

Data Event Taxonomy - UTS iLabs

August 2022

**Overview**

Coles have engaged ClearXP as the learning platform provider for organisational-wide training. The ClearXP platform collects fine-grained analytics data in accordance with the Experience API specification (see <https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md#partone> for background). This document outlines the schema and a sample collection of data events collected when users (also referred to as actors) interact with the platform.

The goal of the UTS iLabs project is to analyse this analytics data to determine common usage patterns and place learners on disparate scales according to exhibited behaviours. The scales themselves are open to interpretation and can be defined by the students performing the analysis, however, the below are two possible examples:

* **Explorer** – the degree to which a user utilises search and filtering functionality.
* **Social** –the degree to which a user shares content, rates or provides feedback.

The output of the analysis could be equated to a Myers-Briggs Type Indicator assessment where each user would be assigned a numeric value for each scale. The combination of these scales may then be used to assign a descriptive behaviour to different learner types.

The data also includes a field indicating whether the user is a manager. Managers have access to more platform functionality than regular users so there is an assumption that their behaviours may diverge and require independent analysis.

The dataset is comprised of:

* Anonymised data from July 1st, 2021 until June 30th, 2022
* 365 files segmented into 24-hour chunks in JSON format
* 39,630,462 records with total size of 20.5 GB

**Caveats**

Note that the system conducts a best-effort approach to calculating data fields and that in some cases, many of these fields may be absent if the system cannot reliably determine how to calculate them.

For example, if the system only has two data points on completely different days for a learner completing the activity then it will be unable to determine the duration taken because duration is the time spent within the activity itself (two data points within a short period of time are required).

In general, data may be more reliable the more recent the record has been recorded as efforts to populate these fields improve.

**Event Taxonomy**

|  |  |
| --- | --- |
| Event | Description |
| **User** initialised **Learning Hub** | Captured when a user launches the Learning Hub |
| **User** experienced **(Screen)** | Captured when the user navigates to a different screen inside the Learning Hub. The following screens exist:   * My Learning * Explore Learning * My Calendar * Dashboard > Team * Dashboard > Activities * Dashboard > Overdue * Dashboard > Due Soon * Profile > Notifications * Profile > My Tasks * Profile > Learning Transcript |
| **User** searched for“(Search Term)” from **(Screen)** | Captured when the user searches for learning, including the screen they were on when they initiated the search. Searches can be initiated from either My Learning or Explore Learning screens. |
| **User** selected **(Filter)** with “(Filter Value)” from **(Screen)** | Captured when utilising the dropdown filters. There are two possible filters:   * Brand Filter * Category Filter   The value selected in the dropdown is also recorded as well as the screen the user was on when they selected the filters. |
| **User** viewed (Count) **Search Results** | Captured after search results are returned including the number of results included. |
| **User** shared **(Activity)** with **(User)** | Captured when the user shares an activity with another user in the system. Learners can share with multiple users at once but a separate event is recorded for each user. |
| **Manager** assigned **(Activity)** to **(User)** | Captured when a manager assigns an activity to a user in their team. |
| **Manager** removed **(Activity)** from **(User)** | Captured when a manager removes a previously assigned activity from a user in their team. |
| **User** registered into **(Activity)** | Captured when a user is registered into an activity. Note that there are two mechanisms that will cause this event:   * When registered by an admin into assigned learning * When self-registering into an activity from search or a notification (in this case the **other** field will reference the Learning Hub). |
| **User** unregistered from **(Activity)** | Captured when a user removes a self-registered activity from their assigned learning. As per registration, the **other** field indicates when this has occurred inside the Learning Hub. |
| **User** rated **(Activity)** (Rating) with (Feedback) | Captured when a user rates a piece of content, includes both a quality rating (from 1 – 5) as well as a free-text feedback response. |
| **User** launched **(Activity)** from **(Screen)** | Captured when a user launches an activity including the screen (or activity) they were on when they launched it. Activities could be launched from any of the following places:   * My Learning * Explore Learning * Search Results * Notifications * My Tasks * Any other Activity |
| **User** completed **(Activity)** with(Score) | Captured when a user completes an activity, sometimes includes a score if the activity reports a score. |

