Activity Data and Paradata

A Briefing Paper

By Lorna M. Campbell and Phil Barker

This briefing introduces a range of approaches and specifications for recording and exchanging data generated by the interactions of users with resources.

Such data is a form of Activity Data, which can be defined as "the record of any user action that can be logged on a computer". Meaning can be derived from Activity Data by querying it to reveal patterns and context, this is often referred to as Analytics. Activity Data can be shared as an Activity Stream, a list of recent activities performed by an individual. Initiatives such as OpenSocial, ActivityStreams and TinCan API have produced specifications and APIs to share Activity Data across platforms and applications.

While Activity Streams record the actions of individual users and their interactions with multiple resources and services, other specifications have been developed to record the actions of multiple users on individual resources. This data about how and in what context resources are used is often referred to as Paradata. A specification for recording and exchanging paradata has been developed by the Learning Registry, an open source content-distribution network for storing and sharing information about learning resources.

Activity data

Activity data is a broad term used to describe:

"The record of any user action (online or in the physical world) that can be logged on a computer."

Jisc, Exploiting Activity Data In The Academic Environment

http://www.activitydata.org/What_is_Activity_Data.html

OVERVIEW

The above definition is taken from the synthesis report of the Activity Data Programme funded by the Jisc in 2011.

Activity data is generated when users interact online with content, systems, and other users. In educational institutions many systems, such as student record systems, virtual learning environments and library information systems store data about the actions of students, teachers and researchers. For example, activity data may be generated when teachers or students interact with a course, search a library catalogue, share or like a useful resource.

The Activity Data Programme Synthesis Report identified three categories of activity data:
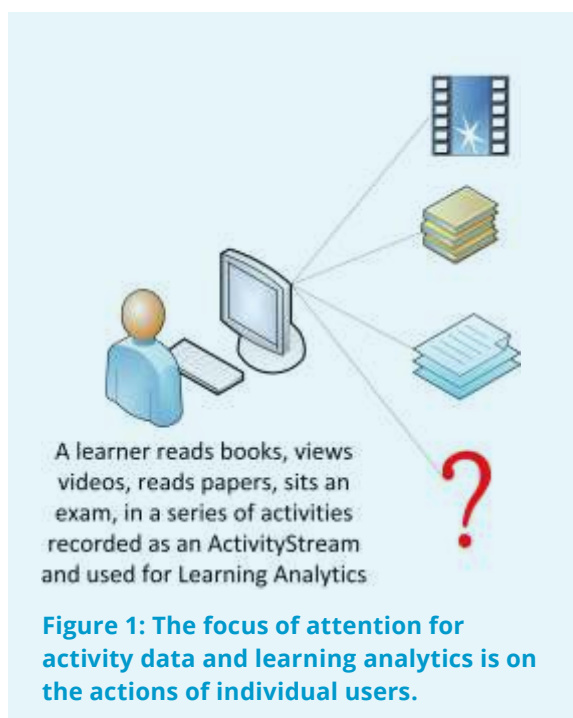
Access – recording access to systems, e.g. log in / log out, passing through routers and other network devices.

Attention – interacting with applications, e.g. page impressions, menu choices, searches.

Activity – records of transactions, e.g. purchases, lecture attendance, book loans, downloads, ratings.

Analytics

Meaning can be derived from activity data by querying and examining the data to reveal patterns and context. Querying data is often referred to as analytics. Analysing activity data can help to increase understanding, improve decision making, tailor interactions, use resources more effectively and improve the user experience. If the activity data being queried is generated by learners' activities and interactions, this may be

A learner reads books, views videos, reads papers, sits an exam, in a series of activities recorded as an ActivityStream and used for Learning Analytics

Figure 1: The focus of attention for activity data and learning analytics is on the actions of individual users.

referred to as learning analytics. Learning analytics can be defined as

"the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs."

LAK11 https://tekri.athabascau.ca/analytics/

LEGAL AND ETHICAL ISSUES

While activity data may be anonymised, attribution (being able to identify individual users) can be useful in analysing and building services based on activity data. However this raises legal and ethical issues around data protection, who owns the data and how it may be used.

