**Title**

# Reconsidering Data in Learning Analytics: Opportunities for Critical Research Using a Documentation Studies Framework

**Authorship**

Kyle M. L. Jones
Indiana University–Indianapolis (IUPUI)
School of Informatics and Computing, Department of Library and Information Science
kmlj@iupui.edu

Chase McCoy
Indiana University–Bloomington
School of Library and Information Studies, The iSchool salo@wisc.edu

**Correspondence**

All correspondence should be directed to Dr. Jones via e-mail at kmlj@iupui.edu.

---

This is the author's manuscript of the article published in final edited form as:

Jones, K. M. L., & McCoy, C. (2019). Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework. Learning, Media and Technology, 44(1), 52–63. https://doi.org/10.1080/17439884.2018.1556216

**Abstract**

In this article, we argue that the contributions of documentation studies can provide a useful framework for analyzing the datafication of students due to emerging learning analytics practices. Specifically, the concepts of individuals being “made into” data and how that data is “considered as” can help to frame vital questions concerning the use of student data in learning analytics. More specifically, approaches informed by documentation studies will enable researchers to address the sociotechnical processes underlying how students are constructed into data, and ways data about students are considered and understood. We draw on these concepts to identify and describe three areas for future research in LA. With the description of each area, we provide a brief analysis of current practices in American higher education, highlighting how documentation studies enables deeper analytical digging.

Keywords*:* Learning analytics, educational data mining, documentation studies, critical data studies

## 1. Introduction

Data are not objective entities, nor can they fully represent an individual's identity, interests, and values. However, in the era of Big Data, these qualities are often attributed to data (Frické 2015). Data create flawed representations of people based on their interactions with technologies, effectively reducing them into fragmentary representations. Moreover, flawed data are framed by social and political interests embedded into data structures by database designers, administrators, and stakeholders who seek to analyze and mine value from data. Given the resources and time being put into Big Data practices and the consequences they have for human life, this is problematic and worth targeted inquiry, especially in the context of education.

The datafication of student life has emerged as a result of learning analytics (LA). LA is the use of data mining and analytic methods to investigate and better understand how students learn, as well as to optimize institutions of learning, their environments, and administrative processes (Siemens 2012; van Barneveld, Arnold, and Campbell 2012). Outspoken advocates of LA argue that only by redirecting data flows to capture student life in granular detail will educational data mining reach its potential for improving learning outcomes among other things (Siemens 2013).

We argue that contributions from documentation studies can provide a useful framework for analyzing the datafication of students with LA. Specifically, the concepts of individuals being "made into" data and how those data are "considered as" (Buckland 2013; Buckland 2014) help frame vital questions concerning student data practices. We begin this article by describing documentation studies. Following, we argue that data should be considered as a type of document. Viewing data as documents and applying the concepts of "made into" and "considered as" to data creation and use, we argue, provides scholars and practitioners a useful lens for understanding student data and critiquing LA practices. After unpacking these approaches, we pull concepts from documentation studies to identify and describe three areas for future research. With the description of each area, we provide a brief analysis of current practices in American higher education, highlighting how documentation studies enables deeper analytical digging.[1]

## 2. Documentation Studies and Data

### *2.1. Documentation studies*

Documentation studies challenge the traditional view that documents and documentation refer to text-based artifacts only, as well assumptions that documents

[1] While our focus is on American higher education, the framework we develop applies nicely to all levels of education where data mining and analytic practices are becoming common. For example, see Crooks (2017) whose work focuses on urban charter schools. Additionally, the framework would apply nicely to non-American contexts, especially the United Kingdom and Australia where higher education systems are facing neoliberal pressures like the United States. However, we recognize that any critique of a particular learning analytics practices needs to take into consideration its geographic location and related contextual factors, including relevant laws, societal norms, and community values (see Prinsloo and Slade 2017).

exist outside of interpretation and the contexts that give them life. As Buckland (1997, 805, emphasis in original) writes, “there was (and is) no theoretical reason why documentation should be limited to *texts*, let alone *printed* texts.” Early documentation theorists—like Otlet (1934), Briet (1951), and Schürmeyer (1935)—pivoted the conversation away from texts towards objects. “One understands as a document,” began Schürmeyer (1935, 537), “any material basis for extending our knowledge which is available for study or comparison.” Effectively, this new view established that any object is a document insofar that it has materiality, is intended to inform others as evidence of something, needs to be processed (i.e., documented), and others besides the documentarian view it as a document (Briet 1951).

