HARVARD

Office of the Vice Provost for Advances in Learning

Harvard VPAL
Private Datasets

September 28, 2015
Rev -
Datavarse Part #

Change Log

Rev	Date	Description
-	2015-09-28	Initial release

TABLE OF CONTENTS

OVERVIEW

This documentation describes all the fields and variable names that will be included in all custom research data requests that have been approved by the VPAL Research team.

The VPAL Research team is able to provide two private datasets, hereinafter referred to as the *"Person-Course-Survey"* dataset, and *"Log Data".* A standard limit of five (5) HarvardX courses will be imposed on each external data request, although there can be some cases where a larger number of course can be provided by special arrangement.

Upon review of the external research data request, and approval by the VPAL Research team, the research data for the respective courses requested will be added to the queue to be processed, encrypted and finally delivered.

Notes:

- The "Person-Course-Survey" dataset will not necessarily include the latest data as of the date and time when the data was processed, meaning the data received could reflect summary statistics a month back

Variable Name and Definitions

The following variables are included in this protected FERPA constrained dataset. This dataset includes all registrants for the courses requested, including variables that come from the edX system, data computed by the research team and survey questions answered by users for a given course.

Notes:

- "administrative" indicates that the variable comes from the edX system or has been computed by the research team
- "user-provided" indicates that the variable comes from questions asked by edX of the student at the time of registration with edX, indicates that the variable comes from survey questions asked by HarvardX or MITx;
- blank values in user-provided columns indicate that although the user created their edX account after the question was asked in the student registration process, the user declined to provide the information, the user did not respond to a survey question;

Private Datasets

Person-Course-Survey Dataset

The Person-Course-Survey dataset is at the level of one row per-person, per-course. So, for example, if one individual enrolled in three MITx or HarvardX courses during the period covered by the dataset (for this release, Fall 2012, Spring 2013, Summer 2013, Fall 2013, Spring 2014 and Summer 2014), that person would have three rows associated with their user ID.

There are two (2) sources of data used to generate the Person-Course-Survey dataset. The first primary source of data is the raw edX data, including Student Info and Progress Data, which contains database exports about all users that have registered for a HarvardX course on the edX platform, and the Tracking Log data, which includes all activities/events associated with users interacting with the HarvardX course on the edX platform. The secondary source of data used to generate the Person-Course-Survey dataset is the Qualtrics Pre-course Survey data. Data from both sources have been combined for all users in all HarvardX courses, and then summarized in our Person-Course-Survey dataset.

The Person Course Survey dataset will be provided to the external researcher as a single encrypted [based on their public GPG key], comma-separated

file [csv format], containing all the registrants for the specific course(s) that was requested (typically up to 5 maximum).

The following is a list of variables for the "Person Course Survey" dataset, which will comply with FERPA constrained datasets:

List of Variables

course_id
administrative; course ID in the standard format as org/number/semester

registered
administrative; Boolean flag, indicating that the user registered in the course (obsolete: should always be 1 or true)

viewed
administrative; Boolean flag, indicating that the user visited the course at least once

explored
administrative; Boolean flag, indicating that the user viewed at least half of the chapters of the course

certified
administrative; Boolean flag, indicating that the user earned a certificate in the course

cc_by_ip
administrative; Two-letter country code corresponding to the modal IP address

countryLabel
administrative; Full name of country corresponding to modal IP address

city
administrative; Full name of city corresponding to modal IP address

region
administrative; Two-Letter code of region or U.S. State corresponding to modal IP address

subdivision
administrative; Full name of region/subdivision or U.S. State corresponding to modal IP address

postalCode

administrative; Postal code corresponding to modal IP address

un_major_region
administrative; UN defined sub-continent major geographic regions corresponding to modal IP address

un_economic_group
administrative; UN defined major geographical economic groups (developing and developed nations) corresponding to modal IP address

un_developing_nation
administrative; UN defined developing nation groups files corresponding to modal IP address

un_special_region
administrative; Special Regions Set 1 files: Latin America and the Caribbean.csv,Sub-Saharan-Africa.csv corresponding to modal IP address