**Analytics Fields**

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| **activity** | String | The ID of the activity that the record pertains to. |
| **actorId** | String | The ID of the actor that the record pertains to. |
| **ancestors** | Array<Object> | Resolved ancestry tree for the activity including the ID of each ancestor activity and the depth of the tree. For example, if this record pertained to accessing a Dashboard Tab, this object would look like the following:  [  { activity: "(**Learning Hub** Activity ID)", depth: 0 },  { activity: "(**Dashboard** Activity ID)", depth: 1 },  { activity: "(**My Team** Activity ID)", depth: 2 }  ] |
| **completion** | Number | Whether this record signifies a completion should be recorded or not (0 = not completed, 1 = completed) |
| **contained** | Analytics Object | Includes a nested record within this one – with the exception of activity references, all other fields in this table can appear as sub-fields of **contained**. Used for compound records such as “Manager shared Activity with User” – the “manager shared activity” portion appear as top-level analytics fields but **contained** will include the “user” portion (along with any other recorded fields). |
| **duration** | Number | The duration spent (in milliseconds) when recording this interaction. |
| **Is Manager** | String | A flag indicating whether the user is a Manager. **Y** indicates the user is a manager and **N** indicates they are not. |
| **Is New Starter** | String | A flag indicating whether the user had been an employee for less than 3 months at the time the record was captured. **Y** indicates the user is a new starter and **N** indicates they are not. |
| **id** | String | The unique identifier for this record. |
| **other** | Array<Object> | Indirect activity references for this record. Typically includes the screen or an activity a user is viewing when the event is recorded against a different activity to the one the user is currently viewing. I.e. “Launching Induction Training” – Induction Training is the activity, but **other** will include the activity the user launched from. |
| **rating** | Number | A number between 1 – 5 indicating how the actor has rated the activity. Present for rating interactions only. |
| **response** | String | Response data attached to this record, if applicable. |
| **timestamp** | Date | The date of the interaction. |
| **Time Zone** | String | The user’s Time Zone at the time this record was stored. |
| **type** | String | The type of activity being tracked stored a unique Inverse Relational Identifier. |
| **verb** | String | The ID for the type of interaction that this record pertains to. |

**Activity Metadata Fields**

Additional metadata for each activity can be found inside the **activities.json** file uploaded to the S3 bucket.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| **Authoring Tool** | String | The industry tool used to build the content. |
| **Brand** | String | The organisational brand that the content pertains to (i.e. Supermarkets, Express, Liquor or SSC) |
| **Category** | String | The Learning Category represented by the activity, i.e. Leadership, Safety, Operations, etc. |
| **CMS Type** | String | The type of content (Video, Page, Portal, etc.) |
| **Department** | String | The organisational Department covered by the learning content (i.e. Bakery, Service, Dairy, etc.) |
| **Duration** | Number | The *expected* duration it will take to complete the content. |
| **Functional Owner** | String | The business function that is responsible for managing this content. |
| **id** | String | The unique identifier for the activity. |
| **Objectives** | String | Any learning outcomes or objectives covered by the content. |
| **Passing Score** | Number | The score required to pass the content (if applicable). |
| **Program** | String | The wider learning program that this activity belongs to. |

**Sample Data**

In its most simple form, an analytics record may look like the following:

***User A experienced Explore Learning***

|  |
| --- |
| {  "verb" : "http://adlnet.gov/expapi/verbs/experienced",  "completion" : 0,  "activity" : "https://organisation.hub.clearlrs.com/explore",  "duration" : 490,  "id" : "0d92ef81-d686-424b-a196-89ae3244d04d",  "timestamp" : "2021-09-13T10:22:06.519Z",  "actorId" : "account|https://organisation.clearlrs.com|RANDOM\_ID",  "Is Manager" : "N",  "ancestors" : [  {  "activity" : "https://organisation.hub.clearlrs.com",  "depth" : 0  },  {  "activity" : "https://organisation.hub.clearlrs.com/explore",  "depth" : 1  }  ]  } |