Briet (1951) famously asked, “is an antelope a document?” The answer: it depends. If the antelope is living in the wild, no. If the antelope is on display at a zoo, yes. The difference being that in a zoo, the antelope is “an object of study, it has been made into a document. It has become physical evidence being used by those who study it” (Buckland 1997, 806). Combined, Briet’s insight and Buckland’s analysis support Lund’s (2009, 424) description of a document as a “physical, social, and mental phenomenon.” That is to say that documents come into being through technical and technological means, they are shaped by a mélange of social factors, and creators and users develop a mental (cognitive, intellectual) relationship with documents.

Early documentation studies more often than not viewed documents through a technical lens (Buckland 2013; Lund 2009), one that focused on resource management processes, especially preservation, organization, representation, and dissemination practices (Buckland 1991; Buckland 1997). As a counterweight, modern documentation studies—such as those by Latour and Woolgar (1979) and Brown and Duguid (1996), among others—have focused on social contexts and political factors, arguing that purported evidence and facts within a document depend on the perceptions of those who both construct and use the document. Such an approach has broadly impacted the study of documents by 1) building from Briet’s liberal definition of a document and 2) expanding the range of problems associated with their creation and use.

### *2.2. Data as documents*

Data are documents, too, as Buckland (2000, para. 15, emphasis added) writes: “any ‘thing’ that is regarded as signifying [is a document]: Books, records, *data*, speech, signs, symbolic objects.” And considering data as documents through a modern documentation studies perspective, provides a lens to explore the conditions that lead to the construction of data and, in turn, how they are perceived and understood. Data are commonly viewed as digital objects derived through the reduction and extraction of social and cultural phenomena from observations, computations, experiments, and record keeping (Kitchin 2014). In this sense, data are “anything recorded in a relational database in a semantically and pragmatically sound way” (Frické 2015, 652). Data are then able to be searched, aggregated, cross-referenced, and examined (boyd and Crawford 2012). However, this technocentric view too often considers data as objective entities, while ignoring the social, cultural, and political aspects of data and the consequences of their use (Bowker 2013; boyd and Crawford 2012; Kitchin and Lauriault 2014).

This paper draws on two views of documents to 1) conceptualize student data as documents and 2) to frame areas for critical research in learning analytics (LA). From a modern documentation studies perspective, the functional (or instrumental) view and the semiotic view enable new ways to study student data (Buckland 2013; Buckland 2014). In the functional view, *anything* can be considered as a document so long as it is "held up as constituting evidence of some sort" (Buckland 2013, 7). This view is concerned with the process through which things are "made into" documents. The semiotic view defines documents as anything that is evidence of something, regardless of a creator's intentions and its format (e.g., textual, non-textual), and explores how documents are "considered as" (Buckland 2013).

Individuals are "made into" plausibly infinite data representations that constitute evidence of something in relation to themselves and others through their technical interactions with information systems (Day 2015; Floridi 2012). And even though the technical avenues by which a person is "made into" data have limitations, the incalculable amount of relations that can be made among datasets continues to add information to and redefine how humans are "made into" data.

Through the semiotic view, data are "considered as" being informative in that the perception of data and the meaning associated with them depends on their creator(s), users, and the context in which they are situated. As Buckland states, the phenomenological aspect of documents (including data) refers to "objects perceived as something, the status of being a document is not inherent (essential) but attributed to an object" (Buckland 2013, 4). The individual that is "made into" numerous and ever growing number of data are, therefore, embedded in social and political contexts that underlie how data are "considered as."

### *2.3. Documenting students in data*

Documents about student life no longer consist of paper files locked away in the registrar's office. In fact, student life is increasingly documented in digital dossiers consisting of academic, social, behavioral, and even emotional markers. And these dossiers encompass the data against which institutions run their descriptive and predictive analytics to develop insights, which feed back into the dossiers to further inform an array of actors about a given student (Alblawi and Alhamed 2017).