There is a legal requirement governed by the Data Protection Act 1998 that users must give their permission whenever data is collected. Employees and learners usually do this when they join an educational institution or a course and sign up to the relevant policies and procedures. However data can only be used for the purposes for which it has been collected, as notified to the data subjects. In order to share activity data, it must either be anonymised, or permission must be requested from the data subjects. If anonymised or permissible data is shared or published it should also be licensed appropriately.

EXAMPLES

At a basic level, activity data can be used to indicate popularity e.g. Facebook likes, retweets, number of times a book is borrowed from an institutional library, how often a learning resource is accessed, number of students signed up for a course.

Activity data can also highlight links between resources and contexts; e.g. showing what resources are popular for specific learning activities and objectives, identifying which reading lists a resource has been added to.

At a more sophisticated level recommender systems can be based on the analysis of activity data; e.g. people like you bought x, y and z, students who completed this module went on to study these modules, teachers who found this learning resource useful also used these.

Activity data can also potentially be used to make behavioural correlations; e.g. between students' library usage, grades and performance, and to make interventions; e.g. by identifying students who are struggling with a module and may be at risk of dropping out.

FURTHER INFORMATION

CETIS Analytics Series (ISSN 2051-9214)
http://publications.cetis.ac.uk/c/analytics
Discovering the Impact of Library Use and Student Performance by Brian Cox and Margie Jantti, published on July 18, 2012
http://www.educause.edu/ero/article/discovering-impact-library-use-and-student-performance
Exploiting Activity Data in the Academic Environment by Tom Franklin, Helen Harrop, David Kay, Mark van Harmlen,
http://www.activitydata.org/
Jisc Activity Data Programme,
http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata
Data Protection Act 1998,
http://www.legislation.gov.uk/ukpga/1998/29/contents

Activity Streams

An activity stream is a list of recent activities performed by an individual. Activity streams are closely associated with social networks and other social media platforms.

The simplest activity stream model is based on the actor–verb–object archetype; e.g. Jane shared a photograph. Other information may be added such as where–when–how; e.g. Jane shared a photograph on flickr 10mins ago via iPhone.

IMPLEMENTATIONS

Facebook. The most commonly cited example of an activity stream is Facebook's Timeline. Some learner management systems (e.g. Canvas) are also starting to develop analytics tools to generate activity stream-like profiles for learners and teachers.

Google has added social media reporting functionality to Google Analytics. The Google Social Data Hub is a platform that social network sites can use to integrate their data in the form of a global Atom/RSS Activity Stream feed which is pushed to the hub using PubSubHubbub (PSHB) an open server-to-server, web-hook-based publish/subscribe protocol.

OpenSocial is a specification, originally developed by Google and MySpace, for a component hosting environment and a set of common APIs for social network applications to access data and functions from social networks. More recently, OpenSocial has been adopted as a general runtime environment for allowing partially trusted components and third party services to run in web applications. OpenSocial 2.0 onwards incorporates support for a range of open web technologies including ActivityStreams.

FURTHER INFORMATION

Facebook Timeline,
http://www.facebook.com/about/timeline
Canvas Analytics, http://www.analyticscanvas.com/
Capturing The Value Of Social Media Using Google Analytics,
http://analytics.blogspot.co.uk/2012/03/capturing-value-of-social-media-using.html

Google Analytics rolling out social network activity streams: Paradata heaven? by Martin Hawksey, http://mashe.hawksey.info/2012/03/google-analytics-rolling-out-social-network-activity-streams-paradata-heaven/
pubsubhubbub, https://code.google.com/p/pubsubhubbub/
Analytics Social Data Hub, https://developers.google.com/analytics/devguides/socialdata/
OpenSocial http://opensocial.org/

ActivityStreams

The ActivityStreams initiative is being developed to address the proliferation of sites generating social activity data feeds. Most of this data is produced in the form of RSS or Atom feeds, however there is considerable diversity in the form of these feeds, which can lead to interoperability problems and places an increasing burden on aggregators. In addition, simple Atom and RSS feeds do not capture the richness and complexity of much social network activity.

