Learning Analytics Community Exchange

Learning Analytics Interoperability – The Big Picture in Brief

an introductory briefing

By: Adam Cooper, Cetis, University of Bolton, UK

Published: 26 March 2014

Keywords: learning analytics, interoperability, benefits

Learning Analytics is now moving from being a research interest to a wider community who seek to apply it in practice. As this happens, the challenge of efficiently and reliably moving data between systems becomes of vital practical importance. System interoperability reduces this challenge. This introductory briefing, which is aimed at non-technical readers who are concerned with developing plans for sustainable practical learning analytics, describes some of the motivations for better interoperability and outlines the range of situations in which standards or other technical specifications can help to realise these benefits.

Contents

Introduction

Learning Analytics is now moving from being a research interest to a wider community who seek to apply it in practice. As this happens, the challenge of efficiently and reliably moving data between systems becomes of vital practical importance.

Whereas traditional forms of analytical processing rely on existing management data, such as student demographics, grades, and recruitment figures, more recent approaches to analytics rely on data that has greater variety and arises from traces left as people use IT systems. This is a central concern for learning analytics, where the data arises from normal use of multiple pieces of software designed for accessing learning resources, social interaction, content creation, etc. In many cases, therefore, practical learning analytics requires that data moves from operational to analytical systems and be put to a different use than originally intended. For example, the data structures in a VLE or LMS are likely to have been designed not for analytics, but to realise teaching and learning use cases - e.g. for accessing video content, participation in forums – in a way is technically scalable and maintainable. When statistical processing or data mining is undertaken, for example to support analysis of learner engagement, data has to be re-interpreted. This situation is further amplified by the necessity of combining data from various sources, or maybe to use cloud-computing based data mining engines, to build, test, and apply useful statistical and predictive models.

This introductory briefing, which is aimed at non-technical readers who are concerned with developing plans for sustainable practical learning analytics, describes some of the motivations for better interoperability and outlines the scope for interoperability between learning analytics systems.

What is Interoperability?

The Institute of Electrical and Electronics Engineers defines interoperability to be:

> *"the ability of two or more systems or components to exchange information and to use the information that has been exchanged."*

A broad interpretation of "systems" that includes people and the activities they undertake using these digital technologies captures the true essence of interoperability as a means to achieve human aims and objectives. When we have interoperability, things "just work". When we do not, digital technologies are frustrating or arduous to use: we see errors, warning messages, and crashes; we are faced with data we do not understand, that is corrupted, or missing; we have to know tricks to "work-around" the technology; we have to use certain software for a given task when we would prefer an alternative.

The most difficult challenges with achieving interoperability are typically found in establishing common meanings to the data. Sometimes this is a matter of technical precision, but culture – regional, sector-specific, and institutional – and habitual practices also affect meaning. Meaning is found in what people do and think and it is often not formally expressed but instead captured in the way software is written or used. An important consequence of the difficulty posed by common meaning is that, except in very constrained settings, useful levels of interoperability are often only found after periods of investigation, negotiation, experience, and refinement.

What are the potential benefits?

In addition to our experience, as consumers, of IT "just working", interoperability affords a number of other kinds of benefit with much wider impact in an institutional or business setting. In any given real-world setting, we observe that one or two of these benefit types tend to be more prominent than the others over the lifetime of a system or data. Some of them are general benefits, for example arising from enabling modular or service-oriented approaches, while others have particular relevance to learning analytics. Analytics interoperability benefits include:

Efficiency and timeliness are the most obvious when considering business use of IT and IT systems operation. Interoperability means that systems work without the need for a person to intervene to re-enter, re-format, or transform data. Information flows become more efficient and the delivery of information to the point of use enables more timely action. Data quality may also be improved in a suitably designed and managed system by reducing the risk of mistakes and unpredictable degradation of data quality.

Independence from consequential disruption: resilience. Although it may be possible to access data directly from the databases that underpin most software, the supplier is unlikely to guarantee that there will be no changes as they release new versions. An interface that is designed to be interoperable – conforming to specified data structures and access methods and de-coupled from the operational database scheme - should be a much more dependable data source.