A more complex example of a compound record may look like the following:

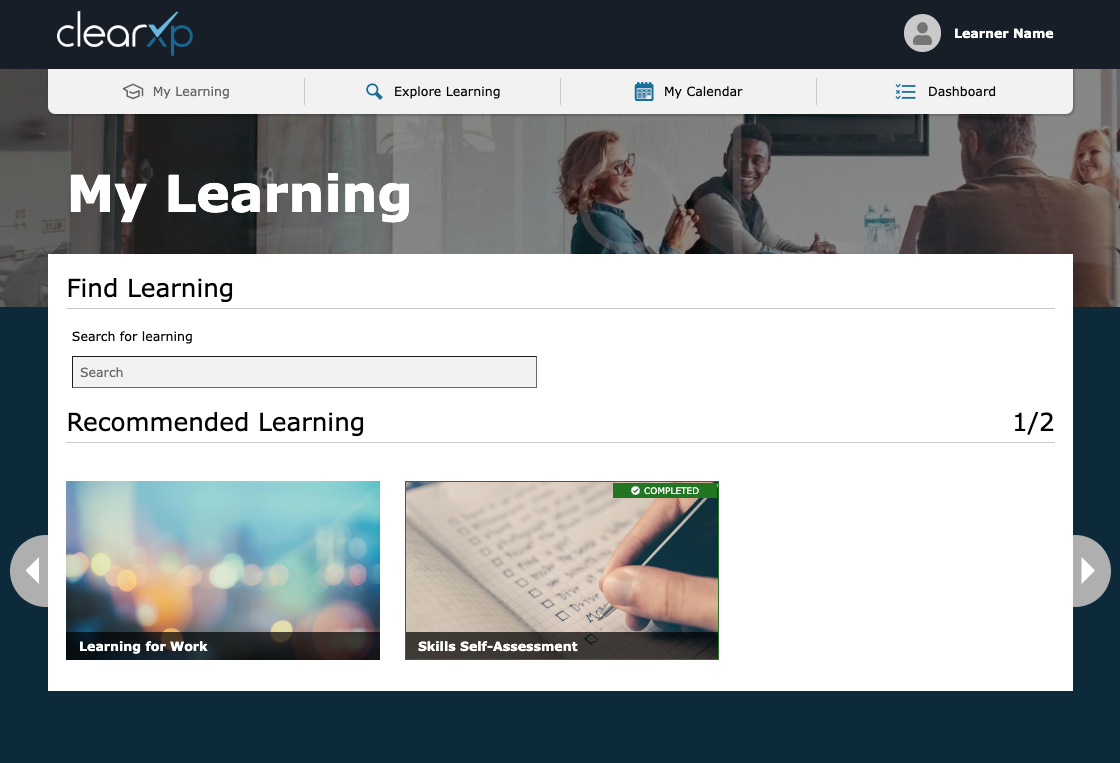
***User A assigned Induction Training to User B***

