



Learning analytics as data ecology: a tentative proposal

Paul Prinsloo¹ · Mohammad Khalil² · Sharon Slade³

Accepted: 13 January 2023 / Published online: 26 January 2023
© The Author(s) 2023

Abstract

Central to the institutionalization of learning analytics is the need to understand and improve student learning. Frameworks guiding the implementation of learning analytics flow from and perpetuate specific understandings of learning. Crucially, they also provide insights into how learning analytics acknowledges and positions itself as entangled in institutional data ecosystems, and (increasingly) as part of a data ecology driven by a variety of data interests. The success of learning analytics should therefore be understood in terms of data flows and data interests informing the emerging and mutually constitutive interrelationships and interdependencies between different stakeholders, interests and power relations. This article analyses several selected frameworks to determine the extent to which learning analytics understands itself as a *data ecosystem* with dynamic interdependencies and interrelationships (human and non-human). Secondly, as learning analytics increasingly becomes part of broader *data ecologies*, we examine the extent to which learning analytics takes cognizance of the reality, the potential and the risks of being part of a broader data ecology. Finally, this article examines the *different data interests* vested in learning analytics and critically considers implications for student data sovereignty. The research found that most of the analyzed frameworks understand learning analytics as a data ecosystem, with very little evidence of a broader data ecological understanding. The vast majority of analyzed frameworks consider student data as valuable resource without considering student data ownership and their data rights for self-determination.

✉ Mohammad Khalil
mohammad.khalil@uib.no

Paul Prinsloo
prinsp@unisa.ac.za

Sharon Slade
sharon.slade@outlook.com

¹ University of South Africa, Unisa, P O Box 392, Pretoria 0003, South Africa

² Centre for the Science of Learning & Technology (SLATE), University of Bergen, Christiesgate 12, 5015 Bergen, Norway

³ EarthTrust, Little Wittenham, Abingdon OX14 4QZ, UK

Keywords Data ecology · Data ecosystem · Data interests · Learning analytics

Introduction

Data, and particularly student learning data, is an integral part of three distinct, but overlapping, discourses and practices, namely: data-driven decision-making (DDDM), educational data mining (EDM) and, more recently, learning analytics (LA). Since its emergence as a research focus and practice in 2011 (Ferguson, 2012), learning analytics has matured and become institutionalized in many contexts. This is particularly the case in the Global North (Gašević, 2018; Leitner, et al., 2017) with evidence of increasing engagement and adoption from institutions within the Global South (Falcão, et al., 2020; Hilliger, et al., 2020). The scope of learning analytics research is wide-ranging and includes, inter alia, studies and frameworks on its implementation (e.g., Greller & Drachsler, 2012; Tsai et al., 2018), analytics to inform pedagogy (Rienties, et al., 2017), identification of students-at-risk (Wong & Li, 2019), uses of student and lecturer-facing dashboards, the increasing importance of multimodal data (Valle, et al., 2021), uses of algorithmic decision-making systems (Prinsloo, 2017), and issues pertaining to ethical and privacy concerns in the collection, analysis and use of student data (Slade & Prinsloo, 2013).

Despite the potential positive impacts on, for example, pedagogy and student support, learning analytics reflects the clear asymmetrical power-relationship between the institution and its students (Broughan & Prinsloo, 2020; Prinsloo & Slade, 2016). There is also the need to understand both the potential and perils of digital data in online networks when, for example, student data from one course is combined with other sources of data (Borgman, 2018). She also makes the point that digital profile data is expandable and shareable, raising concerns about downstream uses of data as it disappears into the black boxes of algorithmic decision-making systems. When data elements are combined, whether from a single data source or several, the possibility of misuse and abuse increases. (Borgman, 2018). As such, students remain largely at the receiving end of the process, and their data interests (Hasselbalch, 2021) as a key stakeholder are often not considered (Prinsloo et al., 2019).

From the beginning, research into learning analytics has included frameworks for implementation, such as the generic frameworks by Greller and Drachsler (2012), Khalil and Ebner (2015) and the SHEILA framework (Tsai, et al., 2018). Other frameworks relating to different aspects of learning analytics also exist, including a framework for data protection (Cormack, 2016), an implementation framework with specific focus on student retention (West, et al., 2016a, 2016b), and the PERLA framework which uses learning analytics to personalize learning (Chatti & Muslim, 2019).

In this article, we explore learning analytics through an ecosystemic lens (e.g., Prinsloo, et al., 2020; Tan & Koh, 2017). Understanding learning analytics as a *data ecosystem within a larger data ecology* where different data interests play out, would have important implications for notions of data stewardship, student privacy, and ethics. We first explore data ecosystems, ecologies and data interests before

evidencing (1) the extent to which selected frameworks recognise learning analytics as a data ecosystem with dynamic interdependencies and interrelationships (human and non-human), (2) whether/how these frameworks acknowledge LA as part of a larger data ecology and (3) how they account for different data interests, with a specific focus on student data interests.

Ecologies, ecosystems and data interests: an overview

Emerging from a critical analysis of data-driven decision making (DDDM), EDM and LA, Piety (2019) proposes a conceptual model to help understand the impact of various elements (e.g., technical components, infrastructures, institutional capacity, and practices within a policy and systemic context) on uses of data in education. The use of data, and the impact of that use, should be understood in the context of the systemic nature of information, and the many hidden and visible factors within systemic technologies and infrastructures (Piety, 2019).

There is increasing academic interest in the notion of ecosystems to foreground the interconnectedness and interdependencies constituting phenomena, with specific attention to relationships and contextual variables (Guggenberger, et al., 2020). Ecosystems are seen as self-organizing arrangements between different independent actors connected through exchanges of value and mutual interest/benefit. There are many different types of ecosystems, including data ecosystems (Oliveira, et al., 2018).

Initial concepts of ecosystems emerged from the field of ecology and were introduced to replace notions of the “complex organism” and “biotic community” (Guggenberger, et al., 2020, p. 3; italics in the original). In making clearer the difference between an ecology and ecosystem, (Guggenberger, et al., 2020, referring to Kast & Rosenzweig, 1972) introduce ecosystems as “the understanding of organisms living together (**ecology**) in delimited borders inhabited by interrelated and interdependent parts and elements (**system**)” (p. 3; bold in the original).

Both data ecologies and data ecosystems are nascent concepts (Oliveira, et al., 2019). For the purpose of this research, following Guggenberger et al. (2020) and others, we understand a data ecosystem (e.g., LA) to be a subset of a broader data ecology characterised by different, competing data interests, each having ramifications for student agency, privacy and data-sovereignty.

Data ecologies

In the same year as the first Learning Analytics and Knowledge Conference (LAK'11), Breaux and Lotrionte (2011) explored cybersecurity in “the new data ecology” (p.1). They speak of the data ecology as emerging, characterized by unprecedented amounts of information sharing, allowing for changes in the availability of data, changing data needs and an increasing integration and combination of data. They further refer to the integration of data from various sources to create novel tools for individual decision-making. In this data ecology, judgments and data

range from private to non-critical with increasing demands for assurance based on controls and nuanced transparency of data and practices. Their version of data ecology includes both automated and semi-automated decisions in data supply chains that are not only highly integrated but also which rely inherently on high quality data. Such decisions based on input from a variety of sources may have far greater consequences than expected (Breaux & Lotrionte, 2011).

The notion of a data ecology also highlights the roles of governments and industry and the understanding of big data within the “larger ecology of markets, organizations, usage, culture, and the production of services” (Shin, 2016, p. 845) as well as consideration of what constitutes a socio-technical system. Data ecologies might be understood as might be understood as a heterogeneous and organic system made up of the multiple ways that people, institutions, technology, and non-human actors exchange information and values (Tang, et al., 2022). (Also see Star & Ruhleder, 1996). This resembles Davenport and Prusak’s (1997) description which refers to “information ecology”, and also Baker and Bowker’s (2007) case study which highlights the mediating function of data in tying together a complex web of standards, relationships, and professional organizations (in Tang, et al., 2022). The range of human actors in a data ecology includes data curators, data submissions, support assistance, and data flow engineers (Nadim, 2016). Central to the notion of a data ecology then are relationships between data, people, technologies, in a particular data context and policy environment, and as a result of driving forces/interests (Zhang, et al., 2021).

In their article, Tang et al. (2022) describe the potential of how data generated in one context or in response to a data query can enter new research and development lifecycles long after the completion of the original study. Queenborough et al. (2010) call this an “ecology of data-sharing”, and Sauvé and Houben (2021) “an ecology of interconnected data devices.” Many of these linkages (and actors) are invisible, resulting in an ecology of the visible and invisible (Denis, 2016; Starr & Strauss, 1999). Other iterations of data ecologies include Blok et al. (2017) mention of “an ecology of data labor” (p. 200), and “an ecology of data maintenance” (p. 201).