Adaptability, both of IT architecture and educational practices, can arise when interoperability is combined with a modular approach; it is faster, cheaper, and less disruptive to change things as needs change. This is particularly relevant for applications, such as learning analytics, where the methods are not yet mature, the institutions that are adopting learning analytics are undergoing rapid change, and the people who are the ultimate users do not yet fully understand the potential. The antithesis of adaptability is lock-in to a black-box integrated system, where the cost and disruption of change inhibits transition to a more fit-for-purpose system. In the absence of change, teachers and learners must endure unsatisfactory software and the potential of learning analytics remain unrealised.

Innovation and market growth is closely related to adaptability, as its supply-side counterpart. Interoperability combined with modularity makes it easier to build IT systems that are better matched to local culture without needing to create and maintain numerous whole systems. The market becomes more attractive to new entrants, hence innovation is increased, by reducing the cost of development and because consumer organisations are better able to undertake incremental change. Interoperability enables the development and viability of a category of learning analytics tools that can work across diverse local system choices.

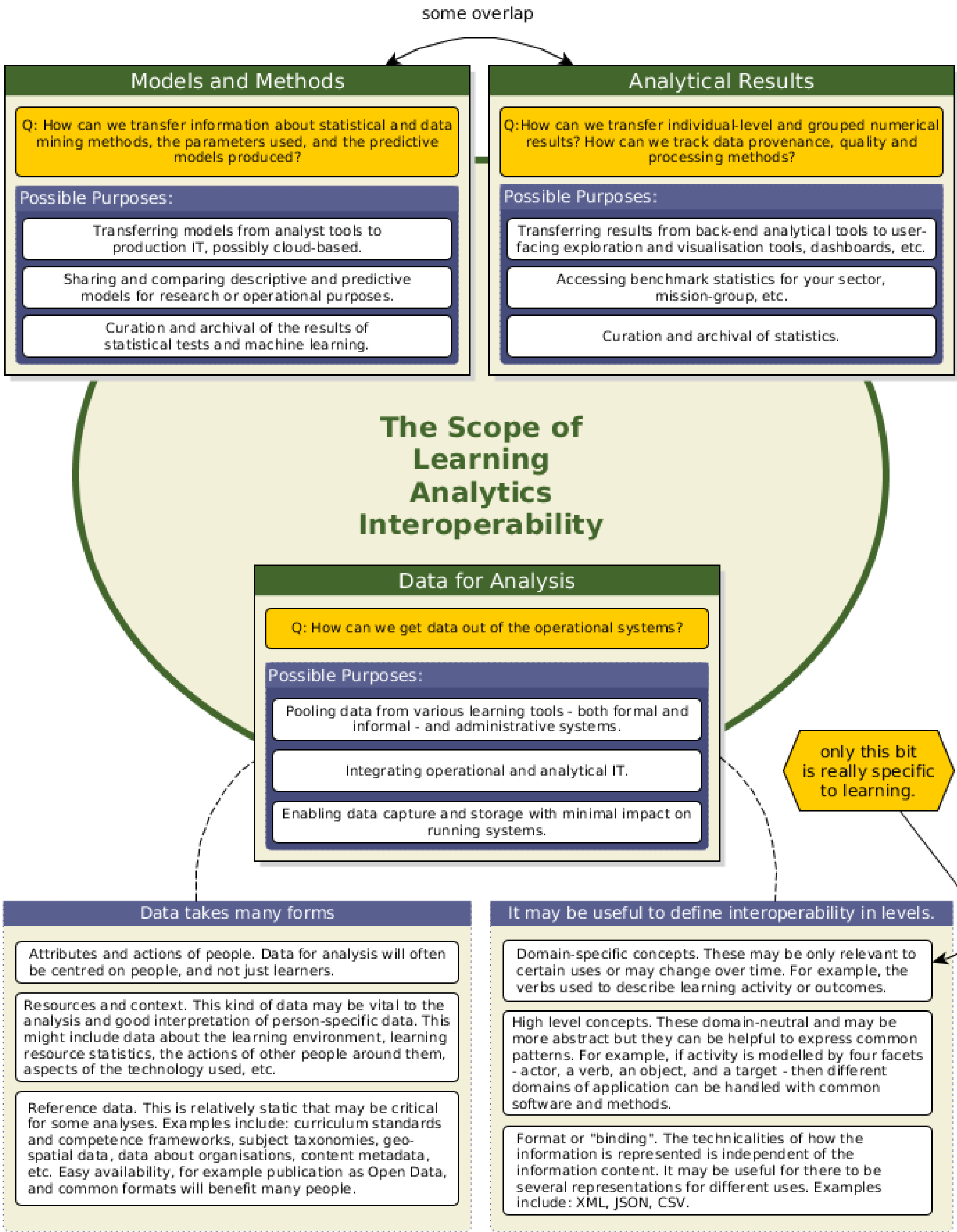
Durability of data has generally been the concern of archive managers and neglected by others. Historical data is valuable for learning analytics but simply preserving it as it is created in operational systems is often unsatisfactory because structures and formats change over time, and the changes are rarely properly documented. This is likely to apply to the raw data generated by operational systems, and also to the results of analysis whether these live in spreadsheets or more structured analytical systems. Archiving data using a stable, documented, and well-defined information model and technical representation improves the chances that historical data and analytical results will be decipherable in the future.

