EcoHealth Alliance

Open Data Kit Quality Control Workflows

Ernest Guevarra

16 November 2021

Outline

- Brief overview of **ODK Central** version 1.3 data review feature
- Demonstration of **ODK Central**'s data review feature
    - manual process
    - semi-automated process using R
- Points to remember when using **ODK Central**'s data review feature

EcoHealth Alliance

ODK Central v1.3 and data review feature

- Allows project manager/s to review submitted data from **ODK Central** and mark each row of data as either **approved**, **has issues**, or **rejected**.
- When marking a submission with any of the review labels mentioned, a comment can be written to describe the reasons/basis for the label assigned.
- Allows project manager/s to edit submitted data from **ODK Central** to resolve issues identified with the data submitted.
- If data has been edited, project manager/s can then re-review and re-assign a review label accordingly (i.e., from **rejected** to **approved**)
- creates history of edits done within **ODK Central** which can be used by project manager/s in the review process

EcoHealth Alliance

Demonstration of ODK Central data review feature - manual process

- login to your **ODK Central** EHA account - only accounts of those with project manager roles will have the review features available
- we will look at an example project on **anthropometric data collection for children 6-59 months old**
- this will be good for quick and simple **eyeball** checks of data
- more sophisticated and elaborate data checks will be too tedious and inefficient through this approach

EcoHealth Alliance

Demonstration of ODK Cenetral data review feature - semi-automated process using R

- access data in R using the {ruODK} package
- perform data quality review and checks (including more sophisticated/elaborate ones) on data using R functions for this purpose with the aim of identifying rows of data with issues
- login to your **ODK Central** EHA account, go to your project and then manually mark and comment on the rows of data identified with issues

EcoHealth Alliance

Access data in R using the {ruODK} package

```r
## Install ruODK from GitHub
remotes::install_github("ropensci/ruODK")

## Load ruODK
library(ruODK)

## Setup connection with ODK Central
ru_setup(
  svc = "https://odk.eha.io/v1/projects/1/forms/anthropometry.svc",
  un = "YOUR_USERNAME_HERE",
  pw = "YOUR_PASSWORD_HERE",
  tz = "GMT",
  odkc_version = "1.3"
)

## Retrieve test data
anthro <- odata_submission_get()

## View data
anthro
```

EcoHealth Alliance

Access data in R using the {ruODK} package

```
## # A tibble: 25 × 24
##    id                start               end                 today               admin_team
##    <chr>             <dttm>              <dttm>              <dttm>              <chr>
##  1 uuid:cd765e1f-54… 2021-11-16 15:10:09 2021-11-16 15:25:32 2021-11-16 00:00:00 1
##  2 uuid:41eebbdd-27… 2021-11-16 15:14:24 2021-11-16 15:25:15 2021-11-16 00:00:00 1
##  3 uuid:9b5bb2dc-e2… 2021-11-16 14:59:33 2021-11-16 15:25:02 2021-11-16 00:00:00 1
##  4 uuid:d3e4412f-f1… 2021-11-16 14:58:31 2021-11-16 15:24:47 2021-11-16 00:00:00 1
##  5 uuid:b3fa396d-ca… 2021-11-16 14:57:32 2021-11-16 15:24:30 2021-11-16 00:00:00 1
##  6 uuid:3d64560b-91… 2021-11-16 14:56:39 2021-11-16 15:24:15 2021-11-16 00:00:00 1
##  7 uuid:75c5ed7e-e0… 2021-11-16 14:55:50 2021-11-16 15:23:58 2021-11-16 00:00:00 1
##  8 uuid:ee3870de-b1… 2021-11-16 14:54:46 2021-11-16 15:23:35 2021-11-16 00:00:00 1
##  9 uuid:b22ee844-f2… 2021-11-16 14:53:50 2021-11-16 15:23:20 2021-11-16 00:00:00 1
## 10 uuid:92d447aa-9a… 2021-11-16 15:16:38 2021-11-16 15:23:02 2021-11-16 00:00:00 1
## # … with 15 more rows, and 17 more variables: anthropometry_id <int>, anthropometry_age <in
## #   anthropometry_weight <dbl>, anthropometry_height <dbl>, anthropometry_muac <int>, anthr
## #   meta_instance_id <chr>, meta_instance_name <chr>, system_submission_date <chr>, system_
## #   system_submitter_name <chr>, system_attachments_present <int>, system_attachments_expec
## #   system_edits <int>, odata_context <chr>
```

