



Understanding and exploring, how XAPI is capable of
improvising online learning methods.

By
Robert Hitler,
Srishti institute of Art, Design and Technology
Capstone project 2019

Acknowledgment

This project would have not happen without help from many people. I am fortunate to get this project in Adobe. I would like thank Mr. Ashish Garg, Mr. Ramesh Srinivasaraghavan and Mr. Gokul Krishna Paravasthu, who were my mentors from adobe. I would like to thank Mr. Pratyush Raman, who was my mentor in college. And I specially thank Mr. Tridib Roy Chowdhury and Mrs. Anuradha Roy Chowdhury, for helping me get this project in adobe.

Contents

Abstract	-	3
Chapter 1:	introduction	4
Chapter 2:	research and insights	6
Chapter 3:	process/methodology	9
Chapter 4:	problem statements and results	14
Chapter 5:	conclusion	19
Reference:	-	21
Appendix:	-	22

Abstract

The world around us is changing as new things are invented and discovered. What would have happened if learning were not part of humans' nature? We wouldn't have gotten civilized or in simple words, we would have lived a life of an animal. It's understood that learning is part of our culture, and we learn things for knowledge, to survive and make things better around us. There are different platforms where learning happened in the past and we are going to see one medium or a platform which is needed for our future of learning; which is online learning. Understand why is, online learning is important in for future and try to understand how can it be structured and improvised. We are going explore the tools and standards or technology used in improvising the learning platform. Our focus here will be to explore a dataset collected from online education or online learning and find key factors that can be identified to make the course better. Then, I will convert the data into a format which can be further used for a greater purpose to understand data.

Chapter 1

Introduction

I have been interested in understanding the structure of education and see how can it be distributed to everyone equally, and I got an answer when I came across the term technology in education, which means online education. Online education has given a platform where anyone can learn things they are passionate about and from anywhere, all one need is a mobile phone or a laptop with an internet connection, but only sometimes one has to pay some amount of fees for accessing some specialized course or tutorials. Everyone may think why we need an online learning platform while one can learn by going to institutions or organization?

“Our charge was to accept the challenge of an Information Age and acknowledge, at the conceptual as well as at the methodological level, the responsibilities of learning at an epistemic moment when learning itself is the most dramatic medium of that change.” The era we are living in is of the information age, where there are much of data and information produced and that gradually leads to discoveries and inventions and that leads to the change in job fields and change in educational structure.

Information is growing rapidly and with new job fields emerging, learning becomes difficult because everything cannot be taught in institutions in this rapidly changing information world. Everyone and everything around us gets adapted to the new changes and similarly, our learning methods are getting adapted to online learning, so that it allows the freedom of choosing what one likes and gives the ease for one to adapt to a new job by learning it online.

Online learning is a virtual platform for collaborating with others like sharing ideas, skills, helping each other globally. We saw how online learning has grown and has become important for our future of learning and now, we get why it's important to structure it and administer it, so let us move to the focus of this project that is exploring the backend of online learning where we will understand what is done to improvise online learning platform. As part of the project's experiment, I am going to get an academic dataset, study and analyze it and pursue factors which will lead to a new understanding or new ideas to improve the online platform. For this experiment of understanding the backend or administration tools and technology behind online learning, I was fortunate to get help from my mentors in Adobe, who helped me to find what I wanted to understand. Adobe Captivate Prime is one of the software products of Adobe company, and I took up a small project which allowed me to experiment.

Chapter 2

Research and insights

As we are trying understanding the backend of online learning, here, we are looking into Adobe captivate prime an LMS (Learning Management System) which provides all the features of administration that helps online learning platforms like MOOC (Massive Open Online Courses) to be structured, published, monitored and supported. LMS (Learning Management System) LMS capable of many things and some important features are:

- Tracking student performance
- Online chat room
- A platform for collaborative learning
- Inbuilt wiki
- Quizzes and assignments
- Displays point board or scoreboard
- Email integration
- Skill gap analysis
- Dropbox (study materials)
- and many more things ...

Mainly, LMS is a platform where it provides full support to an educator or anyone can create course content to create, publish, monitor and update course content and helps a person with full support throughout his/her journey in learning the content. In the past when online learning began, it was integrated with an e-learning standard termed SCORM (Shareable Courseware Object Reference Model), which tracks student activities and make that data shareable between applications and other shareable platforms.

SCORM captures online activities like pass/fail, time (start and end), and completion of the course but doesn't capture every activity on an online course, so e-learning platform needed something more and then xAPI (Experience API) was introduced also called Tin can. This experiment or project is directly related xAPI standard where we will explore its features and see what role it plays in an online learning platform. XAPI is also called the next generation of SCORM, but it can do a lot more than SCORM.

