)
        else write ("<td><i>n/a</i></td>" , file=f, append=TRUE)
        if(file.exists(zipFile)) write(paste("<td><a href=\"", zipLink, "\">data</a></td>", sep=""), file=f, append=TRUE)
        else write ("<td><i>n/a</i></td>" , file=f, append=TRUE)
        write (paste("<td>", Name, " ", dsite, "</td><td>", Located_On_Type, "</td>", sep="") , file=f, append=TRUE)
        write (paste("<td align=\"right\">", No, "</td>", sep=""), file=f, append=TRUE)
        write (paste("<td>", FirstDate, "</td>", sep=""), file=f, append=TRUE)
        write (paste("<td>", LastDate, "</td>", sep=""), file=f, append=TRUE)
        if (is.na(medEC)) write ("<td>n/a</td>", file=f, append=TRUE)
        else write (paste("<td align=\"right\">", medEC, "</td>", sep=""), file=f, append=TRUE)
        if(his) write (paste("<td><a href=", fLink, ">", sta, "</a></td>", sep="") , file=f, append=TRUE)
        else write ("<td>n/a</td>" , file=f, append=TRUE)
        write ("<td>" , file=f, append=TRUE)
        write(sprintf ("%10.5f",Latitude) , file=f, append=TRUE)
        write ("</td><td>" , file=f, append=TRUE)
        write(sprintf ("%9.5f",Longitude), file=f, append=TRUE)
        write ("</td></tr>" , file=f, append=TRUE)
      }
    }
  }

line 701, column 62 (INS) (CR+LF) - 0 chars selected
```

HTMtail

`HTMtail` writes a date and time stamp at the bottom of the page, outputs the javascript for Google Analytics page tracking and closes off the HTML file with </body> and </html>.



R region 2016-12-20 15: New Tab

Not Secure | https://www.dwa.gov.za/iwqs/wms/data/R_reg_WMS_nobor.htm

water & sanitation

Department:
Water and Sanitation
REPUBLIC OF SOUTH AFRICA

Resource water quality data for region R.

WMS water quality sites home

List of 118 sites, other than groundwater, for region R. (Show groundwater list.)

- **plot** Link to pre-generated water quality PNG or PDF time-series plots, which show the median and 90th percentile statistics to minimise the effect of outliers.
- **data** Link to compact, comma-delimited water quality data files (missing data = #N/A; for detection limits, obtain the latest "raw" data from RQIS).
- *n/a* means that this system could not find inorganic chemical data (other types may exist). (Obtain the latest data from RQIS.)
- Site **descriptions** are verbatim from the database: please report errors.
- **Type** is the database classification.
- **First date**, **last date** and number of samples shown here, while not current, are more up to date than in the pre-generated graphs and data.
- **median** Electrical Conductivity in milliSiemens per metre.
- **Flow data** (where applicable) are presented courtesy of DWS' Hydrological Services.
- Hints:
  - To zoom to a particular point on Google Earth, highlight the **Latitude** and **Longitude** and copy and paste them into Google Earth's search box.
  - The Water Quality Guidelines may help in interpreting water quality data.

Water quality	Description	Type	n	First date	Last date	med EC	Flow, if any	Latitude	Longitude
R10 102500 plot data	Tyume River at Goumahashe Reserve	Rivers	230	1971-11-08	2016-03-01	20	R1H001	-32.75944	26.85528
R10 102501 plot data	Keiskamma River at Zanyokwe/Sa Native Trust	Rivers	447	1953-09-14	1995-08-23	21	R1H005	-32.75167	27.09083
R10 102502 plot data	Keiskamma River at Kammas/Naudeshoek	Rivers	320	1971-11-08	1986-06-03	37	R1H013	-33.01167	26.95472
R10 102503 plot data	Tyume River at Kwa Khayaletu/Yantolas	Rivers	849	1971-11-08	2016-10-24	8	R1H014	-32.64000	26.93611
R10 102504 plot data	Farm 7 About 220M U/S of Howard Shaw Bridge on Keiskamma River (NCMP)	Rivers	1110	1971-08-31	2016-10-26	43	R1H015	-33.18536	27.39075
R10 102505 plot data	Sandile Dam on Keiskamma River: Down Stream Weir	Rivers	578	1989-07-04	2016-10-24	19	R1H017	-32.71806	27.10639
R10 1000002410 plot data	Keiskamma River Above St Matthews High School	Rivers	219	2001-08-14	2016-11-01	17	n/a	-32.64031	27.19058
R10 1000002412 plot data	Gxulu River Before Confluence with Keiskamma River	Rivers	208	2002-01-29	2016-11-01	20	n/a	-32.67581	27.14631
R10 1000002413 plot data	Keiskamma River at R352 Bridge	Rivers	212	2002-03-06	2016-11-01	22	n/a	-32.68714	27.15239
R10 1000011018 plot data	Tyume River below Singeni	Rivers	166	2001-11-06	2016-11-01	29	n/a	-32.87597	26.89256
R10 1000011019 plot data	Belmont at R72 Howard Shaw Bridge on Keiskamma	Rivers	182	2005-03-09	2016-11-01	31	n/a	-33.18497	27.39314

Figure 3. Standard table layout. `HTMhead` generates the part up to and including the table header while `SECbody` generates each line within the table.