Analyzing the documentation of students in data adds a new perspective to critique current and future educational data mining technologies, practices, and sociopolitical interests. Furthermore, it complements and extends the usefulness of a robust toolbox of applicable theories and approaches, including: infrastructure studies (see Williamson 2018), critical data studies (see Illiadis and Russo 2016; Selwyn 2015), sociology of quantification (Espeland and Stevens 2008; Hardy 2015), data and information ethics (see Floridi and Taddeo 2016; Rubel and Jones 2016), and others. Asking how students are "made into" and "considered as" data creates a new framework for addressing important critical questions. These questions along with relevant sub-questions in the table below can serve as a guide for approaching student data from a documentation studies perspective.

| **Made Into** | **Considered As** |
|---|---|
| Material | Purposes |
| • Are the data a document or exist as part of a corpus, and how do we know?<br>• What is the data source?<br>• In what form(s) do the data exist? | • What evidence do these data provide, and can we verify their veracity?<br>• What end(s) do these data serve?<br>• What role(s) do these data play in a sociotechnical system? |
| Processes | Values |
| • What visible and invisible documentation processes created the data?<br>• Were the documentation processes purposefully created or ad hoc, were they responsive or forward-thinking?<br>• How do uses of the data support other practices? | • What social value do the data have?<br>• What cultural value do the data have?<br>• What political value do the data have? |
| Human and Non-Human Actors | Issues |
| • Who directly and indirectly created the data?<br>• Who directly and indirectly uses the data?<br>• Who makes secondary uses of the data?<br>• In what ways do algorithms and artificial intelligence affect the data's characteristics? | • Is the data morally justifiable?<br>• Can the data be ethically accessed and used?<br>• How do the data create or perpetuate unjust systems?<br>• How does the existence/use of the data create harms? |

Table 1. Documentation studies framework.

When we ask questions about making students into data, we begin to unpack the material aspects of data, the processes leading to that materiality, and the actors associated (or not) with those processes. When we ask questions about how students are considered as data, we uncover the purpose or role of the data in a social system, along with how data are used as means to various ends—and the potential and actual negative effects data surface. It is with this framework that we will address particular aspects of LA and the documentation of students in data in the following sections.

## 3. The Fallibility of Data

### *3.1. "Cooking" data*

In the current data climate, the breadth and scale of data being collected has led to the belief that "digital traces" are, in fact, comprehensive pictures of human life (boyd and Crawford 2012). These traces, argue Lazer et al. (2009, 721), hold "the potential to transform our understanding of our lives, organizations, and societies" when aggregated, analyzed, and acted upon. Their argument assumes, however, that such traces are full, infallible representations of human life, and it extends the dangerous position that data "exist prior to interpretation or argument" (Kitchin and Lauriault 2014, 3).

Data do not simply exist, rather they are generated (Manovich 2011). Social influences shape data artifacts and ensuing data-based decisions are subject to interpretive human judgement (Ekbia et al. 2015). Bowker (2013) refers to the process of data generation and shaping as data cooking. He argues that viewing data as "raw" is an oxymoron, and that all data should be considered "cooked" and "seasoned" with the values of those doing the so-called cooking.

In summary, taking for granted data as objective documents and treating them as pure evidence of facts does not motivate some to consider the "made into" and "considered as" characteristics of data-as-documents (boyd and Crawford 2012). Specifically, such an approach to data 1) ignores the politics and values that drive the documentation creation processes and 2) fails to underscore the sociotechnical mélange in which data exist. In the following subsection we highlight the emergence of advising analytics, pointing out as we write how the interests of higher education administrators are intertwined with making students into data.

### *3.2. Examples in practice: Advising analytics*

To increase advising capacities and better inform students of their educational options, LA proponents have built descriptive and predictive advising technologies. When combined with analytic affordances, such tools may be able to help advisors diagnose students issues quicker and with exact information (Aguilar, Lonn, and Teasley 2014). Advisors can also use advising analytics to personalize student resources based on their educational and professional profiles (Kraft-Terry and Kau 2016). Three universities—Austin Peay State, Arizona State, and Georgia State—are known for their work with advising tools supported with analytics.

Austin Peay State University uses a recommender system, which identifies courses students should take by, first, determining if they are core and program-specific courses required for graduation and, second, if students are expected to achieve academic success in a given course (Denley 2013). Arizona State University's advising tools calculate a student's college readiness and predicted retention score (Phillips 2013).