Steedman et al. (2020) suggest that it is necessary to see trust, and the trust deficit, as entangled in “complex relations across diverse factors” (p. 818). Emphasising only individual or institutional agency underestimates not only the inherent fluidity and intersectionalities inherent in complex ecologies of trust, but also various collective and macro-level factors. Understanding trust or the trust deficit in digital contexts as entangled in ‘ecologies of trust’ allows us to explain how trust, skepticism, and distrust interact.

We summarize key aspects of a data ecology below:

Essential markers of data ecology:

Can include a variety of data ecosystems

Evidence of links between these different ecosystems, or links to stakeholders or data interests outside of the institutional LA data ecosystem

Evidence of a multidirectional flow of information and values between non-human actors, humans, institutions, and technology

Evidence of *sharing of data with stakeholders outside* of the institutional LA data ecosystem (e.g., platform provider, analysis providers, governments, regulatory bodies, MOOC platforms, apps, service providers)

Evidence of *importing of data from stakeholders outside* of the institutional LA data ecosystem (e.g., social media, multimodal data from multiple devices, service providers)

Data ecosystems

Other research maps various definitions and types of data ecosystems, including elements and characteristics, taxonomies and typologies of data ecosystems and overviews of the different actors and their roles (Guggenberger, et al. 2020). Data ecosystems are defined as socio-technical networks that allow for cooperation between independent entities, including businesses, institutions, and people (Oliveira, et al., 2018). The metaphor of ecosystem is invoked to allow us to imagine social systems that are interdependent in terms of individuals, organizations, resources and material infrastructures and that emerge in systems that are both technology-enabled as well as information-intensive (Oliveira, et al. 2019). These data ecosystems are furthermore technological cultural and social phenomena founded on connections and relationships between a range of technology, businesses, actors, industries and governments (Oliveira, et al., 2019).

Summarizing their analysis of various definitions, Oliveira et al. (2019) discuss data ecosystems in terms of actors and roles and their relationships and resources in which “a loose set of interacting actors ... directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure)” (p. 604). Additionally, each of these actors perform a variety of roles and relates to other actors. Data ecosystems not only enable the sharing of information, data, experiences and knowledge, but also of the communication through which value is created and shared between multiple, different networks of actors within a larger ecology of individuals, organizations and networks.

Central to data ecosystems are the relations and transactions between its autonomous but interrelated actors (human and non-human) flowing from and constituting varying degrees of interdependency. Different data ecosystems have different attributes, qualities and operating standards and are subjected to context-specific regulatory and policy conditions as well as political, cultural, economic, environmental and technological influences and forces (Oliveira, et al. 2019). [Also see the discussion of Cui et al. (2020) on components in data ecosystems, and the meta-dimensions (economic, technical and organizational) of data ecosystems suggested by Gelhaar, Groß and Otto (2021)].

In the context of this article, the analysis of the roles of the different actors in a data ecosystem by Oliveira et al. (2019) is very helpful. These authors list the following actors and roles (Table 1 below):

Benefits of data ecosystems include, inter alia, improvements in political, social, economic aspects as well as in the quality of services and data, as well as mutually beneficial communication between different actors. The ‘success’ of a data ecosystem depends on the extent to which the different actors, and the ecosystem as a

Table 1 An overview of actors and their roles in a data ecosystem

Role	Brief description
Data user	The person who is “responsible for directly or indirectly consuming data” (p. 606), the end-users
Data providers	Who are “not necessarily responsible for data generation” (p. 608). This duty may be delegated to a different position known as a data producer, who is in charge of gathering or producing data. Additionally, this position may gather, assemble, and package data
Re-user	“...using data to develop applications or services aimed at data users” (p. 608)
Keystone actor	Responsible for “driving forces behind the ecosystem as well as providing stability in unstable environments” (p. 608). They also “foster a set of formal directives, rules, and practices to drive a Data Ecosystem” (p. 608)
Service provider	Responsible for producing and supporting software resources such as “tools, applications, services, libraries, or other software products” (p. 608). Service providers might be hired by actors who lack the skills and resources to process data themselves
Policies, laws and rules parties	One of the most important roles—“responsible for creating the rules and policies to encourage and to control the participation of actors” (p. 608). e.g. Governments, Data Ecosystem Founders, or Standardization Institutions
Infrastructure provider	Refers to the function that supports the operations of other functions. Information and communication technologies (ICT) resources and services, such as hosting and storage capacity, are included in infrastructure
Data consultant	Supports other roles in not only analyzing the various possibilities of data, but also identifying the needs of the different actors
Data sponsor	The data sponsor is in charge of advancing the open data project via both government funding initiatives and private contributions
Data curator	“...responsible for the quality and availability of data” (p. 608)

whole can address the potential of a lack of participation and interaction between different actors, a lack of resources and/or technical expertise to sustain the ecosystem, complexities in data collection, discovery and use as well as concerns about liability, privacy and confidentiality (Oliveira, et al., 2019).

Concluding our brief discussion of data ecosystems, it is insightful to note that Oliveira et al. (2019) state that it is “a new field of growing importance” and that theories informing data ecosystems have not yet been well-developed, while the different existing models “cover only a small fragment of how a Data Ecosystem works” (p. 626).

Essential markers of data ecosystem:

Maps the different *actors* (Data user, Data providers, Re-user, Keystone actor, Service provider, Policies, Laws and Rules Parties, Infrastructure provider, Data consultant, Data Sponsor and Data Curator) their *roles*, and the *relationships* and *interdependencies* between them

Clarity about *data flows* in the ecosystem

Acknowledges context—macro-political/regulatory, market/industry and institutional
 Maps the economic, technical and organizational dimensions
 Acknowledges the social, political and cultural nature of the ecosystem
 Is clear about the value created by the data ecosystem
 Clarity around the barriers and challenges
 Provides an overview of infrastructure, standards, service and applications

Data interests

Hasselbalch (2021) distinguishes between ‘human interest’ and ‘data interest’. Human interest refers not only to the outcome of data uses in, for example, algorithmic processes, but to human involvement from the point of defining the problem, setting the scope and limitations of the collection of data and the training of the machine-learning algorithm. Human interest in the design and deployment of algorithms also focuses on human agency and control, and concerns about the replacement of human expertise and labour by AI. An example of human interest in the use of data in AI is value sensitive design (VSD) where specific human values are embedded to address a moral or ethical dilemma. A data interest can be described as a purpose or a motive that is translated into specific qualities of a data technology that arranges data in ways that enable the agency of certain interests in the data that is stored, processed and analyzed by AI (Hasselbalch, 2021). Examples of values that can be designed include “data privacy, accessibility, responsibility, accountability, transparency, explainability, efficiency, consent, inclusivity, diversity, security, and control” (Hasselbalch, 2021). (Also see Slota, et al., 2021; Umbrello & Van de Poel, 2021).

Different stakeholders, ranging from developers to users, institutions or commercial entities, have different (often conflicting) data interests (Delgado, et al., 2021). There are no neutral or impartial data interests. Understandably there are concerns that AI is invested “with the interests of the powerful—governments, public institutions and big data industries” (Hasselbalch, 2021). (Also see Figueras, et al., 2021). Inherent in understanding and addressing conflicting data interests is the notion of ‘informational asymmetries’ impacting on citizens and individuals’ agency. Data interests therefore refer to informational relationships entangled in different socio, political, environmental, legal and economic interests.

Hasselbalch (2021) puts forward five clusters of themes to illustrate human data interests: data as resource, data as power, data as regulator, data as vision and data as risk.

Data as resource In data discourses, ‘data as resource’ finds expression in metaphors such as data as ‘raw material’ that can be processed and turned into products. Individuals may want to protect this resource and be able to exchange it for, e.g., services. Data scientists may see this resource as a training set for algorithms, commercial entities capitalize on the inherent value in data and governments may see data as a resource for managing security and risk. Data as resource also recognizes that “informational asymmetries also create very tangible social and economic gaps

between the data rich and data poor, which is a conflict of interest on a more general structural level of society” (Hasselbalch, 2021). In the context of learning analytics, student data has been described as “the new black” (Booth, 2012), and “the new oil” (Watters, 2013). Monetizing student data forms part of the basis for the platformitization of higher education (Komljenovic, 2021).

Data as power This theme is closely connected to ‘data as resource’. “Distribution of data/information amounts to the distribution of power in society” (Hasselbalch, 2021). Access to data allows individuals or commercial entities to make informed choices specific to their interests. Understanding data as power is seminal to Data Feminism (D’ignazio & Klein, 2020), Critical Data Studies (Iliadis & Russo, 2016) and understandings of student vulnerabilities in higher education (Prinsloo & Slade, 2016).

