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

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

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

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

team	date_survey	location	id	age	sex	weight	height	muac	oedema	age_days	wfaz	hfaz	wfhz
1	2021-11-19	1	21	48	1	13.8	98.0	154	2	1461.0000	-1.34	-1.27	-0.88
1	2021-11-18	1	23	36	2	13.9	95.5	148	2	1095.7500	0.03	0.12	-0.09
1	2021-11-16	1	9	13	1	8.3	73.0	136	2	395.6875	-1.58	-1.62	-1.11
1	2021-11-16	1	8	13	1	8.1	999.0	124	2	395.6875	-1.80	380.02	NA
1	2021-11-16	1	7	24	2	8.0	74.2	144	2	730.5000	-3.07	-3.78	-1.33
1	2021-11-16	1	6	18	2	7.7	70.6	130	2	547.8750	-2.40	-3.48	-0.82
1	2021-11-16	1	5	15	1	7.4	70.0	124	2	456.5625	-3.00	-3.62	-1.61
1	2021-11-16	1	4	15	1	7.2	75.4	130	2	456.5625	-3.23	-1.49	-3.57
1	2021-11-16	1	3	15	1	7.1	67.5	124	2	456.5625	-3.34	-4.61	-1.25
1	2021-11-16	1	25	36	1	14.2	95.6	152	2	1095.7500	-0.08	-0.13	0.00
1	2021-11-16	1	24	36	1	14.0	96.5	142	2	1095.7500	-0.20	0.11	-0.37
1	2021-11-16	1	20	48	1	13.4	99.7	142	2	1461.0000	-1.58	-0.86	-1.63
1	2021-11-16	1	2	13	2	6.4	70.4	116	2	395.6875	-3.04	-1.84	-2.93
1	2021-11-16	1	19	36	1	13.3	97.2	142	2	1095.7500	-0.63	0.30	-1.17
1	2021-11-16	1	18	48	2	13.3	96.3	160	2	1461.0000	-1.41	-1.49	-0.76
1	2021-11-16	1	17	48	1	12.6	96.0	142	2	1461.0000	-2.07	-1.75	-1.60

Perform data quality review/checks

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

anthro_flags
```

team	date_survey	location	id	age	sex	weight	height	muac	oedema	age_days	wfaz	hfaz	wfhz	flag	flag_description
1	2021-11-19	1	21	48	1	13.8	98.0	154	2	1461.0000	-1.34	-1.27	-0.88	0	No flagged measurements
1	2021-11-18	1	23	36	2	13.9	95.5	148	2	1095.7500	0.03	0.12	-0.09	0	No flagged measurements
1	2021-11-16	1	9	13	1	8.3	73.0	136	2	395.6875	-1.58	-1.62	-1.11	0	No flagged measurements
1	2021-11-16	1	8	13	1	8.1	999.0	124	2	395.6875	-1.80	380.02	NA	NA	
1	2021-11-16	1	7	24	2	8.0	74.2	144	2	730.5000	-3.07	-3.78	-1.33	0	No flagged measurements
1	2021-11-16	1	6	18	2	7.7	70.6	130	2	547.8750	-2.40	-3.48	-0.82	0	No flagged measurements
1	2021-11-16	1	5	15	1	7.4	70.0	124	2	456.5625	-3.00	-3.62	-1.61	0	No flagged measurements
1	2021-11-16	1	4	15	1	7.2	75.4	130	2	456.5625	-3.23	-1.49	-3.57	0	No flagged measurements
1	2021-11-16	1	3	15	1	7.1	67.5	124	2	456.5625	-3.34	-4.61	-1.25	0	No flagged measurements

Perform data quality review/checks

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

anthro_for_checking
```

team	date_survey	location	id	age	sex	weight	height	muac	oedema	age_days	wfaz	hfaz	wfhz	flag	flag_description
1	2021-11-16	1	1	20	2	6.1	82.5	127	2	608.75	-4.54	-0.07	-6.03	2	Check weight and height measurements

Questions?

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

EcoHealth
Alliance