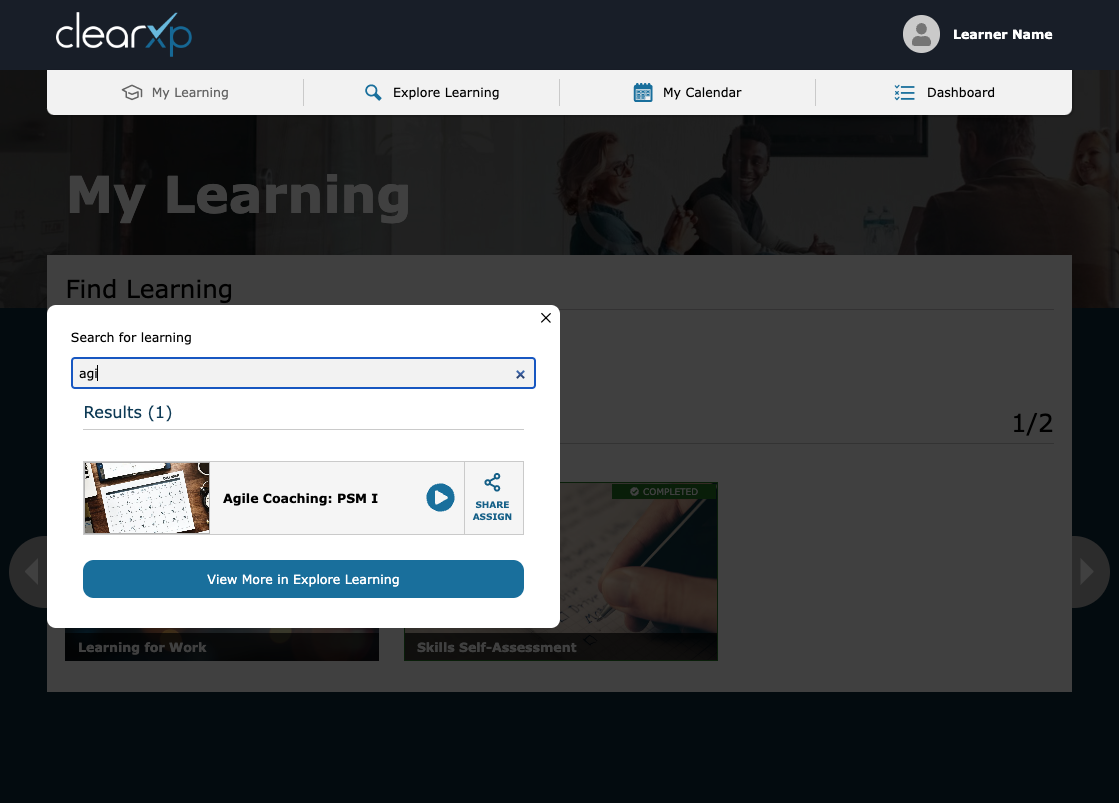
|  |
| --- |
| {  "verb" : "https://w3id.org/xapi/dod-isd/verbs/assigned",  "completion" : 0,  "activity" : "https://organisation.clearlrs.com/activities/ INDUCTION\_TRAINING",  "type" : "http://adlnet.gov/expapi/activities/module",  "duration" : 100279,  "id" : "fad37801-7a55-4624-8c64-9eaf82bf99f6",  "timestamp" : "2021-09-13T10:29:03.749Z",  "contained" : {  "verb" : "http://adlnet.gov/expapi/verbs/registered",  "actorId" : "account|https://organisation.clearlrs.com|RANDOM\_ID\_2"  },  "actorId" : "RANDOM\_ID\_1",  "Is Manager" : "Y",  "Time Zone" : "Australia/Melbourne",  "other" : [  {  "activity" : "https://organisation.hub.clearlrs.com",  "depth" : 0  }  ],  "ancestors" : [  {  "activity" : "https://organisation.clearlrs.com/activities/ INDUCTION\_TRAINING",  "depth" : 0  }  ]  } |