Data as regulator “represents the legal enforcement of the power balancing of data interests” (Hasselbalch, 2021), and in an ideal situation, law and technology will supplement one another. The metaphor of ‘data as regulator’ asserts the role of a particular data design, the implementation of legal frameworks and the realization of legal principles. In other words, the data interests of those with more power regulate what is possible, and in the case of governments, what is acceptable. In the context of LA, data as regulator plays out in many of the data proxies used in determining, for example, students-at-risk (Archer & Prinsloo, 2020).

Data as vision Having access to the decisions and process of algorithmic decision-making systems is core to more trustworthy AI. It is possible to say that the control of visibility, or the architecture of visibility in new technology environments, represents a form of social organization and power distribution. (Hasselbalch, 2021). In the context of education, data as (prophetic) vision forms part of the data imaginary sold to higher education institutions promising “*speedy, accessible, revealing, panoramic, prophetic and smart*” (Beer, 2019, p. 22; italics in the original) insights into student learning.

Data as risk Data interests are also “correlated with the act of assessing the risks of data design” and the extent to which there are potential negative impacts linked to the design on, e.g., the economy, democratic institutions, and individuals. In the light of the fact that data is inherently a risk that should be foreseen and managed, potential risks must be prevented and managed (Hasselbalch, 2021). Criminals can steal data, data can be leaked, and corrupted.

Hasselbalch (2021) concludes his proposal on data interests by pointing to the increasing informational inequalities between (groups of) individuals such as those already marginalized and/or from minority groups, civil servants and developers, as well as stakeholders from broader society inter alia governments, citizens and industries. Only when these asymmetries, power relations and the different interests at play are taken seriously can we develop systems to trade-off between the different data interests.

In addition, there are increasing concerns about the exclusion of the data interests of those from whom data are collected, and who may be most affected by its use by other, more dominating data interests. Such groups include Indigenous peoples, and others outside white male, heteronormative classifications (e.g., LGBTQIA+, blacks, Hispanics, females, etc.). Individuals in these groups may be impacted

negatively through bias, unfairness, obfuscation in algorithmic decision-making, and injustice (e.g., D'ignazio & Klein, 2020; Skinner-Thompson, 2020; Walter, et al., 2021). In the context of higher education, student data interests are often unaccounted for in the discourses and practices surrounding and informing the design and implementation of learning analytics (Broughan & Prinsloo, 2020; Khalil, et al., 2022; Madaio, et al., 2021).

Essential markers of data interests:

Acknowledges the different data interests in the data ecosystem/ecology

Evidence of data as resource

Evidence of data as power

Evidence of data as regulator

Evidence of data as vision

Evidence of data as risk

Evidence of ensconcing the data interests of students

Evidence of the data interests of specific groups

Research methodology and norms

The research questions guiding this research are:

1. To what extent do selected frameworks for the implementation of LA understand LA as an ecosystem that is part of a larger data ecology where different data interests compete?
2. What are the implications for LA, and specifically for students' data interests?

In following Elo et al. (2014), ensuring the trustworthiness of this research encompassed several phases ranging from the data collection, sampling strategy and selecting a coding heuristic, to categorization and abstraction, interpretation and reporting the results and the analysis process.

In selecting LA frameworks for this analysis, this research used the analysis and findings of Khali et al. (2022). Their research analysed 46 LA frameworks. Their research found that LA frameworks share several elements and characteristics such as (1) source (conceptual or empirical); (2) development focus (learner, faculty, course design, research and development, student support and external); (3) application focus (learner, faculty, course design, research and development, student support and external); (4) a form of representation (table, figure and/or other); (5) data sources (virtual learning environment, multichannel, other); (6) data types (system logs, learning artifacts, questionnaire/survey and multimodal); (7) focus (retention/support, pedagogy and other); (8) context (pre-higher education and higher education); and (9) ethics and privacy. [See Khalil et al. (2022) for full description and analysis].

From the 46 frameworks evaluated, we identified 11 frameworks that were the most comprehensive against these applied criteria, *excluding* whether the framework included a graphical presentation, whether there was evidence of its application and the context of the application (e.g., K-12 or higher education). The following 11 (listed alphabetically) were identified as the most comprehensive (Table 2 below):

Finalizing the corpus of 11 frameworks was followed by a deductive, directed content analysis (Assarroudi, et al., 2018; Elo & Kyngäs, 2008; Hsieh & Shannon, 2005) using key elements of data ecologies, ecosystems, and data interests as heuristic as strategy to answer the first research question—To what extent do selected frameworks for the implementation of LA understand LA as an ecosystem that is part of a larger data ecology where different data interests compete?

The analysis process started by “[u]sing existing theory or prior research, researchers begin by identifying key concepts or variables as initial coding categories” (Hsieh & Shannon, 2005, p. 1281). The next step included defining operational definitions for each category using theory in a categorization matrix (Elo & Kyngäs, 2008) or category schemes (Downe-Wamboldt, 1992) (See Fig. 1 below):

Two researchers independently used the coding heuristic and indicated which elements were found in each article. Where there was a difference of opinion, the two researchers discussed the differences and where the coding could not be resolved, the third coder engaged with the code in question and made the final call. Elo et al., (2014) further recommends that in deductive content analysis, a reliability check to be presented. In the case of this research, we calculated a relevant reliability value called Fleiss Kappa as a measure to evaluate agreement (Fleiss, et al., 2013). The reported Fleiss Kappa is ($\kappa = 0.946$, subjects = 220, raters = 2, and $p < 0.005$) indicating a high level of agreement between the two researchers.

In reporting on the analysis (in the next section), the researchers attempted to provide a detailed and rich representation of the process and findings (Elo, et al., 2014).

Analysis and findings

In this section, an overview of the application of the heuristic will be provided before each element of the heuristic is discussed with quotations to illustrate the findings. By way of introduction, it is important to highlight the researchers’ position in following Guggenberger et al, (2020) and others, that we understand a *data ecosystem* (e.g., LA) to be a *subset* of a broader data ecology characterised by different, competing data interests, each having ramifications for student agency, privacy and data-sovereignty.

Figure 2 below presents an overview of the analysis of the 11 frameworks according to the coding heuristic. In the figure ‘O’ represents no evidence found; ▼ presents evidence of the element found in the framework, while ‘X’ records where there was no agreement between the researchers on interpretation of the evidence.

While only one framework (Slade & Prinsloo, 2013) explicitly links LA to being part of a larger data ecology, several refer to linking to other data ecosystems, multi-directional flows of data (including importing from and exporting to other data ecosystems (e.g., Greller & Drachsler, 2012; Hernandez-Leo et al., 2019 and West,

Table 2 Selected frameworks

	Framework	Comprehensiveness count
1	Christopoulos, A., Mystakidis, S., Pellas, N., & Laakso, M. J. (2021). ARLEAN: An Augmented Reality Learning Analytics Ethical Framework	15
2	Christopoulos, A., Pellas, N., & Laakso, M. J. (2020). A learning analytics theoretical framework for STEM education virtual reality applications	15
3	Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics	15
4	Hernández-Leo, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools	15
5	Kazanidis, I., Pellas, N., & Christopoulos, A. (2021). A learning analytics conceptual framework for augmented reality-supported educational case studies	20
6	Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design	15
7	Prasad, D., Totaram, R., & Usagawa, T. (2016). A framework for open textbooks analytics system	16
8	Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas	14
9	West, D., Heath, D., & Huijser, H. (2016a). Let's talk learning analytics: A framework for implementation in relation to student retention	14
10	West, D., Huijser, H., & Heath, D. (2016b). Putting an ethical lens on learning analytics	15
11	Zotou, M., Tambouris, E., & Tarabanis, K. (2020). Data-driven problem-based learning: enhancing problem-based learning with learning analytics	14

Data ecology	Reference to being part of a bigger data ecology
	Mentions links to other ecosystems
	Evidence of multidirectional flow of information between human and non-human actors/providers, apps, service providers
	Evidence of importing data from other ecosystems
	Evidence of exporting data to other ecosystems
Data ecosystem	Maps different actors, and their roles
	Maps relationships between actors
	Maps inter-dependencies between actors
	Provides overview of infrastructure, standards, service, and applications
	Acknowledges context – macro-political
	Acknowledges context –social
Data interests	Acknowledges context – institutional
	Acknowledges different data interests
	Evidence of data as resource
	Evidence of data as power
	Evidence of data as regulator
	Evidence of data as vision
	Evidence of data as risk
	Ensconcing the data interests of students
	Evidence of the data interests of specific groups such as indigenous peoples, LGBTQIA+, blacks, Hispanics females, etc

Fig. 1 Coding heuristic

et al., 2016a, 2016b). What was not always clear in the interpretation of the evidence was whether the linkages to other ecosystems, and the importing of and/or exporting data to other ecosystems, implied that these were outside of the institution.