XAPI (Experience API) is the advanced e-learning standard that is capable of tracking all possible learning activities of a user. It not only tracks the completion of the course but tracks the detailed activities like which course the student enrolled for, how many exercises or sessions

completed, the time when the activity started and ended, is the assignment finished or not, what all the questions attended and not attended, how many times the mouse was clicked, etc. XAPI is also capable of tracking learning activities outside the course, as we all know that learning happens even outside the course in an informal way, which also important to be tracked. XAPI has this special feature where student activity is tracked both online and offline. As xAPI captures every learning experience of people, it called Experience API.

This xAPI standard is used in adobe captivate prime to track all the learning activities of the student or any user. LMS with xAPI is powerful enough to produce a great educational platform. what exactly happens to all the data collected or tracked by the xAPI e-learning standard? All this data collected or captured by xAPI is sent to one central hub called LRS (Learning Record Store) a huge database specifically designed to store xAPI data. The LRS not only receives and store data but also makes the data shareable for a much larger purpose, and we will see that purpose in the later portion of this project. We will also explore the xAPI format in which the data are saved, sent and received by the LRS. This data stored in LRS are anonymized so that anyone can download this xAPI data for analyzing those raw data and try to understand the relationship between data. This data is used for visualization for making sense of the data and

find some pattern which will help improvise course methods and solve the course related problems. For example, sometimes the problem may not be with the course but with the student and this is can be seen when the data is analyzed and visualized.

Chapter 3

Process/ methodology

We explored the technology like LMS, LRS with xAPI standard and now we are going to see the process or journey of raw data collection to analyze data. This is the heart of the project; I mean the important part of the project. For this experiment, I got guidance from my mentors from Adobe, but I didn't get to work with data captured from adobe captivate but followed a similar process with the help of my mentors.

The experiment started by browsing different academic data related sites to download some data for the experiment. As I told before that this data is anonymized and anyone can download it, so I looked into different sites like academic torrents, KD nuggets and Open University (analyze KMI open) and downloaded few data sets. Then, selected one data set to work on, and the selection happened by seeing how rich was the data, types of variables the data contained and whether it suits my need to

continue the work. The dataset that I selected was from academic torrents under Educational Process Mining (EPM), had enough data to study. This dataset contained:

- **Features.txt** - contains the names of the variables
- **Activities.txt** – contains the names of activities (the sites accessed by the student)
- **Activities_info.txt** – contains information about what the activity site means.
- **Features_info.txt** – contains information which tells what exactly the variable in the data means.
- **Readme.txt** – contains the information about the course and the data
- **Exercise_info.txt** – contains the exercise numbers in each session and reference number to access the exercise in the given website.
- **Grades_info.txt** – contains the information about the grades
- **Logs.txt** – this contains the attendance of each student in each session as there were six sessions in total
- **Process folder** – this folder contains the process data, which have thirteen variables like student id, session id, time started, time ended, duration (spent in each activity), mouse wheel scroll, mouse wheel click, mouse right click, mouse left click, mouse movement, keystrokes.
- **Final_grades.xlsx** – contains data with each student's id with grades.
- **Intermediate_grades.xlsx** – contains intermediate grades of each student in each session.

This dataset was rich enough to work on for this project duration and has some elements that I was looking for like duration spent by the student on each activity and mouse activities.

There are two types of data defined; one is intentional (or primary) data type and another is consequential (or secondary) data type. "Primary data is that which is used empirically, that is, analysis of this data leads directly to particular themes or conclusions." And, "Secondary data is that which is used to support the findings of the primary data. The secondary data is that which a 'by-product' of the research was; that is, it was not specifically required for the research." In this dataset, data was all primary data type and only feature.txt, grade.txt, etc....were secondary data types.

The data that are important for this project are those process data, final_grades.xlsx, and logs.txt data. I have given a brief on what it contains and now we will look into it in detail. The process folder holds six subfolders (six sessions) and each folder contains series of text files of the student activity data, so there are 115 students enrolled for the course, so each folder contains 115 student files. And each file contains thirteen variables (named the variables in the table below Table.1) and thousands of rows holding the data of each student; the entire session folder holds the same variables and data type but different session's data. The table

below shows the example table where the thirteen columns are split into two parts, as there is no space to place the complete table.

session_id	student_id	exercise_id	activities	start	end	duration	mouse_wheel
1	1	Es	Other	2.10.2014	2.10.2014	0	0
1	1	Es	Aulaweb	2.10.2014	2.10.2014	218	0
1	1	Es	Blank	2.10.2014	2.10.2014	0	0
1	1	Es	Deeds	2.10.2014	2.10.2014	154117	6

Table.1 process data table (column 1-8)

mouse_w_	mouse_lef	mouse_rig	mouse_mo	keystroke			
0	0	0	84	0			
0	4	0	397	0			
0	0	0	59	0			

Table.1 process data table (column 9-13)

The final_grades.xlsx contains data on the unique ID of the student as it is anonymized, and it has 19 columns and 115 rows as only 115 students enrolled the course. I have a put a sample of the table below in Table.2 and even this table is split into two parts.

