About the "Learn Moodle" anonymized database

Abstract: This document provides a short overview of how to extract and view the open, anonymized data from the "Teaching with Moodle" course from learn.moodle.net.

1   Introduction

Moodle Pty Ltd has made an open data set available for research purposes via research.moodle.net. This document provides an overview of how to obtain and analyze the open, anonymized data from the "Teaching with Moodle" course from learn.moodle.net for the purpose of conducting original research. In keeping with the open source nature of Moodle and the principles of open access and open research, primarily open source tools will be discussed here.

1.1 About the source

The source of this data is the August 2016 session of the "Teaching with Moodle" MOOC, hosted at https://learn.moodle.net/. Sessions of this course are delivered twice per year by employees of Moodle PTY LTD. The description of the course reads, "This free course is designed for anybody who wants to use the Moodle learning platform for teaching, whether it be in a school, a university, a company or just personal interest! The four-week course is also a great opportunity to connect with the vibrant Moodle community dedicated to sharing resources, ideas and anything that could help inspire better teaching practices everywhere." This session was delivered in a fully online format from August 7 to September 4, 2016. There were 6119 students enrolled in this course, and 2 facilitators. Of those 6119 students, participation differed as follows:

Level of participation	Number of students	Percentage of enrolled students
Enrollment	6119	100%
Completed "May we use your data for research?" survey (required to access remainder of course)	2769	45%
Granted permission to use data in research	2167	35%
Earned "Learn Moodle participant Aug 2016" badge (requires a post in "Forum – Introduce yourself!"	1566	26%

Earned "Learn Moodle completer Aug 2016" (requires completion of all activities in course)	735	12%

1.2 About the data

Only data from the August 2016 students agreeing to allow their data to be used in research is included in this data set. A copy was made of the production site https://learn.moodle.net and within that copy, all other courses and users were removed from the site. The remaining data was then anonymised using the master branch of the Local Anonymise plugin for Moodle (https://github.com/dmonllao/moodle-local_anonymise) as of January 10, 2017. Selected tables from the Moodle site database were extracted using psql to csv (comma-separated values) files. Because each user within a Moodle site is assigned a unique identification number which is visible in URLs on the original production site, this value was excluded from extracted tables. Instead, an anonymised username has been provided in each data file to allow reconstruction of relationships between users, activity completion, log entries, badges, and grades. Certain other columns of data were excluded either for privacy reasons or because they did not contain valuable data in this data set (e.g. they represented a feature not used in this course).

Although the course included some non-core plugins and activities, e.g. BigBlueButton, only data from Moodle core plugins are included in this data set. This is due to current limitations of the Anonymise plugin to de-identify sensitive information only in core Moodle tables.

2 Getting and Unpacking the Data

2.1 Get the data set

Get the data set from the Moodle Research Repository at http://research.moodle.net/158/ . You will need to create a login if you have not yet done so, but there is no cost.

2.2 Unpack Data

Six CSV files are included in this data set. For convenience and to limit download bandwidth, they are compressed into one file using the gzip utility. This compressed file can be "unpacked" using an extraction utility that supports gzip (http://www.gzip.org/). Utilities that support gzip are readily available for Microsoft Windows, Apple Macintosh OSX, and Linux systems.

2.3 Examine Files

In the "export" folder, you will find the following files:

File name	Description
mdl_badge_issued.csv	This file contains the records of all badges issued to users during the Aug 16 session of "Teaching with Moodle."
mdl_course_modules.csv	This file contains records describing each activity in the "Teaching with Moodle" course.
mdl_course_modules_completion.csv	This file contains records of each user's completion of each activity in the course.
mdl_grade_grades_history.csv	This table keeps a historical record of individual grades for each user and each item, exactly as imported or submitted by modules.
mdl_logstore_standard_log.csv	This table contains entries for each "event" tracked by the Moodle logging system, and is the source for all the Moodle "log" reports
mdl_user.csv	This is the table containing user records.

A brief description of each file appears in the Appendix. Some of these files are quite large, especially mdl_grade_grades_history.csv and mdl_logstore_standard_log.csv, and may not open in a word processor or spreadsheet software.

