

Data Category	Description	Data Location
1. Course content	XML files defining the course, including e-text, assessment problems, and links to videos (videos themselves are stored on YouTube).	github (LMS workflow) MongoHQ database (Studio workflow)
2. Tracking logs	Text files recording student interaction "events" with the edX platform, including low- level GET and POST requests to the web server, as well as higher-level server records of assessment responses (answers, grades, question location).	Disks on individual servers in the Amazon Web Services cloud
3. Student state data	SQL records ("StudentModule" table) of student state for each "XModule", including last viewed location within course, state of responses (including correctness / incorrectness) for each assessment problem, position within videos and sequences.	Amazon RDS (mysql) databases in the cloud, on "instance" servers
4. Student identification data	SQL records of username, userid, name, user profile, courses enrolled in, mapping to external authentication tables.	Amazon RDS (mysql) databases in the cloud, on "portal" servers
5. Forum data	noSQL documents recording forum contributions, comments, responses, votes, links to course content.	MongoHQ database, accessed by comments service running on Heroku (Fall 2013 and later) mysqlldb (pre Fall 2013)
6. Wiki data	SQL records of wiki entries made by individual students.	Amazon RDS (mysql) databases in the cloud, on "instance" servers

Source: Subcommittee Report on edX Recorded Course Data - Content, Format, and Analysis Procedures, by Joseph K. Blitzein, Isaac Chuang, Dec 17, 2012, HarvardX Research Committee.