Student ID	ES 1.1 (2 points)	ES 1.2 (3 points)	ES 2.1 (2 points)	ES 2.2 (3 points)	ES 3.1 (1 points)	ES 3.2 (2 points)	ES 3.3 (2 points)	ES 3.4 (2 points)
1	2.00	3.00	1.00	0.50	1.00	2.00	2.00	2.00
2	2.00	3.00	2.00	0.50	1.00	2.00	0.00	2.00
4	2.00	3.00	1.00	0.50	1.00	2.00	0.00	2.00
5	2.00	3.00	2.00	1.50	1.00	2.00	2.00	2.00

Table.2 student grade sheets (first 1-9 columns)

ES 3.5 (3)	ES 4.1 (15)	ES 4.2 (10)	ES 5.1 (2)	ES 5.2 (10)	ES 5.3 (3)	ES 6.1 (25)	ES 6.2 (15)	TOTAL (100)
3.00	15.00	10.00	2.00	10.00	3.00	25.00	13.00	94.5
3.00	15.00	2.00	0.00	5.00	1.50	5.00	0.00	44.0
0.00	3.00	4.00	0.00	1.50	0.00	5.00	5.00	30.0
3.00	3.00	2.00	1.50	9.00	1.50	2.00	1.00	38.5
3.00	15.00	10.00	1.00	2.50	0.00	20.00	12.00	78.0

Table.2 student grades sheet (column 10-18)

I will attach the link to the files in the appendix for download.

In this dataset, the log.txt, final_grades.xlsx and process data have the direct correlations and took some time to understand this relationship in detail. In process data, there were fewer than 115 students in each session where it supposed to be 115 files as 115 students enrolled and this justified by the log.txt data which show the attendance and it is noticeable that not everyone attended every session. In the final_grades.xlsx table, the variables are named after the exercise and the first number in each variable show that it belongs to which session, for example, the third column of Table.2 above has the variable ES 1.2 and

the first number 1 shows that the exercise is from session one. Each exercise or exam is given marks and marks in all exercises all together are the total mark for 100. There are two spreadsheets in final_grades.xlsx, so there were two exams conducted. And it's mentioned that some attempted the exam twice, while some only attended either first or second.

I have data in the proper structure, and I understood the different variables in the dataset now moving to the phase where I have the preconceived questions that I think will lead to a new finding or to understand this course better.

Chapter 4

Problem statements and results

In portion above, we explored technologies and standards supporting online learning, and we explored a dataset to find some key factors which will help improvise the learning method or to help students if he/she needs help.

The experiment of analyzing data in this project is to prove that xAPI is capable of tracking and capturing detailed activities of any user and every data have its key factors. In this dataset, data like duration and

activities are captured, which tells us that this much x duration was spent on this Y activity, so here duration acts as a special key factor to understand the activity data. When these key factors related to one another, we can see a pattern that may solve a solution or give a deeper understanding of the methods used.

While analyzing the data, these are the questions I listed down that may get answered and help educators or trainers to understand their student's status and their learning methods.

1. *How many students who had enrolled for the course could finish the assignments and exams?*
2. *Did, all the students who finished the course manage to get good grades?*
3. *When I look into the activity's table, it contains different categories like a text editor, deed's simulation machine and much more part of the course and one of the categories was other, which means the student spent his time outside the course. Did the student spend more time outside or inside the course and how does that affect the grades?*

4. *How many students attended every exercise and exam but couldn't succeed in getting good grades?*

5. *How many people took exams for the second time and still didn't succeed in getting good grades?*

These questions pre-conceived and I know the possible outcome that will be answered when this data is taken to the next phase that is, data visualization. And data that I see as the key factor are mouse activities (mouse click, mouse scroll) and keystrokes.

Does this mouse activities show some pattern when related to the questions? To be more specific, when time spent on activities is related to keystroke and mouse activities, will it show any correlation between data? For example, the relation between mouse activities and time spent on main activities can affect final_grades.xlsx.

