Learning Analytics Interoperability - a survey of current literature and candidate standards

Adam R Cooper,
Centre for Educational Technology and Interoperability Standards, University of Bolton, UK.
a.r.cooper@bolton.ac.uk

The ease with which data can be transferred without loss of meaning from a store to an analytical tool - whether this tool is in the hands of a data scientist, a learning science researcher, a teacher, or a learner – and the ability of these users to select and apply a range of tools to data in formal and informal learning platforms are important factors in making learning analytics and educational data mining efficient and effective. This paper describes, in summary form, the findings of a survey into: a) the current state awareness and research or development into this problem of seamless data exchange between multiple software systems, and b) standards and pre-standardisation work that are candidates for use or experimentation.

interoperability, standards

Scope and Motivation

Defining Terms

Both “learning analytics” and “interoperability” are likely to be understood in different ways. In addition, not all aspects of potential interest in the intersection of these two concepts will be considered. This section seeks to clarify the range of meaning and scope adopted by the author.

Learning analytics has been variously defined (Barneveld, Arnold, & Campbell, 2012) and distinctions drawn between it and educational data mining (Siemens & Baker, 2012) but, for the purposes of this work, the author's preference for an inclusive description (Cooper, 2012) will be adopted. Learning analytics is used as short form for:

> “The process of developing actionable insights in respect of educational aims and objectives through problem definition and the application of statistical models and analysis against existing and/or simulated future data.”

In addition to being an umbrella for the fields of learning analytics and educational data mining as deduced from the activities of researchers and innovators in these two communities, this description is intended to include applications and analytical attitudes that have not yet been explored but which address questions about the purposeful development of knowledge, skills, attitude, habits, etc. There is nothing in this definition to exclude questions at a much larger scale than the individual learner or cohort of learners, for example questions of educational policy at a national level.

Interoperability has also been variously defined but for the purpose of this work, a focus on interoperability of ICT is appropriate. The IEEE provides a commonly accepted definition (Geraci, Katki, McMonegal, Meyer, & Porteous, 1991) of ICT interoperability as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged." The IEEE definition works well for some cases, such as whether this document in electronic form can be read by different pieces of software and presented in a readable form. It also fits the case of back-office or business-system integration. In both cases, everyone except the software developers and integrators expects not to be dealing with the data. This may also be the case for packaged learning analytics software but it is not necessarily so when we consider more exploratory forms of analytics. Rather than the “systems or components” using the information that has been exchanged, this work will also consider that an analyst might be the user of the information, so an adapted form of words is used as a working definition:

> “Interoperability the ability of two or more systems or components to exchange information and for the information that has been exchanged to be used.”

The abbreviation LIA, for “Learning Analytics Interoperability” will be used in the text to refer to the concept outlined above.

Why is Interoperability Relevant?

A thorough discussion of the relationship between interoperability and benefits is not appropriate for this work and the following paragraphs grossly over-simplify the issues in order to provide some brief contextualisation for the survey. The reader who wishes to explore these matters further, including the complex relationship between context and benefit, should consult the literature, for example (Choi & Whinston, 2000) or (Gasser & Palfrey, 2007), for some typical analysis.

Two common metaphors for interoperability are friction-reduction and modularity. In the former case, we can think of interoperability as reducing then energy wasted in manual data transformation or integration. The latter case, modularity, captures the idea that software components can be more easily changed. The essential belief underlying modularity is that if a useful service can be delivered by separated but interoperable systems, A and B, then it must be possible to create a new system C to replace B. In either case. interoperability is achieved by standardisation, either formal, or *de facto* arising from convergent practice.

Friction-reduction and modularity may lead to a number of useful effects, with benefits for research and adopters:

1. A wider range of user-level services become acceptable because the underpinning data is available on a more timely basis. (friction)
2. Innovation is accelerated because good ideas and solutions can be migrated. (modularity)
3. Products and services (internal or external to an organisation) become economically viable and new markets can be opened up. (friction and modularity)

These are general effects and take an IT systems focus. In a similar vein as the short discussion of the definition of interoperability noted, where the possible role of an analyst was pointed out, there are some benefits that become more important when data are the object of interest rather than being only the means by which the system works. An additional useful effect is noted:

4. Combined data sets lead to new insights or improved validity. (friction and modularity)

The statement of the fourth useful includes cases where data comes from different sources but is about the same real-world object or situation, and cases where the data sources are largely concerned with different entities. This is not only relevant for data scientists in an institutional setting; research on learning could be advanced by access to interoperable data at large scale (Duval, 2011).

In addition to these general benefits from interoperability, the setting from which most learning analytics data comes from presents its own problems. As (Krüger, Merceron, & Wolf, 2010) note, “Learning software is not designed for data analysis and mining. Because usage data is not stored in a systematic way, its thorough analysis requires long and tedious preprocessing.” This point is echoed in several of the surveyed items.

Interoperability is relevant but it is not the only means of achieving friction-reduction or modularity. A widely available platform may achieve the same ends. Google Analytics [1] is such an example, although it can be argued that it absolutely relies on interoperability of HTTP, JavaScript, etc. Since Google Analytics can track in-page events it provides the means to avoid the need to build activity tracking capabilities from scratch. In this case, it would still be necessary to address learning analytics interoperability (LAI) questions about the semantics of what is tracked but the technical infrastructure can be taken "off the shelf", even by low budget or experimental projects.

Limitation of Scope

In order to keep the survey to a manageable size, some restrictions on scope have been adopted. Standards such as XML are not considered because they are almost entirely structural and have no domain-relevant semantics for learning analytics and educational data mining. Additionally, the survey only considers data interoperability, and not those aspects of interoperability concerned with communication protocols. Hence, IMS LTI[2] (Learning Tools Interoperability) is excluded, although it could be used to launch end-user analytics tools in tandem with a specified data API.

The scope of interoperability will be further limited to information exchange involving analytics software. The whole chain of exchanges that may occur up-stream of data transfer to the analytics tool is not in scope. For example, the NMEA[3] electrical and data standards are widely used for GPS hardware module communication and a range of scenarios can be imagined in which GPS data is ultimately used for learning analytics but standards such as NMEA are not in scope for this work.