"The ActivityStreams specification aims to define a convenient and consistent way to syndicate social activities around the web.

The activity in ActivityStreams is a description of an action that was performed (the verb) at some instant in time by someone or something (the actor) against some kind of person, place, or thing (the object). There may also be a target (like a photo album or wishlist) involved.

The stream in ActivityStreams is a feed of related activities for a given person or social object."

ActivityStreams wiki
http://wiki.activitystrea.ms/w/page/1359261/FrontPage

Facebook, MySpace, Google, SAY Media, IBM and Microsoft have all contributed to the development of the ActivityStreams specification documents.

ActivityStreams currently consists of three specifications:

- Activity Base Schema (draft)
- JSON Activity Streams 1.0
- Atom Activity Streams 1.0

Plus additional extensions covering audience targeting, responses, verb definitions and priority extensions.

The draft Activity Base Schema defines a base set of object types and verbs. There are currently 90 verbs and 31 common objectTypes. Objects of any specific type are permitted to introduce additional optional or required properties. Activity streams may be serialised using either the JSON or Atom format.

There are generally three different implementation methods to provide ActivityStreams: polling, push-based, and real-time/streaming.

IMPLEMENTATIONS

Although ActivityStreams has been deprecated by Facebook, it is currently implemented in MySpace, Github, Drupal, Yammer and Jira.

A new open source implementation is also being incubated by the Apache Software Foundation to support its use in OpenSocial platforms. Apache Streams aims to develop a scalable server for the publication, aggregation, filtering and re-exposure of enterprise social activities via the ActivityStreams specification.

The Apache foundation is also developing Rave, "a web and social mashup engine that aggregates and serves web widgets" and includes support for ActivityStreams.

FURTHER INFORMATION

ActivityStreams http://activitystrea.ms/
Activity Base Schema (draft)
http://activitystrea.ms/specs/json/schema/activity-schema.html
JSON Activity Streams 1.0
http://activitystrea.ms/specs/json/1.0/
Atom Activity Streams 1.0
http://activitystrea.ms/specs/atom/1.0/
Implementation Scenarios,
http://wiki.activitystrea.ms/w/page/19394614/Implementation-Scenarios
Apache Streams http://streams.incubator.apache.org/
Apache Rave, http://rave.apache.org

Tin Can API

"The Tin Can API (sometimes known as the Experience API) ... makes it possible to collect data about the wide range of experiences a person has (online and offline)."

Tin Can API Overview **http://tincanapi.com/overview/**

OVERVIEW

Tin Can API was initially developed by Advanced Distributed Learning to overcome some of the limitations of their SCORM specification arising from the assumption that learners were working within an LMS or VLE that would deliver content and track their progress. Tin Can API makes no assumption that the learner is in any formal learning environment, but rather allows independent tools to communicate this information (automatically or at the learners prompting) to other systems, for example to a "personal data locker" known as a Learning Record Store (LRS). The data is then available for use by learning analytics systems or as evidence in a learner's eportfolio.

The Tin Can API is a RESTful protocol that transmits statements of the basic form "I did this" in JSON using an "actor", "verb", "object" with "result" in "context" syntax that is reminiscent of an extended ActivityStream. The specification provides an initial range of values for the verb statements such as "experienced", "attended", "created".

IMPLEMENTATIONS

The Tin Can API adopters' page lists many assessment systems, authoring tools, VLEs, games and simulation developers that have implemented the specification in their products.
http://tincanapi.com/adopters/

Rustici Software's SCORM Cloud is a hosted LRS, that lets anyone create their own LRS account.
http://scorm.com/scorm-solved/scorm-cloud-features/

Tappestry is an app for Apple or Android that allows users to record threads of learning activities, including information about the resources used and a reflection on the outcome. Threads can be shared within a group or stored in an LRS **https://www.tappestryapp.com/**

FURTHER INFORMATION

Tin Can API **http://tincanapi.com/**

The layers of Tin Can API **https://www.tappestryapp.com/**

TinCanApi development wiki: (includes technical documentation) **http://tincanapi.wikispaces.com/**

Tin Can API Comparison with ActivityStreams **http://www.adlnet.gov/tin-can-api-comparison-with-activity-streams**

