

Infrastrukturen für Learning Analytics in medien- und projektübergreifenden Lernszenarien

Birte Heinemann¹, Matthias Ehlenz¹, Jens Doveren¹ und Ulrik Schroeder¹

Abstract: Lernen kann an vielen Orten stattfinden. Remote Teaching und digitale Lehre können durch äußere Umstände forciert oder gezielt eingesetzt werden, um das Lernen zu verbessern. Die Bestimmung der passenden Lehrmethoden und Technologien in Abstimmung mit den pädagogischen Zielen ist eine Herausforderung. Der Forschungsbereich Learning Analytics stellt Methoden und Werkzeuge bereit, dieses Passungsproblem zu überwinden und unterstützt Lehrende bei der Wahl adäquater Mittel. Dieser Praxisbeitrag stellt die Erfahrungen bei der Entwicklung von möglichst wiederverwertbaren Learning Analytics-Infrastrukturen für verschiedene Lernorte und Lernmethoden vor.

Keywords: MMLA, Learning Analytics.

1 Motivation

Bei der Digitalisierung von Lehre und Lernen können Daten entstehen (oder gezielt erhoben werden), die genutzt werden, um die Prozesse menschlichen Lernens besser zu verstehen und Lehrangebote durch neue Methoden und eine gezielte Optimierung stetig zu verbessern. Häufig ist in diesem Kontext ein Vergleich des klassischen Lernszenarios mit einer digital gestützten Variante oder verschiedener Methoden untereinander wünschenswert. Ein Kernthema der Forschung ist adäquates Feedback für Lerner zu generieren, besonders wenn die Digitalisierung zu einer Reduktion des Kontaktes mit den Lehrkräften führt, die vorher ein personalisiertes Feedback geben konnten. Der Wunsch nach Evaluierung und Optimierung hat zur Professionalisierung der Forschung und der Entstehung des Forschungsbereichs rund um Learning Analytics geführt. Die Auswirkungen sind auch auf technischer Ebene sichtbar: Datenformate, Frameworks und Standardisierungen wurden und werden entwickelt und Prozesse für ein nachhaltiges Forschungsdatenmanagement, welches zum Beispiel den FAIR Prinzipien folgt werden vorangetrieben. Ein weiteres aktuelles Thema ist die Vernetzung und Synchronisierung verschiedener Technologien, aber auch die Erweiterung um weitere Sensorik und die Erfassung verschiedenster Datenströme. Ziel dieses Beitrags ist die Weitergabe gesammelter Erfahrungen aus verschiedenen Projekten und Ansätzen zur Entwicklung nachhaltiger Infrastrukturen im Bereich Learning Analytics.

¹ RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen,
{heinemann|ehlenz|doveren|schroeder}@informatik.rwth-aachen.de

2 Synergieeffekte projektübergreifender Infrastrukturen

Ein wichtiger Punkt von Forschungsprojekten der Bildungstechnologien ist die Evaluation der Entwicklungen und hierbei die systematische Erhebung von Forschungsdaten über Lernaktivitäten. Die Heterogenität und die Vielfältigkeit des Lernens führen zu einer zeitintensiven Suche zur Identifikation passender Werkzeuge und Modelle. Dieser Prozess kann durch die Entwicklung systematischer, nachhaltige und bedarfsgerechter Infrastrukturen für Learning Analytics unterstützt werden. Gleichzeitig lassen sich hierdurch langfristig projektübergreifende Untersuchungen zu und die Veröffentlichung von Forschungsdaten aktiv unterstützen.

2.1 Bildungstechnologien als vielfältige Projektlandschaft

Die verschiedenen Projekte unserer Institution dienen als Beispiel für die Vielfältigkeit des Forschungsbereiches Bildungstechnologien und geben einen Eindruck der Herausforderung, die projektübergreifende Infrastrukturen darstellen. Im Projekt Open Digital Lab for You² geht es um die Vernetzung von (Lehr-)Laboren in den Ingenieurwissenschaften. Hierbei sollen im internationalen Austausch Laborkonzepte sowie teure Hardware zugänglich gemacht und zwischen verschiedenen Hochschulen gemeinschaftlich genutzt werden können. Dabei geht es sowohl um digitalisierte Lösungen, z.B. in Virtual Reality, hybride Lösungen (durch Remote-Zugriffe auf Laborgerätschaften) als auch die Erfassung von Lernprozessen im realen Laboralltag. Hierbei soll das Laborbasierte Lernen als Prozess in diesen drei Ausprägungen messbar gemacht und auf Vor- und Nachteile der jeweiligen Modalitäten untersucht werden.