3   Data File Formats

Each file incuded in the Export package is a "comma-separated values" text file. Information about Moodle tables is available from within a Moodle site via the xmldb editor, located in the Site Administration→ Development menu. Each file is described briefly below:

mdl_user.csv (2170 records plus header)
This is the table containing user records. The following fields are included:

Field	Description
id	The user id from the original Moodle installation. In the anonymized files, this is always set to 0
confirmed	The account authentication was confirmed. All users in this set confirmed = 1.
deleted	When users are deleted from the Moodle site, their records remain with deleted = 1. All users in this set have deleted = 0 (deleted users have not been included in the export).
suspended	There are no suspended users in this data set.
mnethostid	= 1 Unique remote host ID

username	All usernames have been anonymised to "user"+ a unique string of numbers. These values match the username columns in the other files in this data set. The unique numerical string is re-used in the other anonymised fields, as noted below.
idnumber	Moodle supports the use of an idnumber separate from the username. This was not used in this data set.
firstname	First names have been anonymised to "First name " + that user's unique string of numbers.
lastname	Last names have been anonymised to "Surname " + that user's unique string of numbers.
email	email addresses have been anonymised to the user's unique numerical string + "@example.com"
emailstop	1 if the user chose to disable email notifications, 0 otherwise
icq	anonymised to "icq " + user's unique numerical string, if content found.
skype	anonymised to "Skype ID " + user's unique numerical string, if content found.
yahoo	anonymised to "Yahoo ID " + user's unique numerical string, if content found.
aim	"
msn	"
phone1	anonymised to "phone1 " + user's unique numerical string, if content found.
phone2	anonymised to "phone2 " + user's unique numerical string, if content found.
institution	anonymised to "Institution " + user's unique numerical string, if content found. Note: users from the same institution will not have the same value here, as the value is generated for the individual, not the institution.
department	anonymised to "Department " + user's unique numerical string, if content found. Same note as Institution.
address	anonymised to "Address " + user's unique numerical string, if content found.
city	not anonymised. Defaults to Perth.
country	not anonymised. Defaults to AU (Australia).
lang	The language chosen by the user. Not anonymised.
theme	If user selection of theme is allowed, this contains the theme name. Not used in this data set.
timezone	Not anonymised. Optionally set by the user. Defaults to "99"
firstaccess	a datetime value of the user's first access to the system
lastaccess	a datetime value of the user's last access to the system, or end of the last session.

lastlogin	a datetime value of the user's last login prior to the current session (start of the last session).
currentlogin	A datetime value of the user's most recent login if they are currently logged in
lastip	anonymised to "lastip " + user's unique numerical string

mdl_badge_issued.csv (2354 records, plus header)
This file contains the records of all badges issued to users who were enrolled in the Aug 16 session of "Teaching with Moodle," even badges awarded outside that course. Two fields have been added to this table prior to export: the anonymised username and the badge name.

Field	Description
username	The anonymised username, imported from the mdl_user file to allow links to user records
name	The name of the badge, imported from mdl_badges table prior to data export
dateissued	A datetime value of when the badge was issued to this user
visible	1 if the badge is visible to users, 0 otherwise

mdl_course_modules (75 records plus header)
This file contains records describing each activity in the "Teaching with Moodle" course. These values are not anonymised, but only a limited description of the activity is included in the source table. The name of the activity type (module) has been included from mdl_modules.

Field	Description
name	Module name (from mdl_modules)
id	course module id: a unique identifier of the activity.
course	The course in which the activity appears. 1 indicates the Front Page course. 10464 is the August 2016 session of "Teaching with Moodle."
module	the identifying number for this module type (corresponds to the "name" column)
instance	each activity in the course has a unique instance within its module type
section	the section ID in which the activity appears. The home page has two sections ("12" and "13"). The "Teaching with Moodle" course has sections "58964", "58965", "58966", "58967", "58968", and "56989".

idnumber	Each activity can be assigned a unique alphanumeric string by the course creator. Not used in this data set.
added	A datetime value indicating when the activity was added to the course.
score	
indent	The indent level of this activity within the course sections
visible	1 = the activity is visible to students
visibleold	deprecated
groupmode	1 if the activity defines group use
groupingid	if a specific grouping has been selected for the activity, it is defined here. Not used in this data set.
completion	Whether the completion-tracking facilities are enabled for this activity. 0 = not enabled (database default) 1 = manual tracking, user can tick this activity off (UI default for most activity types) 2 = automatic tracking, system should mark completion according to rules specified in course_modules_completion
completiongradeitemnumber	If an activity requires a certain grade to complete (e.g. quiz) the relevant grade item number is defined here.
completionview	if viewing the activity is one of the components to complete it, this is set to 1.
completionexpected	a datetime value by which the activity was expected to be completed. Not used in this data set.
showdescription	if the activity is set so that its description shows on the course page, this = 1.
availability	complex description of availability criteria, e.g. the user is accessing via a mobile app or has completed a required activity.
deletioninprogress	used internally by the system while an activity is being deleted.

mdl_course_modules_completion.csv (41204 records with header)

This file contains records of each user's completion of each activity in the course. The anonymised username has been imported from mdl_user to allow linking to those records. Note that the maximum possible number of rows approximates 75 (activities) x 2170 (users) = 162,750 completion records. However, not all activities have completion criteria.