Regarding the extent that the frameworks regarded LA as an ecosystem, those from Greller and Drachsler (2012), Hernández-Leo et al. (2019) and West et al. (2016a, 2016b) included all the elements. Interestingly, most frameworks did not acknowledge the macro-political context in which LA functions. All 11 of the frameworks listed the different stakeholders or actors and their respective roles with most mapping the relationships between these. Five of the 11 frameworks specifically mentioned or discussed interdependencies between these actors.

The main emphases pertaining to data interests in all the analyzed frameworks were found to be “data as resource” and “data as regulator”, with seven of the 11 frameworks recognizing a variety of risks inherent in the collection, analysis and use of student data. Only three of the 11 frameworks move beyond student “data as resource” and the use of student data to regulate various aspects of the learning design and facilitation of learning, to mention or discuss ways in which the data interests of students can be considered and protected. Most of the frameworks saw students only as data subjects and as recipients of receiving the benefits of the

Framework	Data ecology					Data ecosystem							Data interests							
	Reference to being part of a bigger data ecology	Mentions links to other ecosystems	Evidence of multidirectional flow of information between human and non-human actors/providers, apps, service providers	Evidence of importing data from other ecosystems	Evidence of exporting data to other ecosystems	Maps different actors, and their roles	Maps relationships between actors	Maps inter-dependencies between actors	Provides overview of infrastructure, standards, service, and applications	Acknowledges context – macro-political	Acknowledges context –social	Acknowledges context – institutional	Acknowledges different data interests	Evidence of data as resource	Evidence of data as power	Evidence of data as regulator	Evidence of data as vision	Evidence of data as risk	Ensconcing the data interests of students	Evidence of the data interests of specific groups such as indigenous peoples, LGBTQIA+, blacks, Hispanics females, etc
Christopoulos et al., (2021)	○	○	✕	▼	○	▼	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○
Christopoulos, et al., (2020)	○	○	✕	▼	○	▼	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○
Greiler & Drachler (2012)	○	▼	▼	○	○	▼	▼	▼	▼	▼	▼	▼	○	○	○	○	○	○	○	○
Hernández-Leo et al., (2019)	○	▼	▼	○	○	▼	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○
Kazanidis et al., (2021)	○	▼	✕	▼	▼	▼	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○
Law & Liang (2020)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Prasad, Totaram & Usagawa, (2016)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Slade & Prinsloo (2013)	▼	▼	○	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
West, Heath & Huisjer (2016)	○	▼	✕	▼	○	▼	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○
West, Huisjer & Heath (2016)	○	▼	▼	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Zotou, Tambouris & Tarabanis (2020)	○	○	▼	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Fig. 2 Overview of the application of the coding heuristic

analysis of their data. Only two frameworks (Hernández-Leo, et al., 2019 and Kazanidis et al., 2021) specifically address the data interests of specific groups (see discussion below).

In the following section, we briefly provide selected examples of the different elements of the coding heuristic.

Evidence of LA as data ecology

• Reference to being part of a bigger data ecology

There is ample evidence of data assemblages that arise from interactions between different data ecosystems and networks (Boyd, 2022; Kitchin, 2014). A data assemblage can be defined as a complex socio-technical system that is composed of many apparatuses and elements that are thoroughly entwined and whose central concern is the production, management, analysis, and translation of data and derived information products for commercial, governmental, administrative, bureaucratic, or other purposes” (Kitchin & Lauriault, 2014, p. 4).

As such, evidence of LA as data ecology will entail reference to part of a bigger, data-exchange/production/aggregation network consisting of multiple, linked servers, and a range of human and non-human actors where students’ data are shared and supplemented with data from other networks/providers/brokers. The analysis of the 11 LA frameworks did *not* provide evidence of an awareness of how LA functions as an ecosystem in a larger ecology of data networks and assemblages. The analysis

found only one framework that points to an LA ecosystem forming part of a bigger data ecology. Slade and Prinsloo (2013) reference student data that are hosted on servers external to the institution according to different “standards, owners and levels of access” (quoting Ferguson, 2012, in Slade & Prinsloo, 2013, p. 1515). They also state that “The distributive nature of networks and the inability to track activity outside of an institution’s internal systems also affect the ability to get a holistic picture of students’ lifeworlds” (p. 1515).

- *Mentions links to other ecosystems*

Higher education institutions as ecosystems host a variety of sub-ecosystems such as the student information system, the learning management system (LMS) and various other sub-ecosystems in administration, student support, the management of residences and so forth (West, et al., 2016a). Some of these ecosystems may be linked while others appear to function independently from others. For example, Christopoulos et al., (2021) refer to “Combining data from diverse e-learning platforms” (p. 4), and Greller and Drachsler (2012) to “The proliferation of *interactive* learning environments, learning management systems (LMS), intelligent tutoring systems, e-portfolio systems, and personal learning environments (PLE) in all sectors of education [that] produces vast amounts of tracking data” (p. 43).

Slade and Prinsloo (2013) point to the reality that

As learners’ digital networks increasingly include sources outside of the LMS, institutions may utilize data from outside the LMS (e.g., Twitter and Facebook accounts, whether study related or personal) to get more comprehensive pictures of students’ learning trajectories. The inclusion of data from sites not under the jurisdiction of an institution raises a number of concerns given that universities have no control of external sites’ policies, and the authentication of student identity is more problematic” (p. 1524). [Also see Kitto, et al., 2015; Liao & Wu, 2022; Wu, 2021].

There are also information flows between students as data subjects, teachers and institutions and finally “[g]overnment agencies [that] may collect cross-institutional data to assess the requirements of Higher Education Institutes (HEI) and their constituencies” (Greller & Drachsler, 2012, p. 46).

- *Evidence of multidirectional flow of information between human and non-human actors/ providers, apps, service providers*

Christopoulos et al. (2021) refer to “[c]ombining data from diverse e-learning platforms can facilitate the evaluation of the instructional decisions” (p. 5) that creates “a loop which would allow the engaged parties to feed the system” (p. 6). They describe three categories of data processes or flows with the *first* category involving clustering information from utilized technological devices e.g., LMS, mobile phones, students’ physical environment (e.g., noise levels and location tracking) and students’ gestures (e.g., facial expressions, eye movement). The *second* category refers to the clustering of information related to pedagogical

performance in a specific discipline (e.g., duration, learning objectives) with students' demographic data and the *third* category clustering information emerging from psychometric tests and sensors (e.g., stress, temperature).

In the specific context of the nexus between learning design and LA, Hernandez-Leo et al. (2019) emphasise the bidirectional interaction between learning design and learning analytics (using a variety of data from a range of sources). "Learning analytics outputs increase their meaningfulness when aligned with pedagogical intentions and learning designs can be strongly influenced by the data analytics available before or during the learning design activity" (p. 11).

- *Evidence of importing data from other ecosystems*

Two of the analysed frameworks refer to importing data from other ecosystems. Christopoulos et al. (2021) refer to data generated from mobile phones, the physical environment (e.g., location tracking), linked to students' demographic data and students' sensory data, which will usually form part of different, and possibly overlapping, data ecosystems. Hernandez-Leo et al. (2019) discuss importing data from physical spaces (e.g., the use of sensor-based technologies) and other institutional platforms, student information systems and surveys that "can complement these sources with information about academic profiles, demographics, and students' satisfaction ratings" (p. 3). They also allude to a range of different data sources such as "virtual learning environments, web tools, attendance registers and student feedback questionnaires" (p. 7). These data are combined with evidence of teachers' designs using "*profiles* (students names, emails, IDs, etc.), *process* (e.g., number of views and editions in resources corresponding to specific activities of the learning design, attendance, submissions), checkpoints (warnings related to the usage of resources, etc.), and *performance* (e.g., comments inserted by teachers) (p. 7; italics in the original).

- *Evidence of exporting data to other ecosystems*

Greller and Drachsler (2012) mention governments who "may collect cross-institutional data to assess the requirements of Higher Education Institutes (HEI) and their constituencies" (p. 46), while Slade and Prinsloo (2013) warn that

When teaching and learning opportunities incorporate social networks outside of the institutional LMS, institutions should also ensure that learners are explicitly informed of the public nature and possible misuse of information posted on these sites, and instructors should consider the ramifications before using such sites" (p. 1525).

In summary, there is a paucity of recognition in the analysed frameworks that LA forms part of a bigger data ecology of an ever expanding and intensifying "data gaze" (Beer, 2019) and interconnecting networks, and brokers, consisting of human and non-human actors and systems.

Evidence of LA as data ecosystem

Maps different actors, and their roles

Table 3 presents an overview of the stakeholders mentioned by some of the analysed frameworks.

Interestingly, while students are specifically mentioned as stakeholder in at least two of the analysed frameworks (Christopoulos et al., 2021; Greller & Drachsler, 2012), and though their specific roles are not described, they are not seen as core participants in shaping LA—rather that LA has their interests at heart.