Aggregation of data from multiple sources that may be both internal and external to an organisation is a key driver of value in analytics. Interoperability eases aggregation most immediately through the easing of access to the data but the real advantages arise from merging or joining datasets, which implies the use of common definitions across several IT systems, even if they do not interact in normal operations. Data joining might be supported by a common set of definitions around course structure, combined with a unified identification scheme. Merging of data from blogs, forums, twitter, wikis, etc to analyse patterns of contribution and consumption would require a common way of expressing social media contributions that is independent of the kind of platform.

Sharing data, especially when there are multiple parties involved, is more tractable if a common language that does not favour any single organisation can be agreed upon. For many organisations, the previously-mentioned benefits are likely to be far more important than the possibility of sharing, especially of original data; legal and ethical concerns are likely to be serious obstacles except maybe between an educational technology platform provider and their clients, or within the regulated confines of educational research. The sharing of derived or aggregated data is, however, more tractable, and could be useful in benchmarking exercises, or the collective development of predictive models that require the largest possible sample size.

The Scope of Learning Analytics Interoperability

The following diagram illustrates some of the different roles for learning analytics interoperability, broken down into three categories. Rather than focussing only on data about learners, it considers a bigger picture that includes interoperability support for the analytical process in addition to the spectrum of data that might be useful. The implication of the big picture is that opportunity to realise benefits may lie in many places, not that realising benefits requires implementation of the whole.



We will now look at some examples where interoperability standards or other specifications exist to illustrate the above diagram. The intention here is not to provide a systematic or value-based summary but to show the variety of approaches to interoperability. Web links to most of the technologies mentioned are given in the section "Further Reading – Technical Details" towards the end of this document. Particular attention is given to interoperability that is based on open, rather than proprietary, standards and specifications. This is because open standards[1] maximise the benefits mentioned previously.

Interoperability of **models and methods** is the aim of PMML, the Predictive Model Markup Language, a mature XML-based specification from the Data Mining Group. Although its emphasis is on predictive methods such as decision trees and logistic regression, it can also be used to convey the results of more common statistical tests. It is supported by a range of software, both open source (e.g. R and Knime, which offer programming and visual environments respectively) and closed source, as well as providers of cloud computing data mining services (e.g. Zementis).

While PMML does deal with **analytical results**, a number of international agencies concerned with statistics collection and publication - including the European Central Bank, Eurostat, and UNESCO - have developed SDMX, the Statistical Data and Metadata eXchange standard, and a number of freely-available software tools. SDMX is typically used for large-scale demographic, social, economic, and environmental statistics and is capable of capturing a lot of information about the data (i.e. metadata) such as the origin, reporting periods, etc. SDMX is an ISO standard and the W3C Data Cube Vocabulary is a closely-related specification. Both SDMX and Data Cube are very complicated, and are not likely to be useful in their entirety for learning analytics applications any time soon. On the other hand, they do provide a resource for the definition of simplified solutions to similar problems. Basing a simple model on SDMX concepts rather than inventing a new approach allows for future growth in scope and for serendipitous interoperability as what were initially distinct cases grow and merge.