Georgia State University "analyzed two and a half million grades earned by students over ten years to create a list of factors that predict which students are less likely to graduate," which the institution's advising system combines with "past student performance data to predict how well each current student will do in all majors and most courses" (Ekowo and Palmar 2016, 3).

Proponents of advising analytics argue that there are "no more excuses" (Crow 2012, 21) not to analyze student data and implement predictive measures, especially for advising, but they often fail to carefully balance the potential harms with the benefits they proselytize. If administrators and other LA proponents continue to make students into and consider students as irrefutable, empirical, factual data, then it is plausible that predictive scores may consequently affect student autonomy (Jones 2018; Rubel and Jones 2016). A prediction that a student will do poorly in a given course could motivate an advisor to suggest easier courses for which a student has a higher rate of predicted success, even if that course does not align with a student's professional or personal interests. At the extreme, if students are predicted to not be retained, administrators may push advisors to provide less of their time and resources, or to suggest to advisees that they pursue their degree elsewhere. This is not a hyperbolic example. Mount St. Mary University's former president advocated for using predictive analytics to "weed out students unlikely to be retained" (Johnson 2017, 1), or as he churlishly put it: "drown the bunnies… put a Glock to their heads" (Svrluga 2016). The president argued that dismissing students unlikely to be retained could improve the university's retention data, which accreditors, prospective students, and national rankings (e.g., *U.S. News and World Report*) use to make judgements about an institution.

### *3.3. Critical questions for future work on learning analytics and values*

The documentation perspective enables an approach that examines advising analytics as an incomplete representation of student retention that is shaped by institutional values, not student interests. As the example in practice above highlights, institutions often pursue advising analytics based on administrative values (e.g., financial cost recoveries). In the Georgia State University example, many have exclaimed about the system more for the nearly $5 million it saved taxpayers and less about how and to what degree students benefited (see University Innovation Alliance n.d.). Using concepts from documentation studies, researchers can examine the consequences of students being "made into" data when one institutional perspective is favored over others, such as the administrative perspective over the advisor or student perspectives, in the "considered as" process. We can generate more critical questions using the "made into" lens regarding the creation of data and the people who are both involved and excluded in the process. The "considered as" concept then enables us to question the purpose and goals of LA, as well as address the emerging effects of these systems in practice. Some questions include:

| Made Into | Considered As |
|---|---|

| | |
|---|---|
| • What is the source of the student data used in the predictive models?<br>• What is the process by which data are determined to be included in analytic models?<br>• Who makes decisions regarding analytic model development? | • How do we know that the predictive scores measure what they purport to?<br>• Whose interests do these analytics serve?<br>• Are the actions the analytics inform justifiable, and what heuristics justify such actions? |

Table 2. Documentation studies questions on data fallibility.

## 4. Apophenia and Data

### *4.1. Seeing patterns where none exist*

Much of the attention in learning analytics (LA) research has focused on advancing analytic strategies. But Bowker (2013) cautions that those who deploy data science methods too often treat data objectively and fail to recognize the imperfections of statistical models. Critics argue that these actions lead to *apophenia*, which is "seeing patterns where none actually exist" (boyd and Crawford 2012, 668). This is certainly problematic since one of the potentials of Big Data methods is the ability to spot patterns that were previously unidentifiable.

One of the causes of apophenic behavior is due to the fact that too often computer scientists identify and analyze data without the aid of individuals for whom the studied phenomena would be familiar (Bowker 2013; Ruppert 2013). Without the proper contextual lens through which to look at data patterns, the patterns themselves stand to become meaningless, if not harmful. If this is the case, how do researchers go about analyzing individuals through data while keeping in mind the potential creep of apophenia into their results? It depends on whose point of view is privileged when data patterns are considered as evidence. The following subsection presents a now infamous LA initiative, where those conducting an LA analysis of student retention failed to consult those who could have curbed the apophenia that emerged.

### *4.2. Examples in practice: Questioning Purdue University's student retention claims*

In 2007, Kim Arnold and Matthew Pistilli, two of Purdue's assessment researchers at the time, developed and patented the predictive application Course Signals. Using a proprietary algorithm, the tool compares students' earned grades in a course, interactions with course materials within a learning management system, and past academic history and personal information to predict their risk of earning an unsuccessful grade. Pistilli and Arnold (2010, 23) explained that the algorithm "was built from the ground up using empirical data at every stage to ensure the most predictive student success algorithm." The popularity around Course Signals stems from its substantial results.