Paradata

Paradata is a form of metadata that records how, and in what context, a learning resource is used. While metadata generally attempts to record objective or authoritative descriptions of a resource, paradata records the opinion of the users and how and where a resource has been used. Paradata is generated as learning resources are used, reused, adapted, contextualized, favourited, tweeted, retweeted, shared. This type of information tends not to be captured by more traditional cataloguing techniques, which aim to describe what a resource is, rather than how it may be used. Paradata complements metadata

People (teachers, learners, managers) view, edit or copy a video in various contexts, in a series of actions that are recorded as paradata about that video
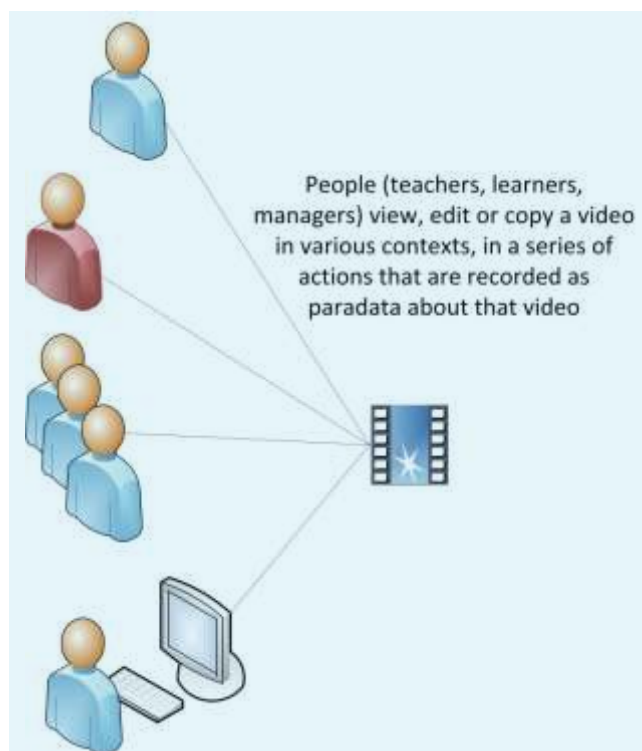
Figure 2 : The focus of attention for paradata is on the resource, the aggregated data from many anonymous users is analysed to provide information about the resource.

by providing an additional layer of contextual information. By capturing the user activity related to the resource, paradata can help to elucidate its potential educational utility.

In this context, the term paradata was first used by the US National Science Digital Library (NSDL) in early 2010 to describe data about user interactions with learning resources within the NSDL's STEM Exchange. Later that year the term was adopted by the Learning Registry, an initiative initially funded by the U.S. Department of Education and the U.S. Department of Defense. The Learning Registry is an open source decentralized content-distribution network of peer-to-peer nodes that can store and forward information about learning resources. The primary purpose of the Learning Registry is to share descriptive metadata and social usage paradata across diverse educational systems.

Paradata differs from ActivityStreams in that it enables complex aggregations of activities to be recorded; e.g. High school English teachers taught using this resource 15 times during the month of May 2011.

(Learning resource paradata should not be confused with survey paradata which is administrative data about the process by which survey data is collected.)

EXAMPLES

On the simplest level paradata can be used to record how users interact with a resource by viewing, downloading, sharing, liking, commenting, tagging, etc.

Information about users may also be recorded; e.g. age, educational level, geographical location, etc.

Paradata can also record contextual information by linking resources with educational standards and curricula, pedagogic approaches and methodologies.

High school English teachers taught using this resource 15 times during the month of May 2011.

```
{
    "activity": {
        "actor": {
            "objectType": "teacher",
            "description": ["high school","english"]
        },
        "verb": {
            "action": "taught",
            "measure": {
                "measureType": "count",
                "value": 15
            },
            "date": "2011-05-01/2011-05-31"
        },
        "object": "http://URL/to/lesson/"
    }
}
```

Figure 3: An example paradata statement, taken from the Learning Registry's Paradata in 20 Minutes or Less guide https://docs.google.com/document/d/12nvvm5ClvLxSWptlo52rTwIDvobiFylYhWLVPbVcesU/edit?hl=en_US

It is worth noting that while learning analytics generally refers to analysis of data about learners, paradata refers to data about learning resources.