A final limitation of scope, which is implicit from having selected an application domain (learning analytics & EDM), is that only interoperability pertaining to the data subjects will be considered. Interoperability as it pertains to the analytics process will not be considered. For example, Predictive Model Markup Language[4] (PMML) is not in scope.

Survey

Structuring the Survey

A simple structure will be used to summarise the extent of coverage of LAI by both the literature and by candidate interoperability specifications. This structure has no basis in theory or roots in a particular pedagogic outlook; it is primarily a means to organise a number of items and secondarily a prompt. As a prompt it might indicate an area of omission in what has been surveyed or a neglected topic in the commentary on LAI.

The structure has two levels:

1. The more static and intrinsic, information about an object.
2. The more dynamic and extrinsic, information about an educationally-relevant event.

---

1 The Google Analytics developers' site describes the architecture and API, https://developers.google.com/analytics/

2 http://www.imsglobal.org/toolsinteroperability2.cfm

3 http://en.wikipedia.org/wiki/NMEA_0183

4 http://www.dmg.org/

	A. The Person	B. Resources[5]	C. Services and Tools	D. Learning Activities[6]	E. Objectives & Assessment
1.	Natural facts about the person and their history prior to the learning episode in focus.	Properties of digital and physical resources and their "tomb stone" meta-data.	Attributes of the services (non IT) and tools (IT) that support education.	Learning designs, lesson and action plans, etc.	Intended learning outcomes. Attributes of an assessment instrument.
2.	Associations of the learner or teacher with other people.	Use of, interest in, or attention to the resources	The log of activity by or interaction of a person with services and tools.	What the learner produces (any medium).	Achievement information, the results of assessment, inferred knowledge.

Table 1: Structure of the survey

Literature

Formal and informal publications were sampled[7] in three ways:

1. The full text of proceedings from the LAK Dataset (Taibi, 2012) was examined for occurrences of "interoperable" or "interoperability".
2. Google Scholar was queried for "learning analytics" + interoperability and, separately, "educational data mining" + interoperability.
3. Google web search was queried with the same terms.

Cases where the reference to interoperability is incidental to learning analytics are excluded, for example where interoperability of the educational content is referred to in a paper that is largely about mining interaction logs.

Academic and Formal Publications

The LAK Dataset revealed five papers, four of which were relevant: one from the International Conference on Learning Analytics and Knowledge (Verbert et al., 2011) , two from the International Conference on Educational Data Mining (Abbas & Sawamura, 2009; A. L. Dyckhoff, Zielke, Chatti, & Schroeder, 2011) and two from a special edition of the Journal of Education, Technology and Society (A. Dyckhoff, Zielke, Bültmann, Chatti, & Shroeder, 2012; Shum & Ferguson, 2012). 16 distinct authors were identified with a group of four being represented in two papers (Chatti, Dyckhoff, Schroeder and Zeilke).

A further eight publications were identified in Google Scholar as containing "*learning analytics"* +

5 These will generally be information resources – books, electronic media, etc – but may also be physical with a practical or emotional value such as a utility object, an art-work or a place.

6 At level 1, this can be thought of as describing how entities from the other columns are to be integrated. Level 2 is taken to the residual of the possible outcomes from the activity that does not naturally fall into a different column.

7 It is acknowledged that relevant material will be missed because it does not use the term "learning analytics" or refer to "interoperability" by name. Hence "sample" is used to describe the process, which is intended to be indicative.

interoperability before relevance appeared to have dropped to nearly zero. Seven of these are of an academic character[8] and all but three of these name at least one author already represented in the LAK Dataset. In addition to these seven was an issue brief from the US Department of Education (Bienkowski, Feng, & Means, 2012). A further eight publications[9] were exposed by querying for *"educational data mining"* + *interoperability*. These publications cover a considerably longer period of time, consistent with the earlier emergence of the educational data mining community compared to learning analytics. Two thirds of these introduce authors not represented in the LAK dataset and all are of a scholarly character, including one book section.

An additional publication (Siemens et al., 2011) is known to the author but was not identified by the above process because it does not mention interoperability by name; it will be included with the above publications because of its clear links with them.

In most of the above pieces of writing, references to interoperability are marginal; few can be said to be "about interoperability". The way LAI is covered by these works will be identified as being of three kinds: assertion or argument in favour of interoperability in general; references to interoperability for a particular purpose or context; interoperability as a significant or key topic.

Assertion and argument about interoperability are usually concerned with the lack of it.

> "Lack of data interoperability among different data systems imposes a challenge to data mining and analytics that rely on diverse and distributed data. Over time, piecemeal purchases of software can lead to significant decentralization of the source of education data..." (Bienkowski et al., 2012)
>
> "LA further requires key stakeholders to address a number of challenges, including questions about handling increasing data volume, heterogeneity, fragmentation, system interoperability, integration, ..." (Chatti, Dyckhoff, Schroeder, & Thüs, 2012)
> "As a precursor to making that [data based research on learning] happen, it is important that we agree on ways to share data sets, in an "open science" approach. ... The main objective is to promote exchange and interoperability of educational data sets." (Duval, 2011)
>
> "How can such a platform [Open Learning Analytics: an integrated & modularized platform] be delivered...? Fundamentally, we require an open platform with standards for adding new "plugins". As long as developers of analytics, recommender services, visual user interfaces, and intervention strategies, comply with these standards, their work can become part of this ecosystem." (Siemens et al., 2011)
>
> Jovanovic et al focus on the database structure within a single system and the lack of syntactic interoperability: "When a developer wants to add some functionality to existing Moodle version he ... adds tables to the database model that will be used to manage data for the new set of functionalities. ... This aspect of the model complicates future extraction of information on students' activities for each new module."