**Platform Screenshots**

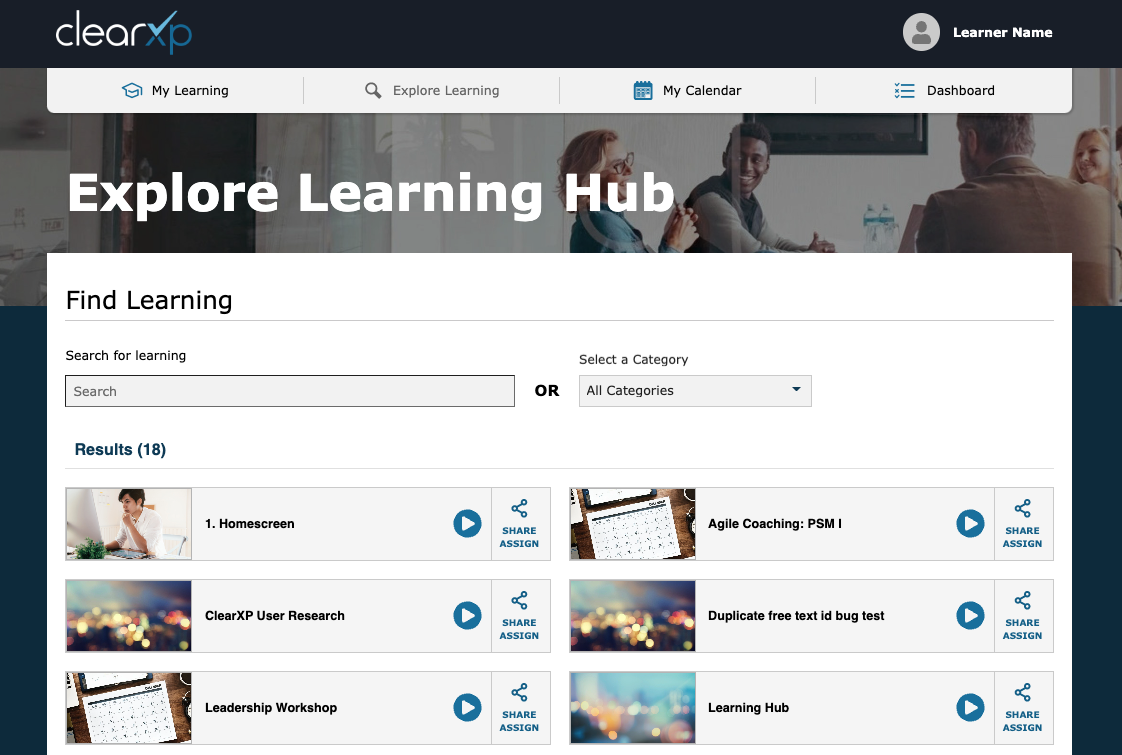
The below screenshots highlight different screens and platform functions and can be used as a reference guide for interpreting data events.