Result

Here, we are pursuing a solution to improvise this course, to do that a deeper understanding of the course and on the student experience is needed. And to get that understanding, these key factors and other data are seen in different combinations. As a result, my experiment will prove that a combination of the relations between data and key factors put

together a new pattern can emerge and may lead to a solution or an understanding of the problem.

This analyzed data after structured and found all the possible relations between data or key factors like mouse activities, count of keystrokes, or duration, this data is sent for data visualization. In data visualization, all the questions are answered and give proper understanding to the target audience (educators or trainers).

But this project is only to prove what can be done with xAPI data, so we are not going to visualize the data here but I am going to take the visualization part as a different project. We saw what xAPI is capable of and how it helps online learning and as part of the experiment, I had to work with xAPI format, where I had to convert the dataset to xAPI format.

The xAPI format looks like this:

```
"actor": {
  "mbox":
"mailto:dilstar7micheal@rediffmail.com",
  "Id": "1"
},
"verb": {
  "id": "http://adlnet.gov/expapi/verbs/viewed",
  "display": {
    "en-US": "viewed"
  },
  "timestamp": {
    "start time": " 2.10.2014 11:25:33",
    "end time": " 2.10.2014 11:25:34",
    "Duration": " 280"
  }
}
```

```

    },
    "object": {
      "id": "http://example.com/activities/ other",
      "definition": {
        "name": {
          "en-US": " other"
        }
      },
      "session": "1"
    }
  }
}

```

The xAPI captures activities and stores it in xAPI format as shown above and get an understanding we will go line by line.

“Actor” is the key or container which saves personal information about the user, where “mbox” holds email-id of the person as it is unique id and the name or other information.

“Verb” stores information about the action performed by the user and other information like time spent. “Display” shows what exactly the user performed, did he watch, or did he write, or did he just scroll the pages, etc....

“Object” is a key which holds values like where the user acted. If he watched something, what did he/she watch, was it video tutorials in the course or other videos outside the course? The object here is *video*.

And xAPI format customized under the actor, verb and object framework. And as I was working with some anonymized data and that data was not downloaded directly from an LRS so data was in a no

extension format and I had to convert those data into a xAPI format for experiment purpose. There were some online tools to convert these data but it not very effective enough, so I wrote a simple code in python to convert these data into a xAPI format. XAPI is written in JSON format but with some specific framework like actor and verb. I will add the code in the appendix.

Chapter 5

Conclusion

Data produced is more and those complex data encoded with number and strings doesn't make sense with a glance. To make sense and derive valid information from those complex raw data we need data analysis and data visualization. This apply to data from online learning as well but to solve problem we need data which is very detailed and that is fulfilled by xAPI.

So I conducted an experiment to see the kind of data is captured from by xAPI and to prove that these data given a relationship between different variables can draw a pattern to understand or can give a solution to solve a problem or to answer a question. In the experiment I divided and grouped different variables and to derive a valid information.

Here I had data related to mouse activities like mouse distance covered while an activity, mouse clicked while an activity, etc. and my finding is to see whether this mouse Activities affect or connect with other variable and draw valid information. Like, if a student is taken good grades or a poor grade does it have anything do with the mouse activities or does this mouse activities show some kind of user behavior so that it help educator or trainer to understand student's or trainee's performance and help them do well in future.

So here my finding remains as a question – which will lead to an answer in my next project of data visualization. This experiment conducted proves that xAPI is capable of collecting different detailed data which will further lead us to different possibilities to solve a problem or to get a deeper understanding of online learning methods, when grouped and matched to see the data in different angle.

Through this project I conclude that detailed data captured by xAPI has the ability to solve kinds of problem in online learning and help to improve online learning, as rich data helps to solve a problem in many different possible ways.

Reference

1. Davidson, Cathy N., and David Theo. Goldberg. *The Future of Thinking: Learning Institutions in a Digital Age*. Cambridge, MA: MIT Press, 2010.
2. Elle.brown@ltgplc.com, "XAPI Layer 2: Record Any Learning Experience (Informal Learning)," XAPI.com, accessed May 05, 2019, <https://xapi.com/layer-2-record-any-learning-experience-informal-learning/>.
3. "Learning Record Store (LRS) vs Learning Management System (LMS) - SCORM," SCORM.com, , accessed February 21, 2019, <https://scorm.com/learning-record-store-vs-learning-management-system/>.
4. HRAMIAK Alison, *A Method for the Analysis of Data from Online Educational Research* 4 (2005): , <http://shura.shu.ac.uk/2372/>.
5. Carliner, Saul. *An Overview of Online Learning*. Amherst, MA: HRD Press, 2004.