A rather different strategy for applying interoperability to the transfer of numerical results is to move the problem from the numbers to their presentation. Widget and portlet approaches allow visualisations and other elements to be composed into a dashboard, provided as cross-platform "apps", or embedded in teaching and learning software. This is the realm of web standards such as OAuth, HTML5, and the W3C Packaged Web Apps (Widgets) specification. As such, they are not specific to analytics interoperability but they may be a practical, and possibly interim, means to improving on closed-environment analytics software while avoiding the complexity of SDMX, for example. These techniques are well understood in the educational technology domain. For example, IMS LTI (Learning Tools Interoperability) builds on top of web standards and represents an easy, and quite widely-supported, way to include context sensitive analytics tools into LMS/VLE systems.

When it comes to **data for analysis**, there are numerous existing specifications and standards that contribute elements of interoperability. Although learning analytics will often rely on data that is

[1] The term "open standards" is used with various meanings, ranging from access to the specification document and rights to use the technology on reasonable and non-discriminatory terms through to a requirement that the standard should have been created in a transparent process open to all interested parties. The Open Stand Principles - *http://open-stand.org/principles/* - provide a good benchmark since they are supported by the most important IT standards organisations, judged by impact.

specific to the education and training domain, for example capturing the idea of person roles, it will also often be possible to use generic data such as the web page access and interation events used in web analytics. The scale of use of these generic techniques makes them candidates for adoption, or adaption, for learning analytics because of the good supply of tools and expertise.

In the world of web analytics, Google Analytics is dominant. It is, of course, an entirely proprietary system but, since it relies on the interoperability of JavaScript in particular, its users realise a number of the benefits outlined in the previous section. Although few, if any, educational establishments would consider tracking learner-level activity with such a service because of a duty of care with respect to personal data, it may be useful for analysis focussed on web sites and content use. The Open Source Software alternative Piwik gives its users more control over the captured data but it is currently less fine-grained in what it can track and it is still essentially a proprietary approach to data collection and access.

Web analytics tools do not, however, cover all of the information that it would be useful to exchange: about people, their activities, the various contexts in which these take place, learning resources, and the relationships between them all. In addition, whereas specifications such as SDMX and PMML have arisen from many years of application, the same is not true for data for learning analytics. Even those cases where an interoperability specification has been used for several years, the use has not usually been for analytics. There is still, in early 2014, a need to experiment, to gather evidence, and to undertake selective refinement and work on new prototype standards driven by analytical need.

Two existing interoperability specifications have been chosen to illustrate different strategies within the learning, education, and training domain; they take different approaches to the split between domain-specific and high-level concepts shown on the diagram. Both specifications relate closely to the centrally-important data about learner activity and both have been conceived with learning analytics use cases. The two specifications are the eXperience API and IMS QTI.

The eXperience API (usually abbreviated xAPI and also called the Tin Can API by its creators) has had a wave of interest and implementation since its conception in 2010. The xAPI is based on the idea of tracking activity and it is based on Activity Streams[2], which was developed to provide a better way of expressing social media activity than existed before. It provides a framework for making statements of the kind “someone does an action to/with something”, for example “Geraldine posted a photo to her album”. The xAPI has been used for social learning platforms and increasingly it is exciting interest as a means of capturing data for learning analytics, although this is not yet well established in practice.

The xAPI only defines the statement pattern and how statements are stored and retrieved; it does not specify, for example, what all the verbs are that might be used in practice. This is both an advantage and a difficulty. The advantage lies in being open to a wide range of uses and, since learning analytics is relatively new outside the research context, we should expect the range of uses to increase considerably over the next few years.The disadvantage is that adopters need to define

[2] The IMS Caliper project, which is currently (March 2014) in the chartering stage adopts a similar approach to tracking activity with a view to using the collected data for analytics.
http://www.imsglobal.org/caliper/index.html

these missing parts. This applies even within a single institutional adoption since data from different systems will generally be combined, but for wider visions of interoperability it indicates that communities of shared practice and educational culture must get together to reach consensus definitions.

IMS QTI (Question and Test Interoperability) has a more-than ten year history of development and implementation. Most of the attention has been on QTI for exchange and rendering of assessments and questions, but the creators of QTI included a number of established high-stakes assessment organisations that have been undertaking well-developed statistical analyses for many years. These analyses are usually referred to as "psychometrics" rather than "learning analytics" but maybe the current interest in the topic will spur wider interest in these mature techniques, and bring more attention to the results reporting capabilities of IMS QTI.