EcoHealth Alliance

Perform data quality review/checks

```r
## Load anthropometric z-score calculator library and data checking library
library(zscorer)
library(nutricheckr)
library(dplyr)

## Calculate anthropometric z-scores
anthro_zscores <- anthro %>%
  mutate(age_days = anthropometry_age * (365.25 / 12)) %>%
  addWGSR(
    sex = "anthropometry_sex",
    firstPart = "anthropometry_weight",
    secondPart = "age_days",
    index = "wfa"
  ) %>%
  addWGSR(
    sex = "anthropometry_sex",
    firstPart = "anthropometry_height",
    secondPart = "age_days",
    index = "hfa"
  ) %>%
  addWGSR(
    sex = "anthropometry_sex",
    firstPart = "anthropometry_weight",
    secondPart = "anthropometry_height",
    index = "wfh"
  )

## View output
anthro_zscores
```

Perform data quality review/checks

```
## ================================================================================
## ================================================================================
## ================================================================================

## # A tibble: 25 × 14
##    team  date_survey         location    id   age sex   weight height  muac oedema age_days
##    <chr> <dttm>              <chr>    <int> <int> <chr>  <dbl>  <dbl> <int> <chr>     <dbl>
##  1 1     2021-11-19 00:00:00 1           21    48 1       13.8   98     154 2          1461
##  2 1     2021-11-18 00:00:00 1           23    36 2       13.9   95.5   148 2          1096.
##  3 1     2021-11-16 00:00:00 1            9    13 1        8.3   73     136 2           396.
##  4 1     2021-11-16 00:00:00 1            8    13 1        8.1  999     124 2           396.
##  5 1     2021-11-16 00:00:00 1            7    24 2        8     74.2   144 2           730.
##  6 1     2021-11-16 00:00:00 1            6    18 2        7.7   70.6   130 2           548.
##  7 1     2021-11-16 00:00:00 1            5    15 1        7.4   70     124 2           457.
##  8 1     2021-11-16 00:00:00 1            4    15 1        7.2   75.4   130 2           457.
##  9 1     2021-11-16 00:00:00 1            3    15 1        7.1   67.5   124 2           457.
## 10 1     2021-11-16 00:00:00 1           25    36 1       14.2   95.6   152 2          1096.
## # … with 15 more rows
```

EcoHealth Alliance

Perform data quality review/checks

```
## Flag z-scores
anthro_flags <- anthro_zscores %>%
  flag_who(hlaz = "hfaz", waz = "wfaz", whlz = "wfhz")

anthro_flags
```

```
## # A tibble: 25 × 16
##    team  date_survey         location    id   age sex   weight height  muac oedema age_days
##    <chr> <dttm>              <chr>    <int> <int> <chr>  <dbl>  <dbl> <int> <chr>     <dbl>
##  1 1     2021-11-19 00:00:00 1           21    48 1       13.8   98     154 2          1461
##  2 1     2021-11-18 00:00:00 1           23    36 2       13.9   95.5   148 2          1096.
##  3 1     2021-11-16 00:00:00 1            9    13 1        8.3   73     136 2           396.
##  4 1     2021-11-16 00:00:00 1            8    13 1        8.1  999     124 2           396.
##  5 1     2021-11-16 00:00:00 1            7    24 2        8     74.2   144 2           730.
##  6 1     2021-11-16 00:00:00 1            6    18 2        7.7   70.6   130 2           548.
##  7 1     2021-11-16 00:00:00 1            5    15 1        7.4   70     124 2           457.
##  8 1     2021-11-16 00:00:00 1            4    15 1        7.2   75.4   130 2           457.
##  9 1     2021-11-16 00:00:00 1            3    15 1        7.1   67.5   124 2           457.
## 10 1     2021-11-16 00:00:00 1           25    36 1       14.2   95.6   152 2          1096.
## # … with 15 more rows, and 1 more variable: flag_description <chr>
```

EcoHealth Alliance

Perform data quality review/checks

```
## Subset anthropometric data to flagged rows
anthro_for_checking <- anthro_flags %>%
  filter(flag != 0)

anthro_for_checking
```

```
## # A tibble: 1 × 16
##   team  date_survey         location    id   age sex   weight height  muac oedema age_days
##   <chr> <dttm>              <chr>    <int> <int> <chr>  <dbl>  <dbl> <int> <chr>     <dbl>
## 1 1     2021-11-16 00:00:00 1            1    20 2        6.1   82.5   127 2          609.
## # … with 1 more variable: flag_description <chr>
```

EcoHealth Alliance

Questions?

EcoHealth Alliance

Thank you!

Slides can be viewed at https://ecohealthalliance.github.io/odk_quality_control or PDF version downloaded at https://ecohealthalliance.github.io/odk_quality_control/odk_quality_control.pdf

R scripts for slides available at https://github.com/ecohealthalliance/odk_quality_control

R script demonstrating how to access data from **ODK Central** using {ruODK} package at https://github.com/ecohealthalliance/odk_quality_control/blob/main/odk_quality_control.R