Appendix

1. Link that to download the data set from -
<http://academictorrents.com/details/e24e083cc337695bb84a2b68707695579c0ab4d8>
2. Code to convert the data to xAPI format is given below:

csv_xapi.py (to convert process data to xAPI format)

```
import csv,json

session_1 =
open("C:/Users/USER/Desktop/Adobe/epm_pbi/final_capstone_files/initial_csv_data/merge_1.csv",'r') #path of
the file that needed to be converted from csv to json

session_json_1 =
"C:/Users/USER/Desktop/Adobe/epm_pbi/final_capstone_files/merge_json_1.json"
#path of the file which will store the json data after
conversion from csv

csv_read = csv.reader(session_1) #csv.reader
reads the file as a csv file

student ={} #defining a empty dictionary

with open(session_json_1,'w') as json_session_1:
#keeping the file open to save the converted data
    for row in csv_read: # loop to read each line in
the csv file and convert in the defined format
        st_data = {'actor':{

"mbox":"mailto:dilstar7micheal@rediffmail.com",
                    "id":row[1],
                    },
                    "verb": {
                    "id":
"http://adlnet.gov/expapi/verbs/viewed",
                    "display": { "en-US": "viewed"
},
                    "timestamp":{'start
time':row[4],
```

```

'end
time':row[5],
'Duration':row[6]}
        "mouseActivities":{
"mouseWheelScroll":row[7],
"mouseLeftClick":row[9],
"mouseRightClick":row[10],
"mouseMovement":row[11]}
        "keystrokes":row[12]
    },
    "object": {
        "id":
"http://example.com/activities/"+row[3],
        "definition": {
            "name": { "en-US": row[3]
},
            "session":row[0]
        }
    },
    },
    },
    student.update(st_data) #updating the
empty dictionary with the formatted st_data dictionary
    d = json.dumps(student,indent=4)
#organizing the data with indentation
    json_session_1.write( d + ',')
#writing the opened json file with the converted data

```

grades_xapi.py (to convert grades sheet into xAPI format)

```

import csv, json #importing csv and json library

final_grades_1 =
open("C:/Users/USER/Desktop/Adobe/epm_pbi/final_capstone_files/final_grades_first_exam.csv",'r')
#opening and reading the file that has be converted in
xapi

```



```

final_grades_2 =
open("C:/Users/USER/Desktop/Adobe/epm_pbi/final_capstone_files/final_grades_second_exam.csv",'r')
#opening and reading the file that has be converted in xapi

final_1_json="C:/Users/USER/Desktop/Adobe/epm_pbi/fin_grade_json_1.json"
#creating a json file to save or store the converted data

final_2_json="C:/Users/USER/Desktop/Adobe/epm_pbi/fin_grade_json_2.json"
#creating a json file to save or store the converted data


fin_1 = csv.reader(final_grades_1)      # defining a variable to read and save csv data using csv library
fin_2 = csv.reader(final_grades_2)      # defining a variable to read and save csv data using csv library


grad_1 ={} # creating a empty dictionary for later use
grad_2 ={} # creating a empty dictionary for later use


with open(final_1_json,'w') as first_exam: #keeping this file open to convert and write csv to xapi
    for row in fin_1:
#defining a for loop to read all the content or element in the file row by row
        exam_data_1 = {"actor":{
#definig a xapi format with a dictionary
            "id":row[0],

            "mbox":"mailto:dilstar7micheal@rediffmail.com"},
            "result":row[17]
        }
        grad_1.update(exam_data_1)      #updating the empty dictionary with data

```

```

        first_data = json.dumps(grad_1,indent=4)
#defining a variable to write and store data from
dictionary with indentation
        first_exam.write(first_data)
#finally writing the converted data to the file we
craeted .

with open(final_2_json,'w') as second_exam:
#keeping this file open to convert and write csv to
xapi
    for row in fin_2:      #defining a for loop to read
all the content or element in the file row by row
        exam_data_2 = {"actor":{                #definig a
xapi format with a dictionary
                        "id":row[0],

"mbox":"mailto:dilstar7micheal@rediffmail.com"},
                        "result":row[17]
                    }
        grad_2.update(exam_data_2) #updating the
empty dictionary with data
        second_data = json.dumps(grad_2,indent=4)
#defining a variable to write and store data from
dictionary with indentation
        second_exam.write(second_data) #finally
writing the converted data to the file we craeted .

```

3. Link to converted xAPI data -

https://drive.google.com/open?id=1Ze2qsJ3qk6WarF1IkATwT_30o1ecaxGU