QTI is a lot more complete in what it defines than xAPI is. This is for good reason; the nature of questions, tests, and the results of them is sufficiently complicated that more ideas must be defined before a specification becomes useful. This makes it less flexible but affords greater clarity about how QTI works. So, while the scores that learners might get on sections of an assessment, or even the order in which they attempt questions could be sensibly captured with xAPI, possibly using terms defined in QTI, a more detailed analysis of interactions is likely to require access to data about the question item structure that would best be expressed in the purpose-build form of IMS QTI.

Conclusion

A lot of attention given to interoperability in the context of learning analytics has, so far, been given to the capture of learner activity. Among the drivers for this are wider trends in analytics and the diversity of IT used for education and training. Some of the best-known examples of business analytics make particular use of the traces captured during consumer interactions with web sites, and with social media. These motivate interest in the almost-magical acquisition of knowledge from apparently inconsequential data. The diversity in IT presents an old problem; it is difficult to analyse data that is spread about. Hence, people are attracted to using a single method of capturing data that originates in many places for storage in a single location. Frequently, progress may be made by adopting common approaches to generic concepts but useful learning analytics will usually require common approaches to domain-specific concepts, which require definition by communities of practice.

This short document has illustrated a wider range of places in which the benefits of interoperability may be found beyond the capture of learner activity. In order to explore these matters more deeply, the Learning Analytics Community Exchange project will produce a range white papers and briefings. These will communicate a systematic analysis of what is currently available and capture practical experience. The papers will be complemented by a number of webinars in which implementers – in educational establishments and in the software industry - will share their experiences, and innovators will present new developments.

The LACE project will also be working with this community of implementers, researchers, and interoperability experts to help build consensus on common data definitions, and identify important missing pieces in the learning analytics interoperability jigsaw.

Further Reading – Technical Details

Activity Streams - http://activitystrea.ms/

Google Analytics - http://www.google.com/analytics/

LTI – IMS Learning Tools Interoperability - http://www.imsglobal.org/toolsinteroperability2.cfm

OAuth - http://oauth.net/

Piwik – http://piwik.org/

PMML – Predictive Model Markup Language - http://www.dmg.org/v4-1/Interoperability.html

QTI – IMS Question and Test Interoperability -
http://www.imsglobal.org/question/qtiv2p1/imsqti_resultv2p1.html

SDMX - Statistical Data and Metadata eXchange - http://sdmx.org/

W3C Data Cube - http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/

W3C Widgets - http://www.w3.org/TR/widgets/

xAPI – The eXperience API (also known as Tin Can API) - http://www.adlnet.gov/tla/experience-api/

About ...

Acknowledgements

The author would like to thank Tore Hoel (Oslo and Akershus University College) and Wilbert Kraan (Cetis, University of Bolton) for reviewing drafts of this document and providing suggestions that have improved the quality of the final version. The remaining weaknesses are wholly my own.

This document was produced with funding from the European Commission Seventh Framework Programme as part of the LACE Project, grant number 619424.

About the Author

Adam is co-director of the Centre for Educational Technology and Interoperability Standards, based at the University of Bolton. He has participated in many of the groups involved in educational technology standards since the year 2000 and is a member of the UK Government Open Standards Board and Information Standards Board for Education, Skills and Children's Services. He has written a number of briefings and white papers on learning analytics, interoperability, and standards development - at *http://publications.cetis.ac.uk/author/adam* - and blogged on the same topics at *http://blogs.cetis.ac.uk/adam/*.

About this document

The persistent URL for this document is: *http://www.laceproject.eu/publications/briefing-01.pdf*

About LACE

The LACE project brings together existing key European players in the field of learning analytics & educational data mining who are committed to build communities of practice and share emerging best practice in order to make progress towards four objectives.

> *Objective 1 – Promote knowledge creation and exchange*
> *Objective 2 – Increase the evidence base*
> *Objective 3 – Contribute to the definition of future directions*
> *Objective 4 – Build consensus on interoperability and data sharing*

http://www.laceproject.eu *@laceproject*