Maps relationships between actors

In one of the earliest LA frameworks, Greller and Drachsler (2012) map the relationships and information flows between the different dimensions in LA ranging from stakeholders, limitations, instruments, objectives and data. The four stakeholders mentioned are the institution, teachers, learners and ‘others’ while the instruments include technology, algorithms, theories and ‘other’. While it does not map the inter-relations in detail, the framework and the mapping of the critical dimensions provided a much-needed basis for further development in LA.

Kazanidis et al. (2021) mention the need for an “analysis of the characteristics of the key stakeholders, their role in the instructional process, the actions that execute, and their relationship to the educational content, the educational context, and the learners” (p. 6). Three frameworks approach this dimension of LA as an ecosystem, albeit differently. For example, Hernández-Leo et al. (2019) refer to different ‘layers’ (e.g., LA, community analytics, and design analytics) and the interactions between these layers, while Law and Liang (2020) map three levels in the use of LA in pedagogical decision-making namely course, curriculum and task level. These three different levels use different LA techniques (e.g., descriptive analytics, inferential analysis, temporal analysis, etc.) different LA

Table 3 Frameworks and their stakeholders

Framework	Actors mentioned
Christopoulos et al., (2021)	Software developers, educators, students, instructional designers and policy makers
Greller and Drachsler (2012)	“...the main stakeholder groups of LA in formal learning situations are learners, teachers, and educational institutions. These may be expanded or substituted by other stakeholder groups, such as researchers, service providers, or governmental agencies” (p. 46)
Kazanidis et al., (2021)	Educators, instructors, data analysts, instructional designers and policy makers
Law and Liang (2020)	Teachers, data scientists, user interface designers, and LA researchers. Emphasises that LA should be co-designed and implemented in collaboration with these stakeholders. They also refer to a “‘learning engineering team’ comprising technologists, education professionals, learning scientists, and design experts” (p. 99)

functionalities (e.g. prediction, learner-oriented feedback, teacher-oriented feedback, etc.) and different data types (e.g., exam grades, LMS log files, quiz grades, etc.). (Also see the layers suggested by Zotou et al. [2020](#)).

Maps interdependencies between actors

Six of the 11 frameworks refer to the interdependencies between different actors and systems. For example, Greller and Drachsler ([2012](#)) point to the interdependencies between the different actors and processes stating, “We would, therefore, strongly welcome if application developers and researchers would not only make their technical environment known and open, but also describe the contextual environment and expectations from the users (e.g., required competences) along the lines of the framework” (p. 54). In their analysis, Hernández-Leo et al. ([2019](#)) discuss the interrelationships and interdependencies not only between actors, but between different layers in the nexus of LA and learning design. Actors referred to include the community of practitioners in the layer of community analytics, learning designers and their tools in the design layer and learners and other participants in the LA layer. The success of the data ecosystem depends on not only expertise and engagement of the different actors in these layers, but also the links and interdependencies between these layers.

Provides overview of infrastructure, standards, service, and applications

Four of the frameworks deal with LA in the context of specific technologies that imply different ecosystems and the need for infrastructure, standards and applications. For example, Christopoulos et al. ([2020](#)) in the context of LA and Virtual Reality (VR) refers to how student data gathered during registration and administrative processes can be linked to learning artefacts and learner behavioral data on the LMS, plus sensory data gathered from the VR equipment. [Also see Christopoulos et al. [2021](#) and Kazanidis et al. ([2021](#)) (Augmented Reality), and Prasad et al. ([2016](#)) (open textbooks)].

West et al. ([2016a](#), [2016b](#)) gathered data on “infrastructure, policies, strategy, governance and concerns related to learning analytics from an institutional point of view” (p. 909) and discuss three main factors regarding LA infrastructures, namely system reliability, system sophistication and relevant expertise. These three factors further include:

- Digital availability and integrity of data
- Integration, continuity, and availability of data systems
- Technical, pedagogical, statistics, and project management expertise
- Data stewardship
- Policy and procedures

Acknowledges context—macro-political

At least four frameworks explicitly acknowledge the macro-political context in which LA functions. Law and Liang (2020) recognise that embedding AI in teaching and learning contexts is not only technical but includes “related human challenges that are cognitive, social, organizational, and political in nature” (p. 2) while Slade and Prinsloo (2013) state that a “sociocritical perspective entails being critically aware of the way our cultural, political, social, physical, and economic contexts and power relationships shape our responses to the ethical dilemmas and issues in learning analytics” (p. 1511).

Acknowledges context—social

The majority ($n=7$) of analysed LA frameworks acknowledged the social dimension of LA as an ecosystem. Evidence of the social element of LA as ecosystem was found in Hernández-Leo et al. (2019) who speak of a community of teachers and other stakeholders—“The notion of community can be considered in a general sense either as an educational center or a cross-institutional community where teachers and collaborators share and jointly contribute to devising educational designs” (p. 5). There is, however, also understanding of student learning in social contexts and Slade and Prinsloo (2013, p. 1524) warn that institutions’ predictive models only explain ““a portion of the wide range of behaviors that constitute the universe of social interactions” (quoting Subotzky and Prinsloo, 2011) and that “[d]ata harvested in one context may not be directly transferable to another” (p. 1524).

Acknowledges context—institutional

It should not come as a surprise that with the exclusion of one framework, all the analysed frameworks acknowledge the institutional context of LA as an ecosystem. For example, West et al. (2016a, 2016b) map a framework for *institutional* adoption of LA and mentions specific elements such as institutional culture, level of sponsorship, governance arrangements, alignment with institutional strategy, sustainability and positioning LA within the institution. These authors further state that “A university’s value base is developed over time and based on various factors including historical foundation, geographic location, student cohort and leadership. Institutional values evolve into an institutional culture that should be reflected in its policies and processes, its focus, and ultimately the decisions that are made” (p. 908).

Data interests

Acknowledges different data interests

Christopoulos et al. (2021) propose that “students, teachers, instructional designers, institutions, [and] industrial agents” (p. 10) should be involved. Greller and

Drachsler (2012), after referring to a range of stakeholders e.g., learners, teachers, educational institutions, researchers, service providers, and governmental agencies state that “Each of the groups has different information needs and can be provided with tailored views on information using LA” (p. 46).

West et al. (2016b) refer to different data interests and power relations pertaining to the collection, analysis and use of student data and state that “In other words, senior management may have very different reasons for wanting to collect and access various types of learning analytics data than lecturers” (p. 914).

Evidence of data as resource

All the analysed frameworks see (student) data as a resource. One of the earliest examples of data as resource is in the early LA framework developed by Greller and Drachsler (2012), where they refer to (student) data as offering “unused opportunities for the evaluation of learning theories, learner feedback and support, early warning systems, learning technology, and the development of future learning applications” (p. 43). As such, teachers and institutions can use student data to plan interventions or adapt their assessment strategies and pedagogies. Slade and Prinsloo (2013) state that the optimal “use of student-generated data may result in institutions having an improved comprehension of the lifeworlds and choices of students, allowing both institution and students to make better and informed choices and respond faster to actionable and identified needs” (p. 1512). They further propose that “higher education cannot afford to not use [student] data” (p. 1521). Pointing to student data as core to LA—e.g., student characteristics, activities, and performance, Kazanidis et al. (2021) state that student data “reveal insights related to their cognitive patterns and behavioral decisions” (p. 2). [Also see Christopoulos et al. (2021) for an overview of (student) data as resource in the nexus between LA and VR/AR, Hernandez-Leo et al. (2019) for a discussion of the use of student data in learning design].

Evidence of data as power

Five of the analysed frameworks refer to (student) data as a form of power. For example, West et al. (2016a, 2016b, p. 907) quotes Swenson (2014) who asks.

who has the power to:

- Make decisions about the learning analytics model and data
- Legitimize some student knowledge or data and not others
- Focus on potential intervention strategies and not others
- Give voice to certain students and not others, and
- Validate some student stories and not others.

The socio-critical approach to the ethical challenges in LA proposed by Slade and Prinsloo (2013) foregrounds ‘data as power’ by referring to the unequal power relations between “learners, higher education institutions, and other stakeholders (e.g., regulatory and funding frameworks)” (p. 1511). These power relations are

inherently unequal and result in “increasing surveillance in teaching and learning environments” which affect “the work and identities of tutors, faculty, and administrators, disrupting existing power relations and instituting new roles and responsibilities” (p. 1515).