**My Learning** screen with Quick Search function

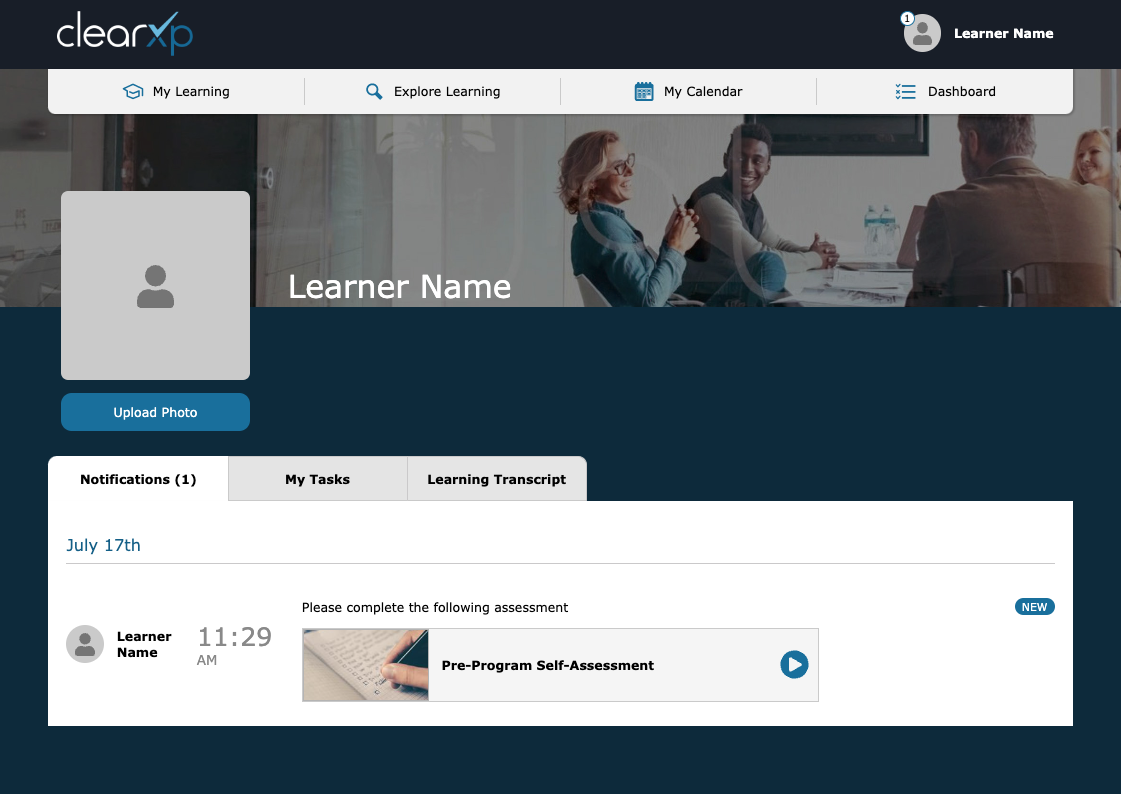
****

****

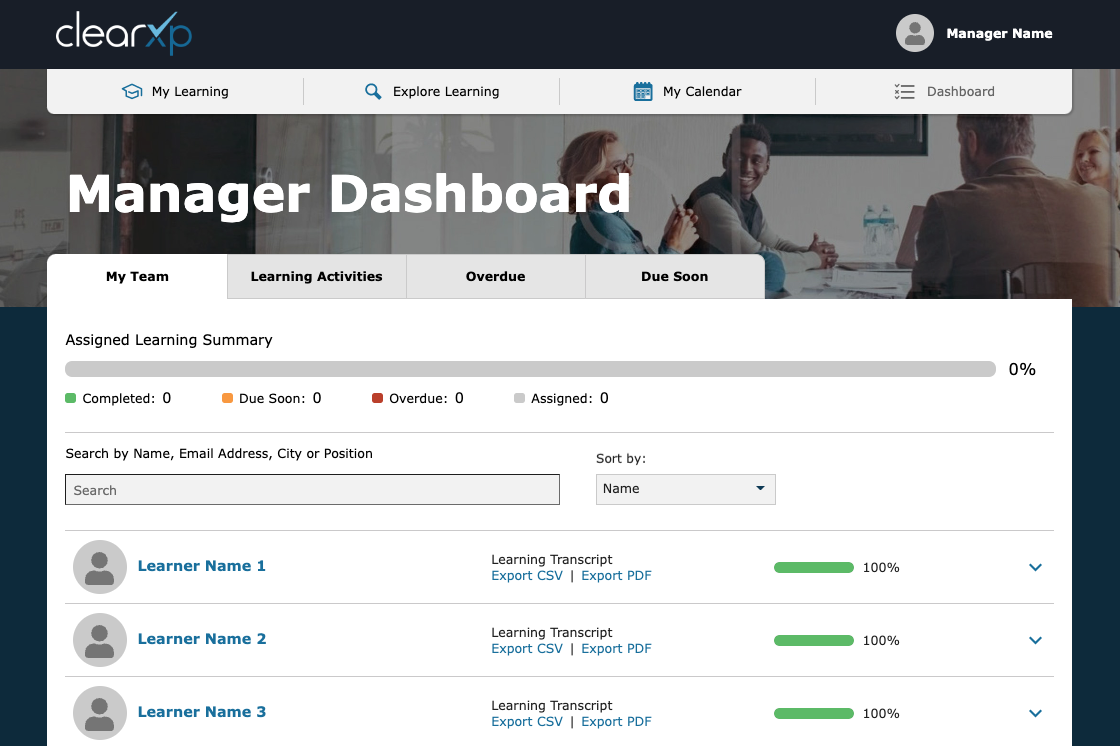
**Explore Learning** screen with Search and Category filtering functions

****

**Profile** screen with sub-screens for **Notifications**, **My Tasks** and **Transcript**

****

**Dashboard** screen with sub-screens for **My Team**, **Learning Activities**, **Overdue** and **Due Soon**

****