In anderen Projekten geht es um die Beforschung von Kollaborationsprozessen beim Lehren und Lernen im Klassenzimmer der Zukunft. Im Mittelpunkt steht hierbei (technisch) interaktive Multi-Touch Tische, das als moderierende Komponente Gruppenarbeitsprozesse fördern und somit das gemeinschaftliche Lernen von Angesicht zu Angesicht ins Zentrum des Lernprozesses rücken soll. Da das technische System hier (im Idealfall) nur den Mediator stellt, findet der eigentliche Lernprozess in der Mensch-Mensch-Kommunikation statt. Zur holistischen Abbildung dieses Lernprozesses reichen also die Interaktionsdaten an der Mensch-Computer-Schnittstelle nicht mehr aus und machen die Integration multi-modaler Datenerfassung erforderlich.

In den Bereich Multi-Modal Learning Analytics fallen darüber hinaus auch weitere, mehr oder weniger neue Medien, so liefern VR-Brillen Datenströme über Lagesensoren, ggf. integrierte Eyetracker und innovative Controller. Smartphones und Tablets gewinnen zunehmend an Bedeutung als Lernmedium. Mobile Learning bietet ebenfalls die Chance Einblicke in (physische) Lernumgebungen und Situationen zu erhalten.

² <https://digilab4u.com/>

Aber auch jenseits von neuer Technik, virtuellen Welten und vernetzten Laboren finden Entwicklungen statt, die gerade während der aktuellen Situation an Stellenwert gewonnen haben. Grundsätzlich gehören traditionelle Lern-Management-Systeme wie Moodle ebenso zum „Kerngeschäft“ im Bereich Bildungstechnologien wie MOOCs und videobasiertes Lernen. Während diese inzwischen in der Hochschullandschaft fest im Alltag etabliert sind, haben sie in Zeiten der COVID-19-Pandemie auch, teilweise sehr kurzfristig, ihren Weg in die Schule gefunden. Neben der Frage, ob bekannte Erkenntnisse aus dem universitären Kontext in dieses doch grundlegend andere System übertragen werden können, stellen sich hier besonders komplexe Fragestellungen, wie zum Beispiel mit Learning Analytics Lehrkräfte in ihrer Aufgabe aktiv unterstützt werden können aber gleichzeitig die Integrität des besonders geschützten Raums Schule auch im Distanzunterricht hinsichtlich des Datenschutzes gewährt werden kann.

2.2 Entwicklung einer gemeinsamen Basis

Die beispielhaft aufgeführten Projekte zeigen einen Ausschnitt, der nur einen Teil der Bildungstechnologien abbildet. Es lässt sich erahnen, welche Herausforderungen die Ambition eine gemeinsame, nachhaltige Basis für Learning Analytics zu schaffen mit sich bringt. Die Betrachtung muss hierbei auf mehreren Dimensionen erfolgen: Neben der technischen Infrastruktur zur Datenerfassung sind dabei insbesondere die Aspekte der Einhaltung von Standards, Spezifikationen, Datenschutzrichtlinien und Best-Practices zu beachten. Ebenso wichtig sind die sinnstiftende Aufbereitung von Daten, die Verbindung verschiedener Datenströme, aber auch die letztendliche Überführung roher Daten in praxisrelevante Erkenntnisse.

Zur technischen Entwicklung wurden bereits in den vergangenen Jahren entsprechende Vorarbeiten geleistet und veröffentlicht, z.B. [Ki15]. In [LES20] die Einrichtung eines zentralen Data Warehouse unter Einbeziehung nachhaltiger Kriterien für die langfristige, sichere Nutzung dokumentiert. In dem Paper wird auch die Wahl für die Datenspezifikation xAPI dokumentiert, welche eine Möglichkeit darstellt die vielfältigen Lernaktivitäten abbilden zu können. Für einen erfolgreichen Einsatz dieser Spezifikation reicht die Einrichtung eines Learning Record Stores allein jedoch nicht aus. Die Offenheit verlangt eindeutige Definitionen, zudem stellen im Sinne von Open Science und FAIR Data Metadaten ein Schlüsselement in der Reproduzierbarkeit und Anschlussverwertung erhobener Forschungsdaten dar. Der in [Eh20] vorgeschlagene Ansatz zur Arbeit mit dieser Spezifikation befindet sich technisch in einem praxistauglichen Zustand und wird aktuell durch die Autoren und neun weitere deutsche Hochschulinstitutionen inhaltlich-semantisch ausgebaut.