Arnold and Pistilli (2012) reported that "students who began at Purdue in fall 2007, 2008, or 2009 and participated in at least one Course Signals course [were] retained at rates significantly higher than their peers who had no Course Signals classes but who started at Purdue during the same semester" (268). For students who had taken two or more Course Signals-enabled courses, five-year graduation data showed that students graduated at a rate that was *24.36 percent higher* than students who did not take a course with Course Signals technology (Tally 2013). Desire2Learn and Blackboard, two prominent learning management system developers, also took note of these findings and modeled their own predictive tools off of the work done by Arnold and Pistilli (Feldstein 2013a).

A single-digit increase in graduation rates is notable, but a double-digit increase is nearly unheard of—this fact alone raised suspicions. In particular, Caulfield (2013a) put forth serious criticisms. Pistilli claimed that taking Course Signals-enabled courses was *the* explanation behind the dramatic increases in retention, but Caulfield argued otherwise. He claimed that a reverse-causality effect was at play, which Essa (2013) was able to simulate. Essa explained:

> [T]he direct causality attributed to Course Signals is erroneous. In fact, the causation is the reverse of what is claimed. Students who take Course Signals courses are not more likely to graduate than non-Course Signals students (at least not directly and at the rates suggested), rather students who graduate are more likely to take Course Signals courses. (para. 2)

Pistilli and Arnold fell victim to apophenia: they saw patterns where they did not exist.

The desire to identify and spread the retention claims was powerful given the sociopolitical retention metrics have in institutional conversations. Their claims gained popularity and even awards (Caulfield 2013b). But at what cost? The work arguably motivated other institutions to invest significant (but unknowable) amounts of money in systems like Course Signals, which now look like sales of "snake oil" (Feldstein 2016). And Purdue's reputational gains may have precipitously dropped among those in the LA community, especially since Pistilli inadequately defended his claims on behalf of the institution (Feldstein 2013b; Straumsheim 2013).

### *4.3. Critical questions for future work on learning analytics informed by contextual factors*

Auditing student data and algorithms through collaboration with those with an intimate knowledge of statistical methods for calculating retention is paramount for curbing apophenia. And as shown in the Purdue example, those familiar with the contextual factors that shape student retention were left out of the conversation but could have been useful collaborators. LA scholars have argued for the importance of considering institutional context in order to create effective data models because doing so "speaks to an in-depth understanding of the institution and how it functions" (Prinsloo, Archer, Barner, Chetty, and van Zyl 2015, 298). To this point, Feldstein's (2016 para. 6) critique noted that the Purdue University situation would not "have happened had there been robust peer review of Purdue's work in the first place" and had different disciplines and higher education perspectives come together to evaluate the analytic processes.

When contextual factors are considered in relation to the data patterns that emerge as part of the documentation process, the "what" found in the data patterns can be further explored to find the "why" without falling victim to apophenia. We pose the following questions from a documentation studies perspective that could be used to study the consequences of similar LA practices:

| Made Into | Considered As |
|---|---|
| • Do the data analytics exist as visualizations or in tabular formats?<br>• Are the statistical modelling processes iteratively developed and vetted at each stage?<br>• Were non-data science actors consulted during the construction of the statistical models? | • Are the data analytics intelligible by those for whom they were design to be used?<br>• Will a diverse set of institutional actors see the usefulness of the data analytics in the same way?<br>• Do the data analytics accidentally target already disenfranchised student populations? |

Table 3. Documentation studies questions on apophenia.

## 5. Creating Data Doubles

### *5.1. Indexicality and data linking*

Architects of data create systems of organization and representation to the end of linking indexical data and, by extension, an index's attributes (Kitchin 2014). Effectively, indexical data enables a seemingly infinite array of data fusions, regardless of the context from which the data were born and the ends to which they were designed to serve. For proponents of data analytics, this characteristic amplifies the possible insights statistical work can bring about. For privacy advocates and learning analytics (LA) critics, indexicality raises serious concerns about the increasing use of 'data doubles' to represent student life (Selwyn 2015).