Special functions

Apart from the main functions in `kml_WMS_mnpts_html.R` dealt with above, the script contains several special service functions (Table 4).

Table 4. Special functions in kml_WMS_mnpts_html.R

Service function	purpose
ver	version numbering and description of changes
fn	compose the required KML and HTML file names based on region and types
ico	set the names and locations of monitoring site marker files
wmaname	set up a list of water management areas in numerical order
Located_On_Group	define groups of Located_On_Type definitions to reduce map legend clutter
lgpIcon	compose the group icon code as defined in ico
mtxt	set the automatic text to display when Maucha diagrams are present
mkey	write the KML code to display a Maucha key at the lower left of the screen

Results

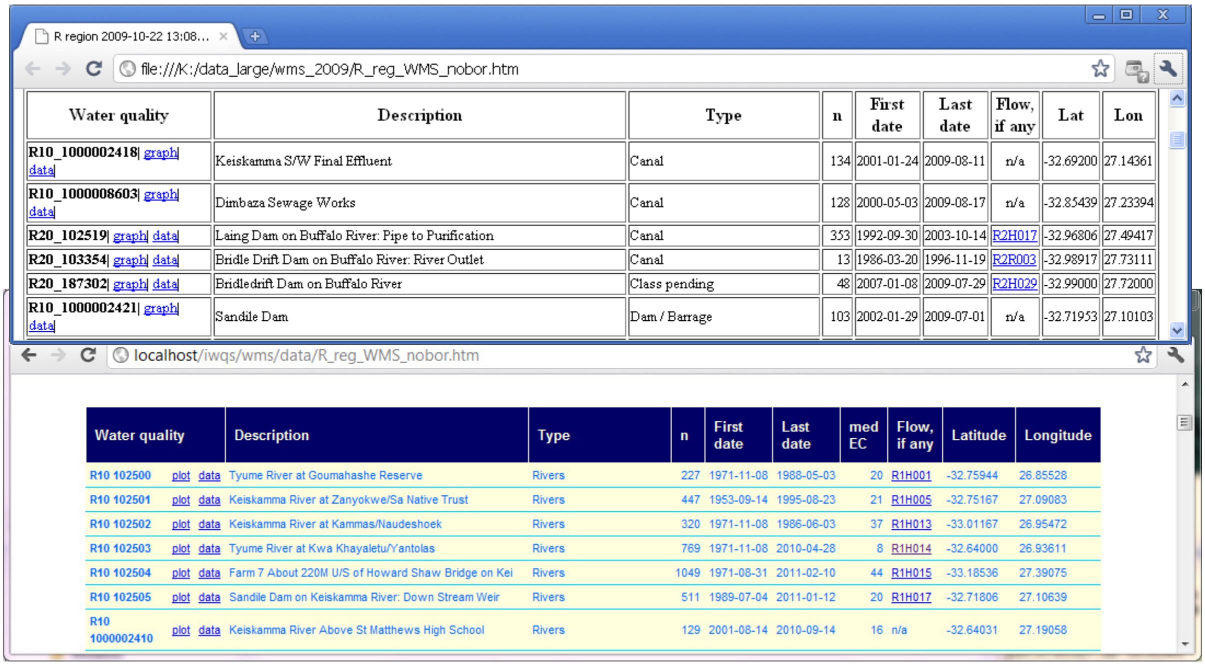
Conversion of the original `awk` script to R was successful, and the first version was ready in June 2011. Testing took place in July and August 2011 and the R-generated system was live at `http://www.dwa.gov.za/iwqs/wms/data/000key.asp` by 13 September 2011. Google Analytics registered 465 visits to the inventory from 15 September to 15 November 2011, and no complaints or queries were received.