Aber nicht nur beim Sammeln der Daten braucht es eine gemeinsame Sprache, auch bei der Auswertung und dem Lesen der Visualisierungen muss der Duktus der Nutzer der LA in den Gestaltungsprozess einbezogen werden. Ansätze dazu sind auf verschiedenen Ebenen anzusetzen. So ist ein wichtiges Kriterium bei der Visualisierung von Daten ob die Informationen an die intendierte Zielgruppe transportiert werden. Hier ist nicht nur sicherzustellen, dass die Auswertung korrekt ist, sondern insbesondere auch, dass die

Fokusgruppe in der Lage ist, die erforderliche Information eigenständig zu extrahieren. Wichtige Ansatzpunkte hierzu liefern Publikationen zu Grundlagen für Visualisierungen wie [Fe08], die erläutern wie Visualisierungen gestaltet sein sollten und was hilft die Kernaussage zu transportieren, die den Lehrenden hilft den eigenen Lehrprozess zu reflektieren und zu verbessern.

Insbesondere “Literacy”³ stellt einen Kernaspekt als Thema dar. Studien werden durchgeführt und Tools entwickelt, um zu ergründen, ob Visualisierungen verstanden werden, z.B. „an assessment tool to measure [data visualization literacy] differences in these abilities across groups with varied levels of experience in science, technology, engineering, and mathematics.” [MHS15, S.84].

Neue Herausforderungen stellen sich in diesem Bereich insbesondere durch die zunehmende Heterogenität und Komplexität der Daten durch die Integration von Multimodalen Learning Analytics, die über die klassische Clickstream-Interpretation hinausgehen, aber auch durch die wachsende Menge der Daten, die insbesondere durch die gegenwärtige Situation der Distanzlehre einen Schub erhalten hat. Heterogenität, Komplexität und Umfang erfordern neue Ideen, um die neuen Erkenntnisse nutzergerecht vermitteln zu können. Ein Ansatz dazu ist zum Beispiel ein mehrschichtigen Storytelling-Ansatz [Ma20].

2.3 Datenerhebung in gemischten Experimentalsettings

Daten zu erfassen ist alles andere als trivial. Kommen erschwerende, weil z.B. sehr heterogene Bedingungen hinzu, ist es oft so komplex, dass es schon ein eigenes Unterprojekt im Forschungsprozess darstellt. Bei der Aggregation von Daten aus Klassenräumen (virtuell und real), Laboren und hybriden Szenarien ergeben sich technische und datenschutzrechtliche Herausforderungen, die bereits untersucht werden.

[Ka21] und [Ch20] beispielsweise beschreiben die Herausforderungen der Aufbereitung nicht-digitaler Lernsituationen oder komplexer digitaler Lernsituationen zur (teilautomatisierten) technischen Auswertung. Hier ist derzeit noch viel Expertenwissen gefragt, da diese Daten häufig noch zeitaufwändig manuell gelabelt werden müssen.

In [Ro20] wird ein Lösungsansatz diskutiert. Dort wurde ein Tool entwickelt, um die Forschenden zu unterstützen, mit dem der Prozess erleichtert wurde. Dieses Tool ist allerdings sehr spezifisch für den gegebenen Anwendungsfall und somit nicht so leicht auf andere Kontexte übertragbar. Insbesondere werden im Annotationsprozess der Videographien unmittelbar high-level-Konstrukte, wie die Strategie der Lernenden, z.B. Phasen der Problemlösungen kodiert. Beschreibung und Interpretation erfolgen im gleichen Schritt. Dieses Vorgehen lässt sich nur schwer in einem generalisierten Format wie der erwähnten xAPI-Spezifikation abbilden. Es ist zielführend hinsichtlich der

³ In allen gegenwärtigen Ausprägungen, z.B. Data Literacy, Information Literacy, Visualization Literacy, Media Literacy und auch Statistical Literacy

gestellten Forschungsfragen, lässt aber eine Nachnutzung der Daten im Sinne von Open Science/Open Data nur bedingt zu.

[Ch20] wiederum sammeln Audiodaten und zugehörige Logs und testen Machine Learning-Verfahren auf diesen. Diese Daten werden genutzt, um Modelle zu trainieren, die diese bezogen auf Kollaboration im Klassenzimmer auszuwerten. Ein ungelöstes Problem ist hier allerdings noch der Mangel an Experten („Human Coders“), die hinreichend große Datenmengen annotieren. Es ist also noch weitere Forschung ist nötig, die insbesondere den Austausch gelabelter Datensätze einbezieht oder vorantreibt.