---

8 (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Dietze, 2012; Drachsler et al., 2012; Duval, 2011; Fazeli, Drachsler, Brouns, & Sloep, 2012; Niemann, Scheffel, & Wolpers, 2012; Palermo, Marr, Oriel, Arthur, & Johnston, 2012)

9 (Chatti et al., 2012; Dietze, 2012; Drachsler et al., 2012; Duval, 2011; Fazeli et al., 2012; Niemann et al., 2012; Palermo et al., 2012)

> (Jovanovic & Vukicevic, 2012)
>
> "One limitation present in many widely used learning environments is the isolated nature of their learner models. Incredibly rich representations of students' knowledge are created, refined, and then discarded at the end of the school year... Coordinating the sharing and interoperability of learner models across learning environments is an area which has been the subject of considerable research... However, despite the increasing sophistication of practice and theory in this area [learner modeling], sharing of data between learner environments has not yet emerged into the most widely used learning environment." (Desmarais & Baker, 2011)

A smaller number of cases where interoperability is only touched upon make reference to a purpose or context in which it would be applicable. As previously, the sense is one of future development.

> "This paper presents an agent-based educational environment to teach argument analysis (ALES). The idea is based on the Argumentation Interchange Format Ontology (AIF) ... [to] refine the learning environment by adding more flexible interoperability."(Abbas & Sawamura, 2009)
>
> "Significant progress had been made towards the introduction of low SES [socio-economic status] student cohort tracking at the university ... discussions have occurred surrounding the interoperability between the data warehouse and the Student Information System" (Palermo, Marr, Oriel, Arthur, & Johnston, 2012)
>
> "'Free and Open' is a key expectation and dynamic within online social learning. ... Data is expected to be accessible, appropriately licensed for remixing and, wherever possible, in machine-readable formats to facilitate interoperability and avoid data or users being locked into a given platform. ... It becomes normal that SLA [social learning analytics] patterns and data are open, shareable resources for reflection, and analysis in alternative tools." (Shum & Ferguson, 2012)
>
> Castro and Alonso make a number of references to an educational ontology and relate this to Learning Object interoperability but fail to describe the ontology and are unclear about its role. (Castro & Alonso, n.d.)

The focussed references to interoperability differ considerably in their character. In some cases, interoperability is identified as an initial requirement for a particular outcome, whereas others take a more general view.

Work to develop an "exploratory learning analytics toolkit", eLAT, (A. L. Dyckhoff et al., 2011; A. Dyckhoff et al., 2012) lists interoperability as one of the software design goals to "ensure compatibility for any kind of LMS by allowing for integration of different data sources." They also note that data created while using informal learning platforms – assumed to be software used for learning but not identified as an LMS – is a relevant concern. eLAT uses a data model (unpublished) that is independent of a particular learning environment and makes available typical kinds of learning environment activity: document usage, assessment/ performance, user activity and communication. An additional paper, not identified by the sampling method described above adopts a similar approach, proposing a fairly simple data model to extract data from the Moodle LMS (Krüger, Merceron, & Wolf, 2010).

Two sources are particularly concerned with existing data models. Social information resource use – rating, tagging and bookmarking content – based around FOAF (Brickley & Miller, 2010) and CAM

("CAM Schema," n.d.) for interoperability is at the heart of the design of a prototype trust-based social resource recommender for teachers (Fazeli, Drachsler, Brouns, & Sloep, 2012). Niemann et al stands out as the only paper that sets out to give an overview of data models, in which they address the problem of choosing a data model for usage data for recommender systems (Niemann, Scheffel, & Wolpers, 2012). They also address CAM and consider the differences in approach and model with Activity Streams (Snell et al., 2011), Learning Registry Paradata ("Learning Registry Paradata Specification V1.0 ," n.d.) and NSDL Paradata ("NSDL's Technical Schema for Paradata Exchange," 2011).

A further source considers particularly the need for there to be interoperable usage data in order for effective recommender systems to be designed (Verbert et al., 2011). The emphasis in this paper is not on the operational aspects but on the research and algorithm design necessary for a recommender system to be effective. The role of interoperability in collections of data for research is also described in one account of the Pittsburgh Science of Learning DataShop, a data repository for the EDM community (Koedinger et al., 2011). This is the only account among those surveyed where interoperability of data from many independent systems is described.

Two papers, both contributed to by Stefan Dietze, are particularly concerned with Linked Data for description of educational resources. One of these is a broad literature survey in which reference to clustering and classification by data mining methods comprises only one section (Dietze et al., 2013), whereas an earlier paper is focussed particularly on resource recommender systems (Dietze, 2012). In both cases, the papers are resource-centric rather than usage-centric, in contrast to the other papers just mentioned.

Computer aided assessment is the subject of a single paper dealing with aggregation of data from multiple assessment item banks in an adaptive testing scenario (Phankokkruad, 2012). Interoperability is mentioned several times but, although XML was used, there is was no use made of existing interoperability standards for e-Assessment such as IMS QTI (Kraan, Lay, & Gorissen, 2012) and no indication that multiple systems were studied to synthesise a common model. The work does not, therefore, really address the interoperability problem and their data model is only given by partial example.

The incorporation of recommendations and adaptive hypermedia into a Java tutoring system, and the role of ontologies, is the subject of one paper, although it only describes a speculative architecture (Klasnja-Milicevic et al., 2011) and no results. The paper states that "open standards, like XML, RDF and OWL needed to be used in order to allow the specification of ontologies to standardize and formalize meaning and to enable the reuse and interoperability." The author's experience in educational technology indicates that XML, RDF and OWL are neither necessary nor sufficient for interoperability and that consensus-based shared conceptual models should be the focus of attention.

An outlier among the publications is a paper describing research into using Foundation for Intelligent Physical Agents (FIPA) standards for communication between the components of an intelligent tutoring system (Medvedeva et al., 2005). FIPA provides a common vocabulary of communicative acts and Medvedeva at al pay little attention to the data that is the subject of these communications, which is the focus of this baseline study.