The KML format is quite versatile and, apart from Google Earth, displays successfully in Google Maps (Figure 5) and ArcGIS.com (Figure 6). It may work in Microsoft Bing Maps.

Discussion

RODBC has opened up opportunities for direct query of the WMS database when assembling data for the KML files. Examples are the latitude and longitude, which previously came from an intermediate file, and the electrical conductivity entry, which would have involved too much intermediate work to be practical in the old `awk` version of the script.

The use of cascading style sheets (CSS) for tabular information has resulted in a neater and more compact display (Figure 4). Although these are nothing new and could have been part of the original system, R provides a more versatile environment for experimenting with and incorporating different methodologies.



R region 2009-10-22 13:08...

file:///K:/data_large/wms_2009/R_reg_WMS_nobor.htm

Water quality	Description	Type	n	First date	Last date	Flow, if any	Lat	Lon
R10_1000002418| graph| data|	Keiskamma S/W Final Effluent	Canal	134	2001-01-24	2009-08-11	n/a	-32.69200	27.14361
R10_1000008603| graph| data|	Dimbaza Sewage Works	Canal	128	2000-05-03	2009-08-17	n/a	-32.85439	27.23394
R20_102519| graph| data|	Laing Dam on Buffalo River: Pipe to Purification	Canal	353	1992-09-30	2003-10-14	R2H017	-32.96806	27.49417
R20_103354| graph| data|	Bridle Drift Dam on Buffalo River: River Outlet	Canal	13	1986-03-20	1996-11-19	R2R003	-32.98917	27.73111
R20_187302| graph| data|	Bridledrift Dam on Buffalo River	Class pending	48	2007-01-08	2009-07-29	R2H029	-32.99000	27.72000
R10_1000002421| graph| data|	Sandile Dam	Dam / Barrage	103	2002-01-29	2009-07-01	n/a	-32.71953	27.10103

localhost/iwqs/wms/data/R_reg_WMS_nobor.htm

Water quality			Description	Type	n	First date	Last date	med EC	Flow, if any	Latitude	Longitude
R10 102500	plot	data	Tyume River at Goumahashe Reserve	Rivers	227	1971-11-08	1988-05-03	20	R1H001	-32.75944	26.85528
R10 102501	plot	data	Keiskamma River at Zanyokwe/Sa Native Trust	Rivers	447	1953-09-14	1995-08-23	21	R1H005	-32.75167	27.09083
R10 102502	plot	data	Keiskamma River at Kammas/Naudeshoek	Rivers	320	1971-11-08	1986-06-03	37	R1H013	-33.01167	26.95472
R10 102503	plot	data	Tyume River at Kwa Khayaletu/Yantolas	Rivers	769	1971-11-08	2010-04-28	8	R1H014	-32.64000	26.93611
R10 102504	plot	data	Farm 7 About 220M U/S of Howard Shaw Bridge on Kei	Rivers	1049	1971-08-31	2011-02-10	44	R1H015	-33.18536	27.39075
R10 102505	plot	data	Sandile Dam on Keiskamma River: Down Stream Weir	Rivers	511	1989-07-04	2011-01-12	20	R1H017	-32.71806	27.10639
R10 1000002410	plot	data	Keiskamma River Above St Matthews High School	Rivers	129	2001-08-14	2010-09-14	16	n/a	-32.64031	27.19058

Figure 4. Updated table appearance using CSS (below).