Ein weiterer, sehr aktueller Ansatz ist die Nutzung von Machine Learning zur Unterstützung von Experten im Labellingprozess. „Model training and label data collection go hand-in-hand in AL, with the model iteratively choosing the most informative next data point to get labeled by a human“. [Ka21, S.158]

3 Lessons learned und Ausblick

Um nachhaltige Infrastrukturen für Learning Analytics aufzubauen sind aufbauend auf unseren praktischen Erfahrungen verschiedene Aspekte von Bedeutung:

- Die Grundlage legen gemeinsame definierte Spezifikationen oder Standards als essenzielle Grundvoraussetzung, im Falle von xAPI bedeutet dies, dass es ein gemeinsames Vokabular für die Beschreibung in Form der Metadaten benötigt wird, siehe [Eh20]
- Die Prinzipien von FAIR und Open Data sind einzuhalten, um eine nachhaltige Nutzung den Austausch, sowie Transparenz sicherzustellen
- Neben dem Austausch von Daten, kann der Forschungsprozess durch den Austausch von Tools und gelabelten Daten vorangebracht werden
- Bei der Entwicklung von Learning Analytics sollten die verschiedenen Stakeholder nicht aus dem Fokus verloren werden, zum einen in der Entwicklung geeigneter Visualisierung, als auch im gesamten Entwicklungsprozess, im Sinne von Human-Centered Learning Analytics, siehe [DL18] oder die Special Section im Journal of Learning Analytics [BFM19]

Die Entwicklung von Schnittstellen, Infrastrukturen und Architekturen, die den gesamten Learning Analytics Prozess unterstützen, stellt eine große Herausforderung dar, bringt aber langfristig und nachhaltig geplant große Vorteile, sowohl für den Forschungsalltag als auch in der Unterstützung aller Stakeholder. Deshalb sollte auch künftig bei der Entwicklung projektspezifischer Werkzeuge stets die Frage einer möglichen Öffnung zu weiteren Kontexten gestellt werden.

Danksagung

Die Autoren bedanken sich für die finanzielle Unterstützung beim Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Verbundprojekts: Open Digital Lab 4 you (Förderkennzeichen: 16DHB2114).

Bibliography

- [BFM19] Buckingham Shum, S., Ferguson, R., Martinez-Maldonado, R.: Human-Centred Learning Analytics. *Journal of Learning Analytics*, 6(2), pp. 1–9, 2019.
- [Ch20] Chejara, P., Prieto, L. P., Ruiz-Calleja, A., Rodríguez-Triana, M. J., Shankar, S. K., Kasepalu, R.: Quantifying Collaboration Quality in Face-to-Face Classroom Settings Using MMLA. In A. Nolte, et al. (Hrsg.), *Collaboration Technologies and Social Computing*. Springer International Publishing, pp. 159–166, 2020.
- [DL18] Dollinger, M., Lodge, J. M.: Co-creation strategies for learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pp. 97–101, 2018.
- [Eh20] Ehlenz, M., Heinemann, B., Leonhardt, T., Röpke, R., Lukarov, V., Schroeder, U.: Eine forschungspraktische Perspektive auf xAPI-Registries. *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.*, 2020.
- [Fe08] Fekete, J.-D., van Wijk, J. J., Stasko, J. T., North, C.: The Value of Information Visualization. In A. Kerren, J. T. Stasko, J.-D. Fekete, C. North (Hrsg.), *Information Visualization (Bd. 4950)*. Springer Berlin Heidelberg, pp. 1–18, 2008.
- [Ka21] Karumbaiah, S., Lan, A., Nagpal, S., Baker, R. S., Botelho, A., Heffernan, N.: Using Past Data to Warm Start Active Machine Learning: Does Context Matter? *LAK21: 11th International Learning Analytics and Knowledge Conference*, pp. 151–160, 2021.
- [Ki15] Kitto, K., Cross, S., Waters, Z., Lupton, M.: Learning analytics beyond the LMS: The connected learning analytics toolkit. *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge*, pp. 11–15, 2015.
- [LES20] Lukarov, V., Ehlenz, M., Schroeder, U.: Building a data warehouse for multimodal learning analytics research projects. *Companion Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pp. 25–28, 2020.
- [Ma20] Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., Buckingham Shum, S.: From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2020.
- [MHS15] Maltese, A. V., Harsh, J. A., Svetina, D.: Data Visualization Literacy: Investigating Data Interpretation Along the Novice—Expert Continuum. *Journal of College Science Teaching*, 45(1), pp. 84–90, 2015.
- [Ro20] Rowe, E., Asbell-Clarke, J., Bardar, E., Almeda, M. V., Baker, R., Scruggs, R., Gasca, S.: *Advancing Research in Game-Based Learning Assessment: Tools and Methods for Measuring Implicit Learning*, 2020.