Data and surveillance practices abstract an identifiable human body and its activities "into various data flows or streams" and then reassemble "them into data doubles to be analyzed and targeted for intervention" (Haggerty and Ericson 2000; Ruckenstein 2014, 69). An individual's data double is configured, reconfigured, and reinterpreted with each recursion brought about by making new data linkages and grafting on new data attributes. As such, the documentation of life in data doubles "configure a certain representation of a person… [and] have their own social lives and materiality, quite apart from the fleshy bodies from which they are developed" (Lupton 2014, 82). We cannot assume, however, that each version of a data double is perfectly constructed.

Data may be inaccurate, and similar flows of data that inform the double may conflict, which effectively embeds anomalies into the double. Moreover, the “rhizomatic,” often opaque connection of complex assemblages of data makes it increasingly impossible for subjects to know what data their double includes and alter these digital records (Caluya 2010, 626; Tsesis 2014).

Taken together, these things highlight a need to understand not only the materiality of data doubles, especially given their role in LA, but also the decision-making criteria for including specific data in data doubles and the processes by which data doubles come to be. It is more often than not the case that data doubles are becoming a part of a student’s official educational record. Bearing in mind how consequential student records are, it is also worth the effort to critically assess how these particular types of documents are defined, by whom, and how such definitions determine student privacy rights—all these things we highlight in the following subsection.

### *5.2. Examples in practice: The Fountain Hopper and student records*

In January of 2015, a group of anonymous Stanford University students writing for the online newsletter the *Fountain Hopper* argued that the admissions process was unfair and possibly discriminatory. Students, they wrote, should have the right to review these records. To enable students to seek access, they built a record request template that exploited what they described as a “tried and tested legal loophole that guarantees you access to your confidential, internal Stanford admissions file” (Fountain Hopper 2015a, para. 2). That loophole was title 20 section 1232g in the U.S. Code, which is more commonly known as the Family Educational Rights and Privacy Act, or FERPA for short. FERPA provides students the right to inspect, review, and request amendments to personally identifiable education records maintained by their institution. Nearly 3,000 students and alumni requested access to their admissions records in a three-month period (Fountain Hopper 2015b). Within 45 days of their request, students began to access their files. They found quantitative scores of “intellectual vitality”; written evaluations by at least two admissions officers, and categories based on whether or not they were legacy students, “VIPs,” athletes, or had other characteristics related to diversity; and logs of every time they used their ID card to enter doors (Pérez-Peña 2015).

In the wake of the press the *Fountain Hopper*’s efforts received, students at other institutions attempted similar efforts to gain access to their records. In response, both Stanford and Yale changed their record retention practices. The former deleted the admissions records of students who had not requested access, and they no longer retain admission files post the admit decision (Gioia 2015); similarly, the latter expunged all admissions and career development records (Pomianowski 2015). Brown University, who was also hit with records requests, claims that “the internal admission file...is not considered part of the academic record and thus does not fall under FERPA” (Lifshits 2015, para. 9), arguing that “each institution can constitute what determines a permanent record” (para. 6).

***5.3. Critical questions for future work on learning analytics and data doubles***

A documentation studies approach homes in on the Kafkaesque nature of educational records, highlighting how such records have shifting materiality and are managed differently by institutions (the "made into" lens). Furthermore, it surfaces how institutional decisions about what to include educational records affects students' legal privacy rights (the "considered as" lens). As institutions continue to link their information systems to student identities, it will become even more unclear how these data doubles will develop and institutions will define an educational record. This is especially true as institutions enter into contracts with educational technology companies to provide services (Polonetsky and Tene 2015). But as these things continue to unfold, we can apply some of the following critical questions to unpack the "made into" and "considered as" aspects of those practices and their effects on educational records:

| Made Into | Considered As |
|---|---|
| • How do we know data is or is not part of a student record?<br>• Does a policy exist detailing what data to include into a student record?<br>• Who, exactly, can access student records within an institution and what usage rights do they have? | • What are the explanatory reasons for including certain data in a student record?<br>• How does a student record and the data it includes represent a student's academic achievement and future potential?<br>• How might leakage of data from a student record harm the reputation of a student, the institution, or specific institutional actors? |

Table 4. Documentation studies questions on data doubles.