LoE
user-provided; Level of education

YoB
user-provided; Year of birth

gender
user-provided; gender, one of 'm', 'f', or 'o' (or NULL)

grade
administrative; Final grade earned in the course, out of 100 points

start_time
administrative; Date of enrollment in the course

last_event
administrative; Date of last event of user, from tracking logs

nevents
administrative; Number of tracking log events

ndays_act
administrative; Number of days with activity, from tracking logs

nplay_video
administrative; Number of video play events from tracking logs

nchapters
administrative; Number of chapters visited by the user

nforum_posts
administrative; Number of forum posts and comments made by user

nforum_votes
administrative; Number of votes received by user's forum posts

nforum_endorsed
administrative; Number of endorsed forum posts by user

nforum_threads
administrative; Number of forum post threads by user

nforum_comments
administrative; Number of forum post comments by user

nforum_pinned
administrative; Number of forum posts by user which were pinned (good indicator of staff)

roles
administrative; Roles played by the user in the course, e.g. instructor, course_staff

nprogcheck
administrative; Number of progress check events from tracking logs

nproblem_check
administrative; Number of problem check events from tracking logs

nforum_events
administrative; Number of all kinds of forum events, from tracking logs

mode
administrative; Mode of registrant, e.g. honor, audit, verified

is_active
administrative; 1 if enrollment active on date of this data, else 0

cert_created_date
administrative; date when certificate was generated (if applicable)

cert_modified_date

administrative; date when certificate was modified (if applicable)

profile_country
administrative; two-letter country code as specified by user in their profile (only available after ~mid-2014)

ntranscript
administrative; Number of video transcript events from tracking logs

nshow_answer
administrative; Number of show answer events from tracking logs

nvideo
administrative; Number of video events (of all kinds) from tracking logs

nseq_goto
administrative; Number of 'sequence goto' navigational events from tracking logs

nseek_video
administrative; Number of 'video seek' events from tracking logs

npause_video
administrative; Number of 'video pause' events from tracking logs

avg_dt
administrative; Average time difference in seconds between consecutive events from tracking logs

sdv_dt
administrative; Standard deviation of difference in seconds between consecutive events from tracking logs

max_dt
administrative; Maximum difference in seconds between consecutive events from tracking logs

n_dt
administrative; Number of consecutive events used in time difference computations, from tracking logs

sum_dt
administrative; Total elapsed time (in seconds) spent by user on this course, based on time difference of consecutive events, with 5 min max cutoff, from tracking logs

prs_Status

user-provided; Type of response collected

prs_StartDate
user-provided; When participant first clicked on Survey link

prs_EndDate
user-provided; When participant submitted response

prs_Finished
user-provided; Finished all survey questions or closed survey without finishing survey

prs_oc_reg
user-provided; How many online courses have you registered for in the / past?
Possible values:
0-12 (12 indicates 12 or more)

prs_oc_comp
user-provided; How many online courses have you completed in the past?
Possible values:
0-12 (12 indicates 12 or more)

prs_fam
user-provided; How familiar are you with historical study? (No familiarity is / required or expected)
Possible values:
0 = Not at all familiar
1 = Slightly familiar
2 = Somewhat familiar
3 = Very familiar
4 = Extremely familiar

prs_complete_psets
user-provided; How much of each of the following elements of this course do you / intend to complete?-Problem Sets
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_vid

user-provided; How much of each of the following elements of this course do you / intend to complete?-Lecture Sequence Videos, How much of each of the following elements of this course do you / intend to complete?-Module Videos
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_vidprob
user-provided; problem solving videos
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_vidques
user-provided; video review questions
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_read
user-provided; How much of each of the following elements of this course do you / intend to complete?-Course Readings
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_ques
user-provided; How much of each of the following elements of this course do you / intend to complete?-Module Questions, Lesson Sequence Questions, Course Reading Questions
Possible values:
0 = None

1 = Few
2 = Some
3 = Most
4 = All

prs_complete_exam

user-provided; How much of each of the following elements of this course do you / intend to complete?-Exams, final project, quizzes, weekly projects
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_lab

user-provided; How much of each of the following elements of this course do you / intend to complete?-Lab Experiments
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_lec

user-provided; How much of each of the following elements of this course do you / intend to complete?-Lecture Sequences, recitation
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_hw

user-provided; How much of each of the following elements of this course do you / intend to complete?-Homework, assessments
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_peer

user-provided; How much of each of the following elements of this course do you / intend to complete?-Peer Grading
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_complete_intact

user-provided; How much of each of the following elements of this course do you intend to complete?-Interactive Activities, analytics competition
Possible values:
0 = None
1 = Few
2 = Some
3 = Most
4 = All

prs_forums

user-provided; How do you intend to participate in the forums?
Possible values:
0 = I will not visit the discussion forums
1 = I will view discussion threads, but will not contribute
2 = I will contribute to discussion threads occasionally
3 = I will contribute to discussion threads frequently

prs_zip_us

user-provided; What is your zip code? US

prs_zip_ca

user-provided; What is your postal code? Canada

prs_zip_uk

user-provided; What is your postal code? UK

prs_zip_au

user-provided; What is your postal code? Australia

prs_zip_nz

user-provided; What is your postal code?