Central to the notion of ‘data as power’ is data ownership (Greller & Draschler, 2012). In this respect, Greller and Draschler (2012) state that “Because the technical systems producing and collecting data are typically owned by the institution, the easiest assumption would be that this data belongs to them. However, which employees of the institution exactly are included in the data contract between a learner (or their parents) and the educational establishment, is as yet unresolved” (p. 49). Further, the authors posit that “the real dangers that the extended and organized collection of learner data may not so much bring added benefits to the individual, but instead provides a tool for HEIs, companies, or governments to increase manipulative control over students, employees, and citizens, thereby abusing LA as a means to reinforce segregation, peer pressure, and conformism rather than to help construct a needs-driven learning society” (p. 54).

Outside of consent for their data to be used, *ownership* of data and the inherent power in ownership is absent from all the frameworks. Though not supporting student data sovereignty, Christopoulos et al. (2021) emphasize voluntary consent, taking the position that “Student data should also not be sold” and that “Learners should have the right for their data to be removed from the system after a given period of time while institutions oversee this entire process” (p. 10).

Evidence of data as regulator

Most of LA frameworks provide explicit evidence of how (student) data acts as regulator—used to shape, inter alia, learning, learning design, assessment, interventions and student support. For example, Christopoulos et al. (2021) discuss how data “enables educational technologists and instructional designers to develop a better understanding of students’ reactions to different types of stimuli under controlled (e.g., classroom) and uncontrolled (e.g., home) circumstances” (p. 6). Data also allows for the employment of associations “to make predictions about the students’ performance and suggestions on the best course of action to improve their learning curve” (Christopoulos et al., 2021, p. 7). Data as regulator is also directly linked to data as resource discussed earlier.

Evidence of data as vision

Interestingly, there were only two frameworks that provided a sense of ‘data as vision’. The initial expectation was that we will find evidence of how data and its analysis provide ‘prophetic’ (Beer, 2019) insights into student behavior. Though there was ample evidence of ‘data as resource’ (as basis for insights) and ‘data as regulator’ (allowing insights to steer action), the two LA frameworks provided a different interpretation or aspect of ‘data as vision. For example, Christopoulos et al. (2021) mention how “coding parameters [in algorithms] should be communicated

transparently to everyone affected" (p.9). The need for transparency and accessibility is also mentioned by West et al. (2016a, 2016b).

Both frameworks call for those who may be impacted to have access to and insight into decisions made by the often opaque 'black box' of algorithmic decision-making systems. 'Data as vision', in the sense of the potential of data to provide insights into student behavior, is covered by 'data as resource' and 'data as regulator'.

Evidence of data as risk

Seven of the 11 LA frameworks deal with 'data as risk'. Greller and Drachsler (2012) moot that "prediction suffers potentially from big ethical problems [...] in that judgements about a person, whether originating from another human or a machine agent, if based on a limited set of parameters could potentially limit a learner's potential" (pp. 47–48). Interestingly, these authors also highlight the responsibility that knowing more brings—"the more access to information about a data subject a data client has, the higher the responsibility is to use this information in a sensitive and ethical way" (p. 51). (Also see Prinsloo & Slade, 2017).

There are, however, other risks including that "ethical risks are the exploitation of such data for commercial and similar purposes, or data surveillance issues (social sorting, cumulative disadvantages, digital stalking) (Greller & Drachsler, 2012, p. 51). These authors also mention a risk not always considered, namely the "inherent danger that we perceive is that the simplicity and attractive display of data information may delude the data clients, e.g., teachers, away from the full pedagogic reality" (Greller & Drachsler, 2012, p. 52).

Several frameworks mention the need for compliance to regulatory frameworks such as the GDPR and the risk of non-compliance (e.g., Christopoulos, et al., 2021), the risks of bias and stereotyping (Slade & Prinsloo, 2013), and the risks of the erosion of trust between students and the institution (West, et al., 2016a, 2016b) when institutions venture into students' social media spaces.

Ensconcing the data interests of students

Other than the issue of data ownership and student data sovereignty, Greller and Drachsler (2012) state that it is of "critical importance for its acceptance that the development of LA takes a bottom-up approach focused on the interests of the learners as the main driving force" (p. 54) as well as flagging the importance of preventing the reconfirmation of "old-established prejudices of race, social class, gender, or other with statistical data, leading to restrictions being placed upon individual learners" (p. 48).

Linked to the danger of perpetuating bias and prejudice, Slade and Prinsloo (2013) ask "whether it is appropriate for students to have an awareness of the labels attached to them? Are there some labels that should be prohibited? As students become more aware of the implications of such labeling, the opportunity to opt out or to actively misrepresent certain characteristics to avoid labeling can diminish the validity of the remaining data set" (p. 1516). Flowing from this, these authors

propose that “In stark contrast to seeing students as producers and sources of data, learning analytics should engage students as collaborators and not as mere recipients of interventions and services” (p. 1519). The authors claim that “Students are not simply recipients of services or customers paying for an education. They are and should be active agents in determining the scope and purpose of data harvested from them and under what conditions (e.g., de-identification)” (p. 1521). (Also see West et al., 2016a, 2016b; Christopoulos, et al., 2021).

Evidence of the data interests of specific groups such as indigenous peoples, LGBTQIA+, blacks, Hispanics, females, etc.

Only two of the 11 LA frameworks specifically mention the data interests of individuals or groups that may face discrimination due to a particular characteristic or combination of identifiers, e.g. indigenous peoples, LGBTQIA+, blacks, Hispanics females, etc.

Greller and Drachsler (2012) state that.

“It is important to remind stakeholders of LA processes that data can be interpreted in many ways and lead to very different consequent actions. To give a drastic example, imagine being confronted with the insight that children from an immigrant background show reading difficulties, backed by supportive data analysis. This may lead to a wide-ranging variety of responses, from developing extracurricular support mechanisms, to segregated classes, up to bluntly racist abuse of various kinds” (p. 51).

The only other consideration of the specific concerns of such individuals or groups is the LA framework of Hernandez-Leo et al. (2019) who foreground the needs of “non-native speakers” (p. 3).

From learning analytics to institutional learning ecosystems: pointers for consideration

In this article we were concerned with understanding the extent to which selected frameworks for the implementation of LA understand LA as an ecosystem as part of a larger data ecology where different data interests compete. As data interests form the basis of data ecosystems and data ecologies, we were also curious as to the extent to which student data interests are core to LA as a data ecosystem and a broader data ecology.

Our analysis has shown that most of the LA frameworks reviewed recognise not only the different actors, but also the relations between them. What is less clear is the extent to which the frameworks take cognizance of the interdependencies between actors, and links to other institutional data ecosystems such as the Student Information System, Library services, and various student support initiatives. While all the frameworks recognise ‘data as resource’ and ‘data as regulator’, they do not recognise the impact of the interdependencies, different data standards in different institutional ecosystems, and the different vested interests in institutional data ecosystems. If ‘data is power’, then there is good reason to believe that the sponsors and custodians of these data ecosystems will be very cognizant of the inherent power

of their data interests and control of (some of) the value chain in institutional data ecosystems.

In concluding our reflection on the analysis and findings, we would like to point out two potential major concerns in institutional LA ecosystems, namely the mis-recognition of students' data interests (including the interest of specific groups constituted on the basis of gender, race, citizenship status, etc.), and a lack of recognition of how LA ecosystems (increasingly) form part of larger data ecologies with different data interests, whether political or commercial.

Students are seen in most of the analysed frameworks as data producers, and not as key stakeholders with their own data interests. While there is mention of consent, and transparency, there is little evidence to suggest that LA as discourse and practice recognizes that institutions do not own student data but should act as caretakers of student data while students retain data sovereignty.

Only one LA framework (Slade & Prinsloo, 2013) pointed to LA as an ecosystem being part of a larger data ecology. While some frameworks discuss the importing and exporting of data from and to other ecosystems, these ecosystems are mostly institutional. There are, however, specific concerns regarding ethics and privacy when institutional data ecosystems become part of external data ecosystems such as found in MOOCs, the increasing emphasis on multimodal data and the sharing of student data with platform providers (e.g., Khalil, et al., 2018; Komljenovic, 2021; Slade, et al., 2019).

Conclusions

Since the emergence of learning analytics in 2011, understanding and improving student learning has been central to the institutionalization of LA. There is evidence that LA can improve course design, increase the effectiveness of pedagogical strategies, allow institutions to identify students-at-risk and offer personalized support. However, there is the danger that LA is presented as a simple panacea which underestimates how it also depends on *more than* the collection and analysis of student data. Evidence shows that student success should be understood in the nexus of often mutually constitutive factors and relations between students (their demographics, prior learning experiences, and efficacy), academic and institutional cultures, strategies and macro-societal factors. Understanding student success *ecosystemically* therefore requires us to see LA as a data ecosystem requires acknowledgement and intentional nurturing of the relationships, interdependencies and data flows between a variety of stakeholders and data interests.