	A. **The Person**	**B.** **Resources**	**C.** **Services and Tools**	**D.** **Learning Activities**	**E.** **Objectives & Assessment**
1.		Linked data for			

		resource description (Dietze, 2012; Dietze et al., 2013)			
2.	Social resource recommender (Fazeli et al., 2012) Argumentation (Abbas & Sawamura, 2009)	eLAT (A. Dyckhoff et al., 2012) Social resource recommender (Fazeli et al., 2012) Resource recommender (Niemann et al., 2012; Verbert et al., 2011; Klasnja-Milicevic et al., 2011)	Resource recommender (Niemann et al., 2012) eLAT (A. Dyckhoff et al., 2012)	Argumentation (Abbas & Sawamura, 2009)	eLAT (A. Dyckhoff et al., 2012) Adaptive testing (Phankokkruad, 2012) Learner modeling (cognitive models) (Desmarais & Baker, 2011)

Table 2: Indicative Coverage of the Application Area Associated with Interoperability

Informal Publications

Google web search added little new information to the picture once co-incidental appearances of "learning analytics" and "interoperability" were discounted. The point of the exercise is sampling to determine whether there are strong message so only the first 150 hits were inspected. From this set only four meaningful items were found that were not also references to the publications or publication-venues (e.g. a call for papers) referred to above. One of these reported on a discussion workshop exploring the relationship between learning analytics and improving educational content (Weston, 2012), a second refers to a conference panel session concerning "... learning analytics, high-stakes assessment and workplace assessment and ... the role that interoperability plays in these areas" but provides no detail, while two others made passing comments about poor coverage of interoperability in the educational technology field generally:

> "In schools we have a far wider range of problems concerning interoperability. Until all data can be made interoperable LA will not happen." (Tolley, 2011)
>
> "Interoperability of plugins and add-ons for these systems [LMSs] is an ideal or a goal, and very far from a present fact. As such, if you're developing or marketing a Learning Analytics product or service, you probably have to start with one of the major existing systems." (Stutt, 2011)

Summary

Although the importance of LAI in both operation and R&D contexts is clear from the literature it is apparent that, subject to the caveats relating to the sampling method:

1. Only a small group of people, largely researchers, have drawn attention to LAI and a significant amount of the literature has been produced by a few people. No references to LAI from software suppliers was noted.
2. There has been little work particularly focussed on LAI.
3. Information resource recommendation has dominated as the aspect of LAI being considered by

those articles with a strong focus on interoperability.
4. There has been very little work based on existing interoperability specifications from the educational technology domain.

Existing Standards

In the following, “standards” is used to refer to technical specifications with a range of depth of, and approach to, consensus process. At one end of the spectrum are specifications that have been through a formal and transparent development process and at the other end are specifications that have been proposed or implemented by a single group. From the point of view of likely utility, the standards with a more inclusive development process are likely to be superior due to the improved focus on shared concepts. At the present time, however, many of the standards listed below have not been used in an analytics or data mining role and so are likely to entirely miss attributes of value for that kind of use that are relatively unimportant for interoperability of the data-generating activities. They are, therefore, candidates with potential for further investigation, or sometimes for initial investigation from a LAI perspective.

The survey of existing standards with potential applicability to learning analytics is broken down into four, with the more semantically-neutral in the earlier sections: logging standards, general web standards, educational technology standards, and standards from governmental or sector bodies. For each standard, a synopsis of its subject matter is given, along with mention of the extent to which they are used, and whether the use is of a LAI character.

Logging

Logging standards typically fall at one end of the spectrum of domain-relevant semantics since they usually define concepts such as an event and an agent but do not specify a vocabulary for the kind of event or kind of agent, or only do so at a coarse level of distinction. A given application can use its own vocabulary within the framework that the standard provides but greater benefits accrue from the use of shared vocabularies.

Extensible Event Stream (XES) was developed for generalised logging tasks in processes such as handling insurance claims, using a complex x-ray machine or browsing a website. It was developed by researchers into work-flow management and process mining who have developed a generic open source tool-set, ProM[10], for process mining that can receive XES as an input format. ProM has been applied to educational data mining (Trcka, Pechenizkiy, & Van der Aalst, 2011), although this work describes use of MXML (“Mining eXtensible Markup Language”), the predecessor of XES. XES is not widely used but ProM provides powerful process-analytical capabilities for desktop use.

PSLC DataShop Tutor Message Format (DataShop) is intended for use in logging activity in a tutoring application; it allows for learner interactions with a tool and tutor responses to be logged. The semantics of interactions are only partially-specified, and these are recommended attribute values, so the format is generally applicable. It is the format used in the Pittsburgh Science of Learning Center DataShop, a repository of learner interaction data, which predominantly contains data from intelligent tutoring systems (ITS). Although the format is not tied to “intelligent”, the approach to skill and level reflects ITS practice. An extension of the format allows learner interactions with media – typically audio and video – to be logged.

Contextualized Attention Metadata (CAM) is essentially a logging schema for user interactions with learning environments. It is not well documented as an interoperability standard but some Java code is

10 http://www.processmining.org/prom/start

available. From the documentation, it is not clear whether CAM offers any capabilities that are not available using a better-documented format and there is insufficient detail to permit adoption.

General Purpose Web Standards and Domain-neutral Tracking Standards

Atom Syndication Format (Atom) was originally developed as a successor to RSS for blog syndication but is potentially useful for LAI because it provides structure and minimal, but extensible, meta-data. around user generated content to complement text mining of the content. Atom could be applied to any series of learner-generated content, whether or not published as a blog, and can be extended with educationally-relevant meta-data. (see LEAP2A, below). The availability of software libraries for Atom is an advantage.

Friend of a Friend (FOAF) allows the encoding of basic demographic, membership and human-relationship information, both direct ("knows") and indirect through shared endeavour ("Project") – the FOAF Core – with links to things of interest and a person's social web presence. Although FOAF is not widely used on the social web due to the dominance of closed hubs such as Facebook, the FOAF model may be useful and its semantic units can be re-used as part of a composite. It is consistent with Linked-Data/Semantic-Web approaches. FOAF is intended to be general purpose and lacks nuance in, for example, what it means for one person to know another but nuance can be added by extension.