SPECIFICATIONS

Although Learning Registry paradata is informed by the ActivityStreams approach, it differs from ActivityStreams in that it enables the description of aggregations of activities. Like ActivityStreams, there are three main parts to a basic paradata statement - an actor does verb to an object, e.g. "A teacher taught the lesson located at this URL." However paradata also adds:

Descriptions which provide context to actors, verbs or objects.

Measurements which provide data about magnitude, e.g. the number of times a verb occurred over a period of time.

Dates which record when an action took place.

Paradata can be regarded as an extended and altered version of JSON ActivityStreams.

ACTIVITY DATA AND PARADATA

IMPLEMENTATIONS

The JLeRN Experiment was a Jisc funded project which explored the feasibility of setting up a Learning Registry node and contributing and analysing data, in order to better understand the potential of the Learning Registry in the UK Higher Education context. The project set up a test node, successfully published data to it and built a Node Explorer tool based on the LR slice API, which is now available on Github.

JLeRN Experiment, **http://jlernexperiment.wordpress.com/**
JLeRN Node Explorer,
   http://jlernexperiment.wordpress.com/tag/node-explorer/
Node Explorer on Github, **https://github.com/jlern**

Sharing Paradata Across Widget Stores (SPAWS) was a Jisc funded project involving the University of Bolton, the Open University, KU Leuven and IMC. The aim of the project was to build on the Learning Registry to share usage data, such as reviews, ratings, and download statistics between web app stores of widgets and gadgets for educators. The project team successfully created an open source software library that developers can use to add "paradata sharing" to app stores, and integrated it into Edukapp, a cross university web app store.

SPAWS Project, **http://scottbw.wordpress.com/tag/oerri/**
SPAWS Software Library,
   https://github.com/scottbw/spaws
Edukapp, **http://code.google.com/p/edukapp/**

Kritikos, which was created by the ENGrich project based at the University of Liverpool, is a customised search engine for visual media relevant to engineering education. Using Google Custom Search (with applied filters such as tags, file-types and sites/domains) as a primary search engine for images, videos, presentations and Flash™ movies, Kritikos pushes and pulls corresponding metadata and paradata to and from the Learning Registry. A user-interface enables academics and students to add further data about particular resources and how they are being used. This information is then published to a Learning Registry node and used to order any subsequent searches.

Kriticos Visual Media Search for Engineering
   Education, **http://engrich.liv.ac.uk/**
ENGrich Case Study,
   http://jlernexperiment.wordpress.com/2012/10/17/taster-a-soon-to-be-released-engrich-learning-registry-case-study-for-jlern/

FURTHER INFORMATION

NSDL Network Paradata,
   http://nsdlnetwork.org/stemexchange/paradata
Learning Registry, **http://www.learningregistry.org/home**
Learning Registry Technical Guides,
   http://www.learningregistry.org/documents
Learning Registry Paradata Specification V1.0
   https://docs.google.com/document/d/1IrOYXd3S0FUwNozaEG5tM7Ki4_AZPrBn-pbyVUz-Bh0/edit?hl=en_USI

About this Briefing Paper

Title: Activity data and paradata

Authors: Lorna M Campbell and Phil Barker

Date: 1 May 2013

URI: **http://publications.cetis.ac.uk/2013/808**

For more information on the Jisc CETIS publication policy see **http://wiki.cetis.ac.uk/JISC_CETIS_Publication_Policy**

About CETIS

CETIS are globally recognised as leading experts on interoperability and technology standards in learning, education and training. We work with our clients and partners to develop policy and strategy, providing impartial and independent advice on technology and standards. CETIS are active in the development and implementation of open standards and represent our clients in national, European and global standards bodies and industry consortia, and have been instrumental in developing and promoting the adoption of technology and standards for course advertising, open education resources, assessment, and student data management, opening new markets and creating opportunities for innovation.

For more information visit our website: **http://jisc.cetis.ac.uk/**