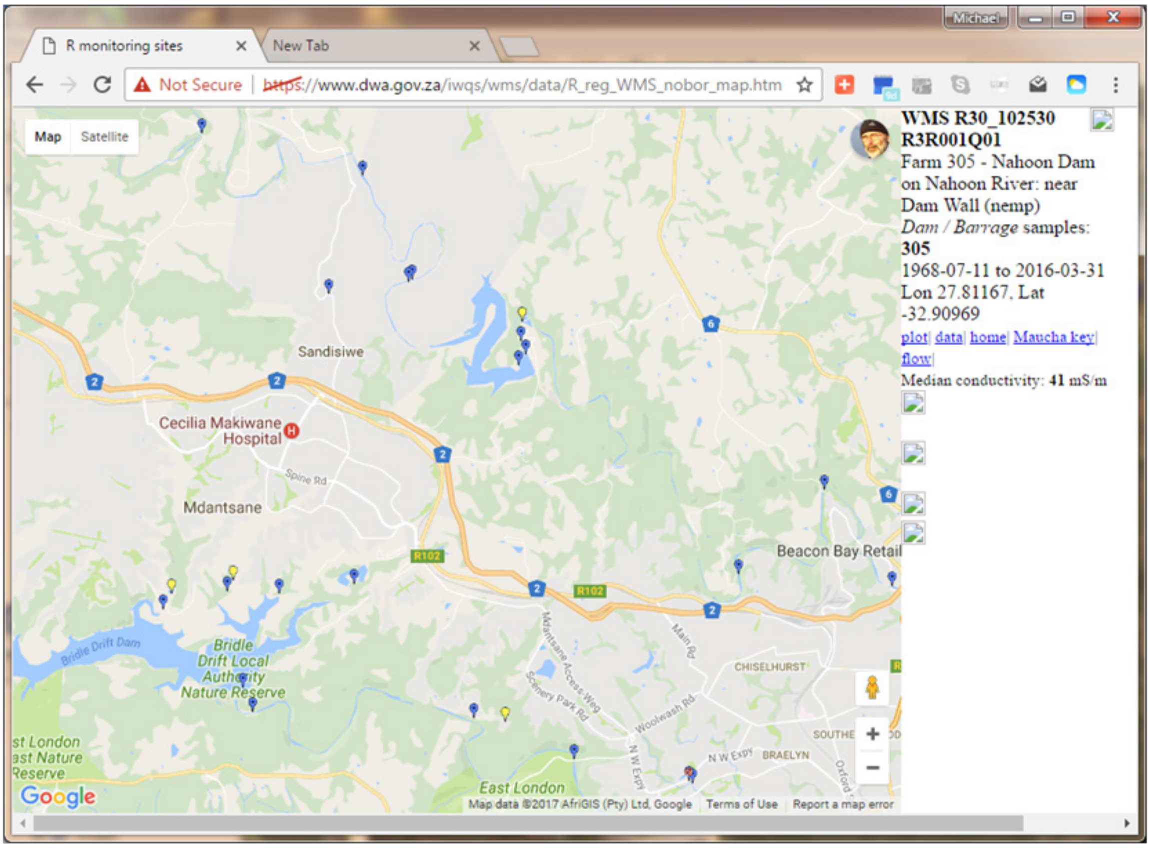


Figure 5. KML file from Figure 2 displayed in Google Maps. This layout is less informative than the previous version of Google Maps described in Silberbauer 2011.

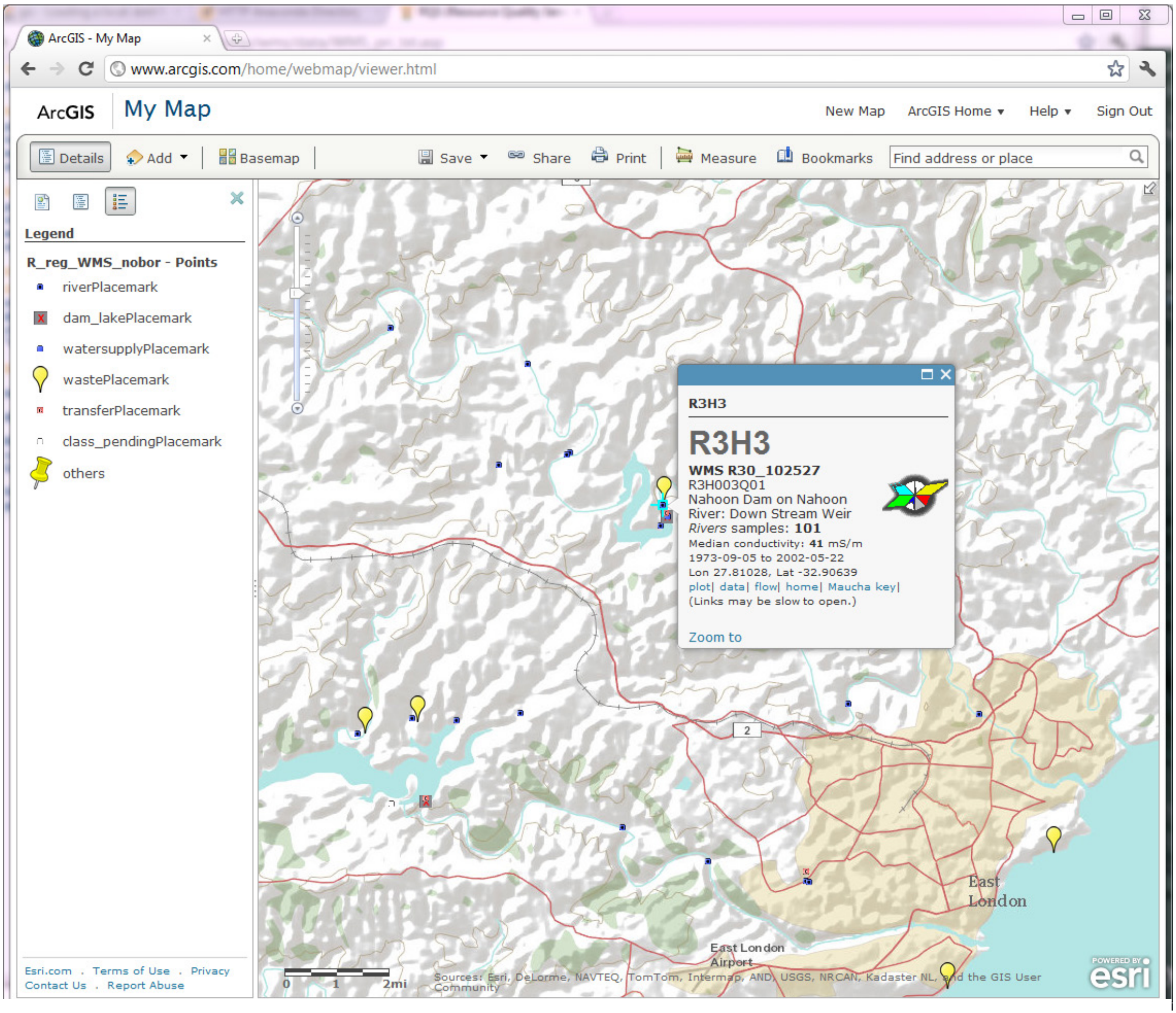


Figure 6. KML file from Figure 2 displayed in ArcGIS.com.

Conclusion

The conversion of kml_WMS_mnpts_html.awk to kml_WMS_mnpts_html.R resulted in a more manageable and extensible piece of code. Modifications are now more easily implemented.

This document should serve as a useful reference for anyone wanting to port the spatial inventory to another platform, or planning to implement a similar system for other data sets.

References

Crampton, J. W. 2008. Keyhole, Google Earth, and 3D Worlds: An Interview with Avi Bar-Zeev. *Cartographica* **43**:85-93.

Google 2011. KML 2.2 reference. http://code.google.com/apis/kml/documentation/kml_tags_beta1.html

McDonald, R. D. 1989. New numbers for gauging stations, monitoring points and systems as well as for data and information TR 141 (in Afrikaans). Technical report, Department of Water Affairs, Pretoria.

R 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Silberbauer, M. J. 2011. KML-based online spatial inventory of water chemistry data—R version (Conversion to R of the application that creates the online spatial inventory of water chemistry data). First draft. RQS, Pretoria. N/0000/00/RD1/2011

Silberbauer, M. J. 2017. Barcode time-series plots of inorganic water chemistry variables—RODBC version of the application that creates Barcode graphs of the main water chemistry variables. RQIS, Pretoria. N/0000/00/RD2/2011.

Silberbauer, M. J. and Geldenhuys, W. G. 2008. Using Keyhole Markup Language to create a spatial interface to South African water resource data through Google Earth. In Proceedings of the FOSS4G 2008 conference. http://www.foss4g.org/index.php/foss4g/2008/paper/view/74/28.

Silberbauer, M. and Geldenhuys, W. 2009. Google Earth - a spatial interface for SA water resource data. *PositionIT*, (April/May 2009): 42-47.