Semantically-Interlinked Online Communities (SIOC) is concerned with representing information from diverse kinds of online community: message boards, wikis, blogs, etc. It deals with a level of detail in the structure and content and its relationship to people beyond what FOAF and Atom can describe. SIOC was conceived of to link online communities but it has potential as a generalised representation of the way learners create, revise, or respond to written items in a social or collaborative context. It does not yet appear to have been exploited as a means of getting such data into a learning analytics or EDM tool.

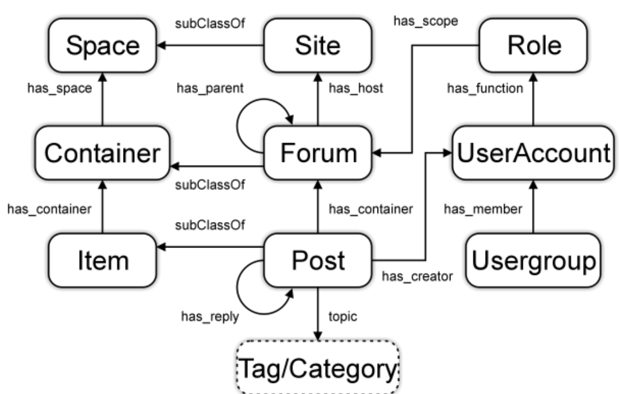

Figure 1: SIOC overview (from the specification)

Activity Streams (AS) provides a framework for recording activity and could be considered to be a form of logging. It has only a minimal set of defined concepts to allow an open-ended set of actions to be recorded; it defines the concepts needed to capture an actor, a verb, an object and a target but does not, for example, define a range for the verb other than a default value of "post". Activity Streams can be encoded using JSON or Atom. Whereas SIOC has content structure at its heart, Activity Streams have the actor (normally a person) at the centre; the verb, object and target are all optional according to use. Both could be used to convey information about a series of events in which some writing is posted/submitted. Activity streams encompasses more passive actions such as using a resource, bookmarking it, etc.

Attention Profiling Markup Language (APML) was an early effort to standardise sharing of attention profiles including browsing history, blogging, social bookmarks, tweets, etc. It failed to gain traction and has become moribund[11].

Educational Technology Standards
The first entry in this section continue the theme in the previous section: encoding information about learner use of, or attention to, resources.

National Science Digital Library Paradata and Annotation Schema (NSDL) is concerned with the exchange of: a) annotations on learning resources, including facts about specific contexts of use such as a pedagogic method, curriculum standard, teaching tips, etc, and b) a range of different kinds of usage of learning resources, including rating, recommending, downloading, tagging, citing, etc. Documentation is incomplete and use appears to be limited to NSDL for which they have defined their own XML Schema. NSDL participates in the Learning Registry[12] which has a draft paradata approach based on Activity Streams but does not require its use.

Mozilla Open Badge Initiative (OBI) is described by Mozilla as "a new online standard to recognize and verify learning" and is oriented towards the issuing and display of achievement and popular use may offer little data for analysis. The data model does, however, offer a simple structure for encoding achievement and its links to both evidence and to curriculum standards. The OBI standard and supporting software is available under open licence.

Information Model for Learning Outcomes and Competences (InLOC) permits definitions of intended learning outcomes and competences and structures over such definitions, including similarity or identity relations, to be exchanged. It includes a consideration of level and credit and mapping to subject classification. It has not been designed for analytics and by itself would only support curriculum analysis but the data InLOC can encode about what has, or has not yet, been learned has the potential to enrich analysis of achievement. It could, for example, be coupled with OBI to express what individuals learn in relation to conceptual structures of what can be learned.

e-Portfolio Portability and Interoperability (Leap2A) provides a model for describing what a learner has achieved, created, done or experienced along with information about themselves, their abilities and qualities. It permits supporting information or evidence from any source to be referred to and may include both reflections on the past and plans for the future. A binding to Atom is the preferred encoding. Leap2A was not designed for analytics and has not been used for that purpose but it could provide the means to convey content and structure and would be suited to cases where the content is created in a learner-centred process of reflection and development.

IMS Question and Test Interoperability Results (QTI) is the counterpart to a standard for e-Assessment content and has received less implementation attention than the latter. It provides the means to express results at assessment, test and item level and can express candidate responses, the associated outcome and the range of possible outcomes. Although there is little evidence of implementation of QTI Results to date, the standard development team included people experienced in high-stakes assessment.

ADL Experience API (XAPI) arose from the Tin Can Project and was formerly known as the Tin Can API. The data payload of the API is based around the approach and high level entities of Activity Streams with the addition of concepts such as context and result to make it better suited to an educational setting. The vocabularies for Activity Stream verbs and activities are not defined in the core standard but a draft list is currently under discussion[13]. The present list includes outcomes as well as activities and captures a

---

11 The wiki containing the specification is not available [access last attempted 29th April 2013].
12 http://www.learningregistry.org
13 http://tincanapi.wikispaces.com/Verbs+and+Activities [accessed April 30th 2013]

wide range of stereotype acts for both formal and informal learning settings. The Tin Can Project team have identified analytics and presented a webinar under the aegis of the Society for Learning Analytics Research[14].

IMS Learner Information Services Outcomes (LIS) is one component of an IMS standard designed primarily for the exchange of data between university student records systems and learning management systems. It is intended to exchange a student's assessment outcomes within a course or module, which is identified in the data. It does not deal with item level results or responses and is suited to summative assessment.

European Learner Mobility Achievement Information (EuroLMAI) was developed to support mobility in Europe by specifying a common means of exchanging information about educational achievement, although it is more widely applicable than this motivation. It is now a European Standard. It provides a structured means to express the information typically found in a “transcript” for a period of learning and includes high level information about the programme and the outcomes. It would be most use as a common model for expressing heterogeneous prior achievement.

Learning Analytics Tookit (eLAT) is the subject of ongoing work and, unlike previous entries, spans all kinds of data entity present in a typical learning management system (LMS) as the basis for a practical exploratory learning analytics toolkit for teacher/tutor use. eLAT does not qualify as a standard since it has not been published in a form that could be adopted and has been developed in the context of a single institution but it has been designed for use with two LMSs and a flexible view of the tools that would be clients for the LMS data so it has a strong interoperability focus.