prs_zip_nl

user-provided; What is your postal code?

prs_zip_00

user-provided; What is your postal code?

prs_fluent_reading

user-provided; How fluent are you in English, the language of this course?- Reading
Possible values:
0 = Weak
1 = Basic
2 = Intermediate
3 = Proficient
4 = Fluent

prs_fluent_writing

user-provided; How fluent are you in English, the language of this course?- Writing
Possible values:
0 = Weak
1 = Basic
2 = Intermediate
3 = Proficient
4 = Fluent

prs_fluent_speaking

user-provided; How fluent are you in English, the language of this course?- Speaking
Possible values:
0 = Weak
1 = Basic
2 = Intermediate
3 = Proficient
4 = Fluent

prs_fluent_listening

user-provided; How fluent are you in English, the language of this course?- Listening
Possible values:
0 = Weak
1 = Basic
2 = Intermediate
3 = Proficient
4 = Fluent

prs_father_ed

user-provided; What is the highest level of education that your mother and father completed?-Father
-1 = Other / Don't know
0 = None
1 = Elementary / Primary School
2 = Junior secondary / junior high / middle school
3 = Secondary / High School
4 = Associate's Degree
5 = Bachelor's Degree
6 = Masters or Professional Degree
7 = Doctorate

prs_father_ed.1

user-provided; Father/Stepfather or Mother/Stepmother
-1 = Other / Don't know
0 = None
1 = Elementary / Primary School
2 = Junior secondary / junior high / middle school
3 = Secondary / High School
4 = Associate's Degree
5 = Bachelor's Degree
6 = Masters or Professional Degree
7 = Doctorate

prs_mother_ed

user-provided; What is the highest level of education that your mother and father completed?-Mother
-1 = Other / Don't know
0 = None
1 = Elementary / Primary School
2 = Junior secondary / junior high / middle school
3 = Secondary / High School
4 = Associate's Degree
5 = Bachelor's Degree
6 = Masters or Professional Degree
7 = Doctorate

prs_mother_ed.1

user-provided; Father/Stepfather or Mother/Stepmother
-1 = Other / Don't know
0 = None
1 = Elementary / Primary School
2 = Junior secondary / junior high / middle school
3 = Secondary / High School
4 = Associate's Degree

5 = Bachelor's Degree
6 = Masters or Professional Degree
7 = Doctorate

prs_working

user-provided; Are you currently employed? # Are you currently working in a job or business?
Possible values:
-1 = Unsure
0 = No,
1 = Yes

prs_school

user-provided; Are you currently enrolled in school?
Possible values:
-1 = Unsure
0 = No,
1 = Yes

prs_work_ftpt

user-provided; Are you working full-time or part-time in your job or business?
Possible values:
0 = Part-time
1 = Full-time

prs_school_lev

user-provided; Please indicate the kind of school or school program you are enrolled in:
Possible values:
-1 = Other
0 = Not enrolled
1 = Primary / elementary school
2 = Junior secondary / junior high / middle school
3 = Secondary / high school
4 = 2 year college
5 = 4 year college
6 = Graduate or Professional School

prs_school_ft

user-provided; Are you enrolled as a full-time or part-time student at this school?
Possible values:
0 = Part-time
1 = Full-time

prs_harvard

user-provided; Are you affiliated with Harvard?
Possible values:
-1 = Other
0 = Not affiliated
1 = Student
2 = Alumnus / Alumna
3 = Staff
4 = Faculty

prs_MIT

user-provided; Are you affiliated with MIT?
Possible values:
-1 = Other
0 = Not affiliated
1 = Student
2 = Alumnus / Alumna
3 = Staff
4 = Faculty

prs_teach

user-provided; Are you currently, or have you ever identified yourself as, an / instructor/teacher?
Possible values:
-1 = Unsure
0 = No,
1 = Yes

prs_teach_crs

user-provided; Are you, or have you ever, taught material related to this course?
Possible values:
-1 = Unsure
0 = No,
1 = Yes

prs_teach_now

user-provided; Are you currently employed as an instructor/teacher?
Possible values:
-1 = Unsure
0 = No,
1 = Yes

prs_teach_elem

user-provided; In what settings did your instruction take place? (Please click all / that apply)-Elementary / Primary School
Possible values:
0 = No
1 = Elementary / Primary School