The analysis has shown that most of the frameworks analysed here do acknowledge LA as part of *institutional* ecosystems, and to a lesser extent, as part of intra-institutional ecosystems. There is, however, a lack of understanding of LA as part of an increasingly commercial data ecology, directly impacting on students' privacy and their right to data sovereignty. Student data is much more than a resource to regulate learning and forms an integral part of the asymmetrical power relationship between institutions, commercial interests and students. We must consider student

data interests as an integral part of LA as an ecosystem where students are not data subjects, but equal partners with rights to data sovereignty.

Finally, throughout this article we argued that understanding learning analytics as a data ecosystem *within a larger data ecology* has implications for data stewardship, student privacy, and ethics. We cannot ignore, as illustrated in this article, the different data interests of a range of stakeholders. While it falls outside of the scope of this article to consider and map the ethical implications in learning analytics as part of a broader data ecology, this analysis provided some glimpses such as our discussion of ‘data as risk’ (Greller & Drachsler, 2012) referring to the risks inherent in predictive analytics, the commercial interests in monetizing student data and the ‘flattening’ of the student experience to what we can measure and quantify. While ethics were discussed in most of the frameworks used in this analysis (Table 3), there is lack of critical consideration of how current understanding of the ethics in learning analytics will change if we would consider learning analytics to be part of a broader data ecology. The analysis presented in this article does, however, suggest the need for a thorough investigation of the ethical implications of learning analytics as part of a broader data ecology, and specifically, institutions’ moral and legal obligations to have students’ best interests at heart.

Limitations

We acknowledge the limitations in using the research by Khalil et al. (2022) in their analysis of LA frameworks as the baseline for this study as well as acceptance of the criteria they used in identifying 11 frameworks as the most comprehensive. In this paper we differentiate between data ecologies and data ecosystems but both these concepts are relatively new and are often used interchangeably.

We also acknowledge that interpretation is integral to a deductive content analysis and as such, we recognize the inherent subjectivity inherent in the interpretation and analysis. In mitigating this subjectivity, the researchers were as transparent as possible regarding their interpretations and the coding that followed from the interpretation. We believe that other researchers using the coding heuristic would replicate, to a large extent, the main findings of this research.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital).

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Archer, E., & Prinsloo, P. (2020). Speaking the unspoken in learning analytics: Troubling the defaults. *Assessment and Evaluation in Higher Education*, 45(6), 888–900.
- Assarroudi, A., Heshmati Nabavi, F., Armat, M. R., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: The description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, 23(1), 42–55.
- Baker, K. S., & Bowker, G. C. (2007). Information ecology: Open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems*, 29, 127–144. <https://doi.org/10.1007/s10844-006-0035-7>
- Beer, D. (2019). *The data gaze*. Sage.
- Blok, A., Marquet, C., Courmont, A., Minor, K., Young, M., Hoyng, R., & Nold, C. (2017). Data platforms and cities. *Tecnoscienza. Italian Journal of Science & Technology Studies*, 8(2), 183–190.
- Booth, M. (2012). Learning analytics: The new black. EDUCAUSEreview. Retrieved from <https://er.educause.edu/articles/2012/7/learning-analytics-the-new-black>.
- Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, 33(2), 365–412.
- Boyd, C. (2022). Data as assemblage. *Journal of Documentation*. <https://doi.org/10.1108/JD-08-2021-0159>
- Breaux, T. D., & Lotrionte, C. B. (2011). Towards a privacy management framework for distributed cybersecurity in the new data ecology. In I. Soibelman (Ed.), *2011 IEEE International Conference on Technologies for Homeland Security (HST)*. IEEE.
- Broughan, C., & Prinsloo, P. (2020). (Re) centring students in learning analytics: In conversation with Paulo Freire. *Assessment and Evaluation in Higher Education*, 45(4), 617–628.
- Chatti, M. A., & Muslim, A. (2019). The PERLA framework: Blending personalization and learning analytics. *International Review of Research in Open and Distributed Learning*, 20(1), 243–261. <https://doi.org/10.19173/irrodl.v20i1.3936>
- Christopoulos, A., Mystakidis, S., Pellas, N., & Laakso, M. J. (2021). ARLEAN: An augmented reality learning analytics ethical framework. *Computers*, 10(8), 92. <https://doi.org/10.3390/computers10080092>
- Christopoulos, A., Pellas, N., & Laakso, M. J. (2020). A learning analytics theoretical framework for STEM education virtual reality applications. *Education Sciences*, 10(11), 317. <https://doi.org/10.3390/educsci10110317>
- Cormack, A. N. (2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, 3(1), 91–106.
- Cui, Y., Kara, S., & Chan, K. C. (2020). Manufacturing big data ecosystem: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, 62, 101861. <https://doi.org/10.1016/j.rcim.2019.101861>
- D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.
- Davenport, T. H., & Prusak, L. (1997). *Information ecology: Mastering the information and knowledge environment*. Oxford University Press.
- Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2021). Stakeholder participation in AI: Beyond add diverse stakeholders and stir. arXiv preprint [arXiv:2111.01122](https://arxiv.org/abs/2111.01122). <https://doi.org/10.48550/arXiv.2111.01122>.
- Denis, J. (2016, August). Data and its invisible work. In Levin, L., & Medina, L.R., (eds), *Science+Technology by Other Means-4S/EASST Conference* (pp. 1–7). Retrieved from <https://halshs.archives-ouvertes.fr/halshs-01364311/>.
- Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International*, 13(3), 313–321.
- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*. <https://doi.org/10.1177/2158244014522633>

- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115.
- Falcão, T. P., Mello, R. F., Rodrigues, R. L., Diniz, J. R. B., Tsai, Y. S., & Gašević, D. (2020). Perceptions and expectations about learning analytics from a Brazilian higher education institution. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the 10th International Conference on Learning Analytics and Knowledge*. Association for Computing Machinery.
- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317.
- Figuerras, C., Verhagen, H., & Cerratto Pargman, T. (2021). Trustworthy AI for the People? In M. Fourcade & B. Kuipers (Eds.), *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. Wiley.
- Gašević, D. (2018). Include us all! Directions for adoption of learning analytics in the global south. In *Learning Analytics for the Global South*, (pp 1–22). Retrieved from <http://dl4d.org/wp-content/uploads/2018/03/Learning-Analytics-Main-Paper-2.pdf>.
- Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. In Uylenbroek, J.J.M., & Sol, H., (eds), *Proceedings of the 54th Hawaii International Conference on System Sciences* (p. 6113). Retrieved from <https://scholarspace.manoa.hawaii.edu/bitstream/10125/71359/0597.pdf>.
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology and Society*, 15(3), 42–57.
- Guggenberger, T. M., Möller, F., Haarhaus, T., Gür, I., & Otto, B. (2020). Ecosystem types in information systems. In *ECIS*. Retrieved from https://aisel.aisnet.org/ecis2020_rp/45/.
- Hasselbalch, G. (2021). A framework for a data interest analysis of artificial intelligence. In *First Monday*. Retrieved from <https://firstmonday.org/ojs/index.php/fm/article/view/11091>.
- Hernández-Leo, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50(1), 139–152.
- Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y. S., Muñoz-Merino, P. J., & Pérez-Sanagustín, M. (2020). Identifying needs for learning analytics adoption in Latin American universities: A mixed-methods approach. *The Internet and Higher Education*, 45, 100726. <https://doi.org/10.1016/j.iheduc.2020.100726>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 2053951716674238
- Kazanidis, I., Pellas, N., & Christopoulos, A. (2021). A learning analytics conceptual framework for augmented reality-supported educational case studies. *Multimodal Technologies and Interaction*, 5(3), 9. <https://doi.org/10.3390/mti5030009>
- Khalil, M., & Ebner, M. (2015). Learning analytics: Principles and constraints. In *EdMedia+ Innovate Learning*, (pp. 1789–1799). AACE.
- Khalil, M., Prinsloo, P., & Slade, S. (2018). User consent in MOOCs—micro, meso, and macro perspectives. *International Review of Research in Open and Distributed Learning*. <https://doi.org/10.19173/irrodl.v19i5.3908>
- Khalil, M., Prinsloo, P., & Slade, S. (2022). A comparison of learning analytics frameworks: A systematic review. In A. Wise (Ed.), *LAK22: 12th international learning analytics and knowledge conference*. Association for Computing Machinery.
- Kitchin, R., & Lauriault, T. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. Retrieved from <https://ssrn.com/abstract=2474112>.
- Kitchin, R. (2014). *The data revolution*. SAGE.
- Kitto, K., Cross, S., Waters, Z., & Lupton, M. (2015). Learning analytics beyond the LMS: the connected learning analytics toolkit. In J. Baron, G. Lynch, & M. Maziarz (Eds.), *Proceedings of the 5th International Conference on Learning Analytics and Knowledge*. Association for Computing Machinery.
- Komljenovic, J. (2021). The rise of education renters: Digital platforms, digital data and rents. *Learning, Media and Technology*, 46(3), 320–332.
- Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. *Journal of Learning Analytics*, 7(3), 98–117.

- Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—a literature review. In A. Peña-Ayala (Ed.), *Learning analytics: fundamentals, applications, and trends* (pp. 1–23). Springer.
- Liao, C. H., & Wu, J. Y. (2022). Deploying multimodal learning analytics models to explore the impact of digital distraction and peer learning on student performance. *Computers and Education*, 190, 104599. <https://doi.org/10.1016/j.compedu.2022.104599>
- Madaio, M., Blodgett, S. L., Mayfield, E., & Dixon-Román, E. (2021). Confronting Structural Inequities in AI for Education. arXiv preprint [arXiv:2105.08847](https://arxiv.org/abs/2105.08847). <https://doi.org/10.48550/arXiv.2105.08847>
- Nadim, T. (2016). Data labours: How the sequence databases GenBank and EMBL-Bank make data. *Science as Culture*, 25(June), 1–24. <https://doi.org/10.1080/09505431.2016.1189894>
- Oliveira, M. I. S., Barros Lima, G. D. F., & Lóscio, B. F. (2019). Investigations into data ecosystems: a systematic mapping study. *Knowledge and Information*, 61(2), 589–630.
- Oliveira, M. I. S., Oliveira, L. E. R., Batista, M. G. R., & Lóscio, B. F. (2018). Towards a meta-model for data ecosystems. In M. Janssen & S. A. Chun (Eds.), *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. Association for Computing Machinery.
- Piety, P. J. (2019). Components, infrastructures, and capacity: The quest for the impact of actionable data use on P–20 educator practice. *Review of Research in Education*, 43(1), 394–421.
- Prasad, D., Totaram, R., & Usagawa, T. (2016). A framework for open textbooks analytics system. *Tech-Trends*, 60(4), 344–349.
- Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. In Wise, A., Winne, P. H., & Lynch, G. (eds), *Proceedings of the 7th international learning analytics and knowledge conference* (pp. 46–55).
- Prinsloo, P. (2017). Fleeing from Frankenstein's monster and meeting Kafka on the way: Algorithmic decision-making in higher education. *E-Learning and Digital Media*, 14(3), 138–163.
- Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.
- Prinsloo, P., Slade, S., & Khalil, M. (2019). Student data privacy in MOOCs: A sentiment analysis. *Distance Education*, 40(3), 395–413.
- Prinsloo, P., Slade, S., & Khalil, M. (2020). Implementing learning analytics: An ecosystemic perspective (translated by Junhong Xiao). *Distance Education in China*, 4, 1–11.
- Queenborough, S. A., Cooke, I. R., & Schildhauer, M. P. (2010). Do we need an EcoBank? The ecology of data-sharing. *Bulletin of the British Ecological Society*, 41(3), 33–35.
- Rienties, B., Cross, S., Marsh, V., & Ullmann, T. (2017). Making sense of learner and learning big data: Reviewing 5 years of data wrangling at the open University UK. *Open Learning: The Journal of Open, Distance and e-Learning*, 32(3), 279–293.
- Sauvé, K., & Houben, S. (2021, March). Towards an ecology of interconnected data devices. In *CHI'21 Workshop Human-Data Interaction through Design*. Retrieved from https://eprints.lancs.ac.uk/id/eprint/156375/1/CHI21_HDI_workshop_submission_KS_SH.pdf.
- Shin, D. H. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy*, 40(9), 837–854.
- Skinner-Thompson, S. (2020). *Privacy at the margins*. Cambridge University Press.
- Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In Hsiao, S., & Cunningham, J. (eds), *Proceedings of the 9th International Conference on learning analytics and knowledge* (pp. 235–244).
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.
- Slota, S. C., Fleischmann, K. R., Greenberg, S., Verma, N., Cummings, B., Li, L., & Shenefiel, C. (2021). Many hands make many fingers to point: Challenges in creating accountable AI. *AI and SOCIETY*, 5, 1–13.
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1), 111–134. <https://doi.org/10.1287/isre.7.1.111>
- Star, S. L., & Strauss, A. (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1), 9–30.
- Steedman, R., Kennedy, H., & Jones, R. (2020). Complex ecologies of trust in data practices and data-driven systems. *Information, Communication & Society*, 23(6), 817–832.

- Swenson, J. (2014). Establishing an ethical literacy for learning analytics. In *Proceedings of the Learning Analytics & Knowledge conference 2014 LAK'14* (pp. 246–250). New York: ACM Press.
- Tan, J. P. L., & Koh, E. (2017). Situating learning analytics pedagogically: Towards an ecological lens. *Learning Research and Practice*, 3(1), 1–11. <https://doi.org/10.1080/23735082.2017.1305661>
- Tang, X., Mendieta, E., & Litzinger, T. A. (2022). Developing an online data ethics module informed by an ecology of data perspective. *Science and Engineering Ethics*, 28(2), 1–22.
- Tsai, Y. S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics*, 5(3), 5–20.
- Umbrello, S., & Van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, 1(3), 283–296.
- Valle, N., Antonenko, P., Dawson, K., & Huggins-Manley, A. C. (2021). Staying on target: A systematic literature review on learner-facing learning analytics dashboards. *British Journal of Educational Technology*, 52(4), 1724–1748.
- Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous data sovereignty in the era of big data and open data. *Australian Journal of Social Issues*, 56(2), 143–156.
- Watters, A. (2013). *Student data as the new oil: MOOCs, metaphor and money*. HackEducation.
- West, D., Heath, D., & Huijser, H. (2016a). Let's talk learning analytics: A framework for implementation in relation to student retention. *Online Learning Journal*, 20(2), 1–21.
- West, D., Huijser, H., & Heath, D. (2016b). Putting an ethical lens on learning analytics. *Educational Technology Research and Development*, 64(5), 903–922.
- Wong, B. T. M., & Li, K. C. (2019). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, 2, 1–22.
- Wu, J. Y. (2021). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement. *Computers and Education*, 163, 104066. <https://doi.org/10.1016/j.compedu.2020.104066>
- Zhang, L., Downs, R. R., Li, J., Wen, L., & Li, C. (2021). A review of open research data policies and practices in China. *Data Science Journal*, 20(1), 534.
- Zotou, M., Tambouris, E., & Tarabanis, K. (2020). Data-driven problem-based learning: Enhancing problem-based learning with learning analytics. *Educational Technology Research and Development*, 68(6), 3393–3424.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paul Prinsloo is a Research Professor in Open and Distance Learning (ODL) in the Department of Business Management, College of Economic and Management Sciences, University of South Africa (Unisa). He is a Visiting Professor at the Carl von Ossietzky University of Oldenburg, Germany, a Research Associate for Contact North I Contact Nord (Canada), a Fellow of the European Distance and E-Learning Network (EDEN), member of the Executive Committee for the Society of Learning Analytics Research (SoLAR) and serves on several editorial boards. In the South African context Paul has a B3 research rating confirming his considerable international reputation for the high quality and impact of his research outputs. Paul is an internationally recognised speaker, scholar and researcher and has published numerous articles in the fields of teaching and learning, student success in distance education contexts, the ethical collection, analysis and use of student data in learning analytics, and digital identities. He blogs at <https://opendistanceteachingandlearning.wordpress.com/> and his Twitter/Mastodon alias is @14prinsp.

Mohammad Khalil is a senior researcher and lecturer in learning analytics at the Centre for the Science of Learning & Technology (SLATE) within the Faculty of Psychology at the University of Bergen in Norway. Khalil, as an Erasmus Mundus granted scholar, has a Ph.D. with distinction from Graz University of Technology in Learning Analytics in Massive Open Online Courses (MOOCs). He worked as a Postdoc at Delft University of Technology for 2 years in Technology Enhanced Learning. Khalil leads several international projects in learning analytics, including Erasmus+ and PederSather. Khalil is currently an associate editor of the International Journal of Emerging Technologies in Learning (iJET).

Besides learning analytics, his key research interests include AI in Education, Educational Data Mining, and Privacy and Ethics.

Sharon Slade worked in open, distance education for almost 20 years, as a senior lecturer at the Open University in the UK. She led the work on development of the University's Policy on the ethical use of student data for learning analytics, arguably the first of its kind in Higher Education worldwide. She has since contributed to further framework developments, notably with Jisc, Stanford University and Ithaka S+R, New America and the International Council for Open and Distance Education. Sharon was an academic lead for learning analytics projects within the Open University, leading work around ethical uses of student data, operationalisation of predictive analytics and approaches aiming to improve retention and progression. Keynotes and publications include papers around student consent, the obligation to act on what is known, examining the concept of educational triage and broader issues around an ethics of care. She now works on data insight at Earth Trust, an educational and environmental charity near Oxford.