InBloom is a US non-profit organisation that is developing a technology platform to “make personalised learning a reality for every US student”. Among the aims on the path to achieving this is the integration of data from multiple school IT systems and the use of this data to give teachers a better view of how students are performing and what learning materials and activities would best suit them. As for eLAT, their data model lacks detail but is broad in coverage.

Education Agencies and Governmental

Given the considerable resource of data collected by, or on behalf of, governments and their agencies with responsibility for regulating[15] education, it is remarkable that references to this kind of data in the learning analytics literature are so rare. In contrast, this data is often the focus of business and IT staff in higher education establishments. Regulatory data collection and the standards that have been developed to support it are quite distinct from the standards in the previous two categories in that:

1. they embed culturally- and historically-based differences in the way education is organised and managed;
2. their use is normally mandatory or a business necessity within a given sector of education;
3. the data is usually subject to much higher levels of quality control than most data of relevance to learning analytics (with the obvious exception of high stakes assessment data).

National bodies have negligible incentive to converge on a universal approach so these standards will certainly remain highly localised. This limits aspiration for an international effort on learning analytics interoperability based on these standards but the practical value of wide-spread conformity and higher levels of data quality indicates good potential in spite of idiosyncrasies.

---

14 http://www.solaresearch.org/storm/open-webinars/megan-bowe-tincan/

15 The term “regulating” is used to include both formal and informal kinds of regulation. It is assumed that data is collected with the intention of someone using it to decide on a course of action that influences the system that is monitored.

Bearing in mind point 1, above, a list of standards will not be given.

Other
Argument Interchange Format (AIF) allows for the exchange of argumentation in a form suited to researchers in argumentation, artificial intelligence and agent based systems. As such, it has a view of argumentation structure and semantics that exceeds that which is common in educational technology discourse or systems. For systems that scaffold argumentation or where free argument can be marked up, AIF could be applicable.

Summary
Table 3 lists the standards given in the previous section, along with a reference to the specification or most complete public description.

Abbreviation	**Name**	**Category**	**Reference**
AIF	Argument Interchange Format	-	(Chesñevar et al., 2006)
AS	Activity Streams	General Web	(Snell et al., 2011)
Atom	Atom Syndication Format	General web	(Nottingham & Sayre, 2005)
CAM	Contextualised Attention Metadata	Logging	(“CAM Schema,” n.d.)
DataShop	Tutor Message Format	Logging	(PSLC, 2013)
ELAT	Exploratory Learning Analytics Toolkit	Ed. Tech	(A. Dyckhoff et al., 2012)
EuroLMAI	European Learner Mobility Achievement Information CWA 16132[16]	Ed. Tech	(CEN, 2010)
FOAF	Friend of a Friend	General web	(Brickley & Miller, 2010)
inBloom	inBloom Technology Application Developer Documentation	Ed. Tech	(InBloom, 2013)
InLOC	Integrating Learning Outcomes and Competencies	Ed. Tech	(Grant, 2013)
Leap2A	Leap2A ePortfolio Portability and Interoperability	Ed. Tech	(Grant, 2011)
LIS	IMS Learning Information Services Outcomes Management Service Information Model	Ed. Tech	(Smythe, 2011)
NSDLA	National Science Digital Library	Logging?	(“NSDL comm_anno,” 2011)

16 This work was subsequently developed into a full European Standard, EN 15981:2011, that should be consulted as the definitive source; the CWA is referred to because it is available free of charge.

	Annotation		
NSDLP	National Science Digital Library Paradata	Logging?	("NSDL comm_para," n.d.)
OBI	Mozilla Open Badge initiative	Ed. Tech	(Mozilla, 2013)
QTI	IMS Question and Test Interoperability Results Reporting	Ed. Tech	(Kraan et al., 2012)
SIOC	Semantically Interlinked Online Communities	General web	(Bojārs & Breslin, 2010)
XAPI	Experience API (formerly known as Tin Can API)	Ed. Tech	("Experience API v0.95," 2012)
XES	Extensible Event Stream	Logging	(Günther & Verbeek, 2012)

Table 3: List of standards

Table 4 summarises the coverage of the standards. Logging standards are signified by "L", general web standards are signified by "G", and educational technology standards are signified by "E". The abbreviations are as used in Table 3 and in the previous sections.

	A. The Person	**B. Resources**	**C. Services and Tools**	**D. Learning Activities**	**E. Objectives & Assessment**
1.	G: FOAF *	*		E: NSDLA, Leap2A	L: DataShop E: InLOC
2.	L: XES, DataShop G: FOAF, SIOC, AS E: XAPI	L: XES, DataShop, CAM G: FOAF, SIOC, AS E: eLAT, NSDLP, XAPI	L: XES, DataShop, CAM G: SIOC, AS E: eLAT, XAPI	G: Atom, SIOC, AS E: Leap2A AIF	L: DataShop E: eLAT, EuroLMAI, LIS, OBI, QTI, XAPI

Table 4: Indicative coverage of standards (* - standards for common person information and learning resource meta-data. have been omitted from the survey)

Conclusions

The aim of this piece of work was to help establish a baseline for learning analytics interoperability, in terms of: a) the prior work and interest in the topic but also, b) the range of candidate standards for exploration in a LAI setting.

It is clear from the literature that many of the issues have been identified and some moves have been made to improve the status quo. In general, it would be expected that interoperability would not figure much in relatively new field of activity but learning analytics has the movement of data so close to its core that the motivation for tackling the problems is clearly greater. Given this, it is reasonable to assert that LAI should receive more attention than it has so far, especially if we are to move from research to pilots to so to large scale adoption.

There are both existing standards that could be trialled or taken as a starting point, and some gaps in provision. Since the data must come from learning systems, new work should start from an analysis of what exists and strike a balance between the idiosyncrasies of single applications and the complexity of a universal model. Opportunities should be taken, both by workers in the learning analytics and educational data mining field and by workers in the various existing standards and pre-standards initiatives, to collaborate in order that the most rapid progress towards LAI in practice can be made.