Field	Description

username	The anonymised username, imported from the mdl_user file to allow links to user records
id	A unique value per activity x user combination
coursemoduleid	The id of the activity from the mdl_course_modules table
completionstate	The new state of the activity completion for this user. If this is "0" it indicates that the user had previously completed the activity, but has been set to an incomplete state (by the user, by the instructor, etc.) Available states: 0 = not completed [if there's no row in this table, that also counts as 0] 1 = completed 2 = completed, show passed 3 = completed, show failed
viewed	Tracks whether or not this activity has been viewed by this user. NULL = we are not tracking viewed for this activity 0 = not viewed 1 = viewed
timemodified	A datetime value of when the user's completion state changed

mdl_grade_grades_history.csv (87738 records plus headers)

This table keeps a historical record of individual grades for each user and each item, exactly as imported or submitted by modules. Use MAX(usermodified) and GROUP BY itemid, username to restrict the data set to the most recent grade for each user per item.

Note that if every activity in the "Teaching with Moodle" course has been graded for every participant, there would be 75 (activities) x 2170 (users) = 162,750 grade records. However, not all activities have grades.

Field	Description
username	The anonymised username, imported from the mdl_user file to allow links to user records
id	The unique id of this grade history record
action	created/modified/deleted constants
oldid	Deprecated
source	What caused the modification? manual/module/import/...
timemodified	the time this grade was last modified
loggedusername	the anonymized username of the person who last modified this grade
itemid	The grade item this grade belongs to
rawgrade	If the grade is a float value (or has been converted to one)
rawgrademax	The maximum allowable grade when this was created, stored here to record the value at the time the grade was stored, because teachers might change this for an activity.

rawgrademin	The minimum allowable grade when this was created, stored here to record the value at the time the grade was stored, because teachers might change this for an activity.
rawscaleid	If this grade is based on a scale, which one was it? Stored here to record the value at the time the grade was stored, because teachers might change this for an activity.
usermodified	The userid of the person who last modified this grade (set to 0 in this data set)
finalgrade	The final grade (cached) after all calculations are made
hidden	show 0, hide 1 or hide until date
locked	not locked 0, locked from date
locktime	automatic locking of final grade, 0 means none, date otherwise
exported	date of last grade export, 0 if none
overridden	indicates grade overridden from gradebook, 0 means none, date means overridden
excluded	grade excluded from aggregation functions, date means when excluded
feedback	grading feedback
feedbackformat	format of feedback text
information	optional information
informationforma t	format of information text

mdl_logstore_standard_log.csv (346,584 records plus headers)
This table contains entries for each "event" tracked by the Moodle logging system, and is the source for all the Moodle "log" reports (e.g. https://docs.moodle.org/32/en/Logs). For an overview of the logging system introduced in Moodle 2.6, see https://docs.moodle.org/dev/Logging_2 Also see https://docs.moodle.org/dev/Event_2#Properties

Field	Description
username	The anonymised username, imported from the mdl_user file to allow links to user records
id	The unique id of this log entry record
eventname	The event class name in path format, delimited by \ e.g. '\core\event\course_viewed'
component	The component that generated the log entry, e.g. core or mod_(module name), using frankensytle

action	What action was taken, e.g. viewed, searched, deleted, etc.
target	The type of object the action was directed toward
objecttable	The suffix of the database table the object is stored in
objectid	The id of the object within its own table
crud	Create/Read/Update/Delete, a simple way of categorizing actions
edulevel	1 = Teaching - actions performed by a teacher, e.g. updating a resource; 2 = Participating - actions performed by a student, e.g. posting in a forum (including a teacher posting in a forum); 3 (or 0) = Other - actions performed by a user with a role other than teacher or student, or performing an action not directly related to learning, e.g. viewing or updating a profile. See https://docs.moodle.org/dev/Event_2#Level_property for more information.
contextid	The ID of the context in which the event occurs
contextlevel	The level of the context in which the event occurs. This tells you if it was a course, activity, course category, etc.
contextinstanceid	Based on context level this may be course id , course module id, course category, etc. (event->contextinstanceid)
courseid	This is used only for contexts at and below course level - this can be used to filter events by course (includes all course activities)
relatedusername	Used when the event can act on someone other than the person who triggered it, i.e. an unenrol event would likely contain the id of the user being unenrolled. (Set to anonymized username of relevant user in this data set.)
anonymous	Was this event anonymous at the time of triggering?
other	Any other fields needed for event description - scalars or arrays, must be serialisable using json_encode(). Floating point numbers cannot be used.
timecreated	Time when the event was triggered.
origin	How the event was triggered, e.g. web, cli (command-line interface), cron, ws (web services), etc.
ip	IP address of the user who triggered the event (anonymised in this data set)
realusername	Anonymized username of the user when using logged-in-as feature