prs_teach_hs
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Secondary School / MIddle School/High School
Possible values:
0 = No
1 = Secondary School / High School

prs_teach_college
user-provided; In what settings did your instruction take place? (Please click all / that apply)-College or University
Possible values:
0 = No
1 = College or University

prs_teach_out
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Outside the scope of traditional schools.
Possible values:
0 = No
1 = Outside the scope of traditional schools

prs_teach_ta
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Support staff, e.g., Teaching Assistant.
Possible values:
0 = No
1 = Support staff, e.g.: Teaching Assistant

prs_teach_other
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Other
Possible values:
0 = No
1 = Yes

prs_teach_home
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Homeschool
Possible values:

0 = No
1 = Homeschool

prs_teach_tutor
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Tutor
Possible values:
0 = No
1 = Tutor

prs_teach_comm
user-provided; In what settings did your instruction take place? (Please click all / that apply)-Community Center
Possible values:
0 = No
1 = Community Center

prs_reason_lc
user-provided; People register for HarvardX courses for different reasons. Which of the following best describes you?
Possible values:
0 = Have not decided whether I will complete any course activities.
1 = Here to browse the materials, but not planning on completing any course activities (watching videos, reading text, answering problems, etc.).
2 = Planning on completing some course activities, but not planning on earning a certificate.
3 = Planning on completing enough course activities to earn a certificate.

prs_country_l
user-provided; List of Countries

prs_motiv_engage
user-provided; How important were the following reasons in choosing to register / for this course? -Engaging in lifelong learning
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_curious
user-provided; How important were the following reasons in choosing to register / for this course? -Curiosity about online learning
Possible values:

0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_career

user-provided; How important were the following reasons in choosing to register / for this course? -Advancing my career
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_educ

user-provided; How important were the following reasons in choosing to register / for this course? -Advancing my formal education
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_bestprof

user-provided; How important were the following reasons in choosing to register / for this course? -To learn from the best professors and universities
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_comm

user-provided; How important were the following reasons in choosing to register / for this course? -To better serve my community
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_opport

user-provided; How important were the following reasons in choosing to register / for this course? -To access learning opportunities not otherwise available to me
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_cert

user-provided; How important were the following reasons in choosing to register / for this course? -To earn a certificate
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_participate

user-provided; How important were the following reasons in choosing to register / for this course? -To participate in an online community
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

prs_motiv_learn

user-provided; How important were the following reasons in choosing to register / for this course? -To learn about course content
Possible values:
0 = Not Important
1 = Slightly Important
2 = Somewhat Important
3 = Very Important
4 = Extremely Important

roles_isBetaTester

administrative; 1, if user is a beta tester

roles_isInstructor
administrative; 1, if user is an instructor

roles_isStaff
administrative; 1, if user is staff

forumRoles_isAdmin
administrative; 1, if user is an forum administrator

forumRoles_isCommunityTA
administrative; 1, if user is a community TA

forumRoles_isModerator
administrative; 1, if user is a forum moderator

forumRoles_isStudent
administrative; 1, if user is a student

Log Data

The Log Data contains data at the level of one activity/event per user. An event performed by a user on the edX platform when taking a HarvardX course is recorded as Log Data and stored as JSON documents.

[excerpt taken from edX Research Guide as of 2015-08-28]
Events are emitted by the server, the browser, or the mobile device to capture information about interactions with the courseware and the Instructor Dashboard in the LMS, and are stored in JSON documents. In the data package, event data is delivered in a log file.

The Log Data will be provided to the external researcher as a single encrypted [based on their public GPG key], compressed gzipped file (e.g.: "Request_ID#_v#_YYYY_MM_DD.log.gz"), containing all events for all users for the specific course(s) requested (typically up to 5 maximum). A directory for each course requested will be created, and each directory will contain log files separated by day (e.g.: "YYYY_MM_DD.log.gz"), where each daily log file will contain all events for all users during that day based on the event timestamp.

The following is a list of variables from the Log Data dataset, which will be modified from the original source dataset in order to comply with FERPA constrained datasets. Further information on the log event fields are available in the edX Research Guide documentation. See the latest edX Research Guide documentation for the latest details.

List of Modified Variables

user_id [modified => hashed id]

[excerpt taken from edX Research Guide as of 2015-08-28]
The user_id of the user who caused the event to be emitted. This string is empty for anonymous events, such as when the user is not logged in.

List of Variables [to be included in dataset to External Researchers]

See the latest edX Research Guide documentation for the latest details.