Capturing learner activity at relatively fine levels of granularity is clearly a common theme and there are already a large number of standards for capturing learner activity or attention (see Table 4). It seems likely that there are opportunities to consolidate development work and move towards the adoption of a core with minimal defined semantics overlain by more semantically-specific vocabularies or extensions. A new work item proposal to the CEN Workshop Learning Technologies that was accepted by the European Commission in April 2013 to create an information model for capturing social and context data in Technology Enhanced Learning (Klerkx, 2011) may be a useful means of achieving this through an open consensus process,

In spite of the fact that granular usage data is the low-hanging fruit for LAI, there clearly are standards that can support a more semantically-rich level of analysis on the learners' argumentation and written word.

References

Abbas, S., & Sawamura, H. (2009). Developing an argument learning environment using agent-based its (ales). *Proceedings of The Second International Conference on Educational Data Mining*, 200–209. Retrieved from http://pdf.aminer.org/000/266/271/evaluating_a_learning_environment_for_case_based_argumentation_skills.pdf

Barneveld, A. Van, Arnold, K. E., & Campbell, J. P. (2012). Analytics in Higher Education : Establishing a Common Language. *Business*, (January), 1–11.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: an issue brief. *Washington, DC: Office of Educational Technology, ...* Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Enhancing+Teaching+and+Learning+Through+Educational+Data+Mining+and+Learning+Analytics:+An+Issue+Brief#0

Bojārs, U., & Breslin, J. G. (2010). SIOC Core Ontology Specification. Retrieved April 23, 2013, from http://sioc-project.org/ontology

Brickley, D., & Miller, L. (2010). FOAF Vocabulary Specification 0.98. Retrieved from http://xmlns.com/foaf/spec/20100809.html

CAM Schema. (n.d.). Retrieved from https://sites.google.com/site/camschema/home

Castro, F., & Alonso, M. (n.d.). Learning Objects and Ontologies to Perform Educational Data Mining. Retrieved from http://world-comp.org/p2011/FEC8576.pdf

CEN. (2010). WORKSHOP AGREEMENT European Learner Mobility Achievement Information (EuroLMAI). European Committee for Standardisation. Retrieved from http://www.cen-wslt.din.de/sixcms_upload/media/3378/CWA16132.pdf

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318. doi:10.1504/IJTEL.2012.051815

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., & Willmott, S. (2006). Towards an argument interchange format. *Knowledge Engineering Review*, *21*(4), 293–316. doi:10.1017/S000000000000000

Choi, S.-Y., & Whinston, A. B. (2000). Benefits and requirements for interoperability in the electronic marketplace. *Technology in Society*, *22*(1), 33–44. doi:10.1016/S0160-791X(99)00034-2

Cooper, A. (2012). What is "Analytics"? Definition and Essential Characteristics. *CETIS Analytics Series*, *1*(5). Retrieved from http://publications.cetis.ac.uk/2012/521

Desmarais, M. C., & Baker, R. S. J. D. (2011). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*(1-2), 9–38. doi:10.1007/s11257-011-9106-8

Dietze, S. (2012). Linked Data as facilitator for TEL recommender systems in research & practice. *Proceedings of the 2nd Workshop on Recommender Systems for Technology Enhanced Learning*, 7–10. Retrieved from http://ceur-ws.org/Vol-896/keynote.pdf

Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H. Q., Giordano, D., Marenzi, I., & Numes, B. P. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program: electronic library and information systems*, *47*(1), 60–91. doi:10.1108/00330331211296312

Drachsler, H., Dietze, S., Greller, W., D'Aquin, M., Jovanovic, J., Pardo, A., Reinhardt, W., et al. (2012). 1st International Workshop on Learning Analytics and Linked Data. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12* (p. 9). New York, New York, USA: ACM Press. doi:10.1145/2330601.2330607

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. Retrieved from http://dl.acm.org/ft_gateway.cfm?id=2090118&ftid=1075994&dwn=1

Dyckhoff, A. L., Zielke, D., Chatti, M. A., & Schroeder, U. (2011). eLAT: An Exploratory Learning Analytics Tool for Reflection and Iterative Improvement of Technology Enhanced Learning. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *4th International Conference on Educational Data Mining* (pp. 355–356). International Educational Data Mining Society.

Dyckhoff, A., Zielke, D., Bültmann, M., Chatti, M. A., & Shroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *Journal of Educational Technology and Society*, *15*, 58–76. Retrieved from http://www.ifets.info/journals/15_3/5.pdf

Experience API v0.95. (2012). Advanced Distributed Learning Initiative. Retrieved from http://tincanapi.wikispaces.com/file/view/Experience API Release v0.95.pdf/384712964/Experience API Release v0.95.pdf

Fazeli, S., Drachsler, H., Brouns, F., & Sloep, P. (2012). A Trust-based Social Recommender for Teachers. *Proceedings of the 2nd Workshop on Recommender Systems for Technology Enhanced Learning*, 49–60. Retrieved from http://ceur-ws.org/Vol-896/paper4.pdf

Gasser, U., & Palfrey, J. (2007). When and How ICT Interoperability Drives Innovation When and How ICT Interoperability Drives Innovation. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1033226

Geraci, A., Katki, F., McMonegal, L., Meyer, B., & Porteous, H. (1991). *IEEE Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries*. (F. Katki, L. McMonegal, B. Meyer, J. Lane, P. Wilson, J. Radatz, M. Yee, et al., Eds.)*IEEE Std 610* (p. 1). Inst of Elect & Electronic. doi:10.1109/IEEESTD.1991.106963

Grant, S. (2011). The Leap2A Specification for e-Portfolio Portability and Interoperability. Retrieved from http://www.leapspecs.org/2A/core-specification

Grant, S. (2013). InLOC Information Model for Learning Outcomes and Competences. Retrieved from http://wiki.teria.no/display/inloc/Information+Model

Günther, C. W., & Verbeek, E. (2012). XES Standard Definition v1.4. *Fluxicon Process Laboratories*. Technische Universiteit Eindhoven. Retrieved from http://www.xes-standard.org/_media/xes/xesstandarddefinition-1.4.pdf

InBloom. (2013). inBloom Technology Application Developer Documentation v1.2. Retrieved from https://inbloom.org/sites/default/files/docs-developer/index.html

Jovanovic, M., & Vukicevic, M. (2012). Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. *International Journal of Computational Intelligence Systems*, (April 2013), 37–41. Retrieved from http://www.tandfonline.com/doi/full/10.1080/18756891.2012.696923

Klasnja-Milicevic, A., Vesin, B., Ivanovic, M., & Budimac, Z. (2011). Integration of recommendations and adaptive hypermedia into java tutoring system. *Computer Science and Information Systems*, *8*(1), 211–224. doi:10.2298/CSIS090608021K

Klerkx, J. et al. (2011). CEN WS-LT “Workshop on Learning Technologies ”Proposal for : An Information Model for capturing Social and Context Data in Technology Enhanced Learning. Retrieved from http://wiki.teria.no/download/attachments/15401021/CEN_WSLT_Proposal_Social_Context_data_V1.4_clean_without_budget.pdf

Koedinger, K. R., Baker, R. S. J. d., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2011). A Data Repository for the EDM Community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 43–55). CRC Press.

Kraan, W., Lay, S., & Gorissen, P. (2012). IMS Question & Test Interoperability Results Reporting v2.1. Retrieved April 23, 2013, from http://www.imsglobal.org/question/qtiv2p1/imsqti_resultv2p1.html

Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. *Proceedings of the 3rd International Conference on Educational Data Mining*, 131–140. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.174.6315&rep=rep1&type=pdf

Learning Registry Paradata Specification V1.0 . (n.d.). Retrieved from https://docs.google.com/document/d/1IrOYXd3S0FUwNozaEG5tM7Ki4_AZPrBn-pbyVUz-Bh0/edit#heading=h.lqc5y1f59stc

Medvedeva, O., Chavan, G., & Crowley, R. (2005). A data collection framework for capturing ITS data based on an agent communication standard. *Proceedings of the 20th Annual Meeting of the American Association for Artificial Intelligence.*, 23–30. Retrieved from http://www.aaai.org/Papers/Workshops/2005/WS-05-02/WS05-02-004.pdf

Mozilla. (2013). Badge Specification (Assertions). Retrieved from https://github.com/mozilla/openbadges/wiki/Assertions

Muehlenbrock, M. (2005). Automatic action analysis in an interactive learning environment. *The 12th international conference on artificial intelligence in education*, 73–80. Retrieved from http://www.muehlenbrock.de/publications/Muehlenbrock-WS AIED 2005.pdf

Niemann, K., Scheffel, M., & Wolpers, M. (2012). An Overview of Usage Data Formats for Recommendations in TEL. *Workshop on Recommender Systems for TEL*, 95–100. Retrieved from http://ceur-ws.org/Vol-896/paper9.pdf

Nottingham, M., & Sayre, R. (2005). The Atom Syndication Format (RFC4287). Retrieved from http://tools.ietf.org/html/rfc4287

NSDL comm_anno. (2011). Retrieved from https://wiki.ucar.edu/display/nsdldocs/comm_anno

NSDL comm_para. (n.d.). Retrieved from https://wiki.ucar.edu/display/nsdldocs/comm_para

NSDL's Technical Schema for Paradata Exchange. (2011). Retrieved from http://nsdlnetwork.org/stemexchange/paradata/schema

Palermo, J., Marr, D., Oriel, J., Arthur, J., & Johnston, D. (2012). Tracking student success: Using an action learning approach to better understand the how, what, where and why. *Journal of Institutional Research*, *17*(1), 39–50.

Phankokkruad, M. (2012). Association rules for data mining in item classification algorithm: Web service approach. *2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP)*, 463–468. doi:10.1109/DICTAP.2012.6215408

PSLC. (2013). Guide to the Tutor Message format (v4). Retrieved from http://pslcdatashop.web.cmu.edu/dtd/guide/tutor_message_dtd_guide_v4.pdf

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology and Society*, *15*, 3–26. Retrieved from http://oro.open.ac.uk/34092/

Siemens, G., & Baker, R. S. J. (2012). Learning Analytics and Educational Data Mining : Towards Communication and Collaboration. *LAK12*.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., & Ferguson, R. (2011). *Open Learning Analytics : an integrated & modularized platform Proposal to design , implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*.

Smythe, C. (2011). IMS Outcomes Management Service Information Model v1.0. Retrieved from http://www.imsglobal.org/lis/lisv2p0/OMSv1p0InfoModelv2p0pd.html

Snell, J. M., Atkins, M., Norris, W., Messina, C., Wilkinson, M., & Dolin, R. (2011). Activity Streams Specifications 1.0. Retrieved from http://activitystrea.ms/specs/

Stutt, T. (2011). Two More Takes on the LMS market. *Tim's Learning Analytics blog*. Retrieved from http://blogs.ischool.berkeley.edu/tim/2011/11/10/two-more-takes-on-the-lms-market/

Taibi, D. et al. (2012). LAK Dataset. Society for Learning Analytics Research. Retrieved from http://www.solaresearch.org/resources/lak-dataset/

Tolley, R. (2011). Learning Analytics. *eFolio in the UK and Europe (Blog)*. Retrieved from http://efoliointheuk.blogspot.co.uk/2011/02/learning-analytics.html

Trcka, N., Pechenizkiy, M., & Van der Aalst, W. (2011). Process Mining from Educational Data. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 123–142). CRC Press.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Datasetdriven research for improving recommender systems for learning. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 44–53). New York, New York, USA: ACM Press. doi:10.1145/2090116.2090122

Weston, C. (2012). Learning analytics for better learning content. *Ed Tech Now Blog*. Retrieved from http://edtechnow.net/2012/03/31/learning-analytics-for-better-learning-content/