**6. Conclusion**

Students are "made into" innumerable data representations through their interactions with technical artifacts. These data are considered to constitute some sort of evidence regarding students and their behavior. However, as has been shown, the evidence these data represent should be called into question given that data are not objective representations. Rather, they are generated, shaped, and cooked into some representational form to stand as evidence. Data are subject to manipulation, which shape how these data are "considered as" regardless of how the data had been intended to be considered. While the plasticity of data has been acknowledged in critical data studies, more attention needs to be given to the social and political shaping of data in learning analytics (LA) research with a documentation studies approach.

Using the "made into" and "considered as" views of data as a framework, we put forward three critical areas for future LA research in this article. LA research should begin to acknowledge and address the fact that data are not pre-factual; that *apophenic* behavior can lead to data misjudgments; and data doubling obfuscates students' understanding of what constitutes their education records and how to access data about themselves. The critical areas we covered herein differ from previous critical research on LA (see Selwyn 2015) in that these areas illustrate the unintended consequences of LA when applied in practice. The practice orientation of these critical areas provides a variety of topics that researchers should begin to explore.

We do not deny that LA research has contributed to improvements in pedagogy, technological design, and learning outcomes. Furthermore, we recognize that LA scholars are reflective and critical of their own work. We argue, however, that LA research writ large would be further improved and more self-aware by considering a documentation studies perspective. Our analytical framework enables researchers and practitioners to examine the data they use in model design, algorithms, and data dashboards, along with practices informed by data analytics. LA holds potential to impact higher education—for the good and for the bad—and our documentation studies framework provides a way to ensure that bigger data become better data.

## References

Aguilar, S., S. Lonn, and S. D. Teasley. 2014. “Perceptions and Use of an Early Warning System During a Higher Education Transition Program.” In *Proceedings of The Fourth International Conference on Learning Analytics and Knowledge*, 113–117. New York, NY: ACM. doi: 10.1145/2567574.2567625.

Alblawi, A. S., and A. A. Alhamed. 2017. “Big Data and Learning Analytics in Higher Education: Demystifying Variety, Acquisition, Storage, NLP and Analytics.” *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, 124–129. IEEE. doi: 10.1109/ICBDAA.2017.8284118.

Arnold, K. E., and M. Pistilli. 2012. "Course Signals at Purdue: Using Learning Analytics To Increase Student Success." In *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, 267-270, Vancouver, Canada: ACM. doi:10.1145/2330601.2330666.

Bowker, G. C. 2013. “Data Flakes: An Afterword to ‘Raw Data’ is an Oxymoron.” In *“Raw data” is an Oxymoron*, edited by L. Gitelman, 167–171. Cambridge, MA: MIT Press.

boyd, d., and K. Crawford. 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” *Information, Communication & Society* 15 (5): 662–679. doi: 10.1080/1369118X. 2012.678878.

Briet, S. 1951. *Qu’est-ce que la Documentation?* Paris, France: ÉDIT.

Brown, J. S., and P. Duguid. 1996. “The Social Life of Documents.” *First Monday* 1 (1). http://firstmonday.org/ojs/index.php/fm/article/view/466/387

Buckland, M. K. 1997. “What is a ‘Document?’” *Journal of the American Society for Information Science* 48 (9): 804–809. doi: 10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V

Buckland, M. K. 2000. “The Academic Heritage of Library and Information Science: Resources and Opportunities.” Paper presented at the opening plenary session, Association for Library and Information Science Education 85th Anniversary Celebration, San Antonio, TX, January 11-14.

Buckland, M. K. 2013. “Document Theory: An Introduction.” In *Records, Archives and Memory: Selected Papers from the Conference and School on Records, Archives and Memory Studies,* edited by M. Willer, A. J. Gilliland, and M. Tomić, 223–237. Zadar, HR: University of Zadar. http://people.ischool.berkeley.edu/~buckland/zadardoctheory.pdf.

Buckland, M. K. 2014. “Documentality Beyond Documents.” *The Monist* 97 (2): 179–186. http://www.jstor.org/stable/44012677.

Caluya, G. 2010. “The Post-Panoptic Society? Reassessing Foucault in Surveillance Studies.” *Social Identities: Journal for the Study of Race, Nation and Culture* 16 (5): 621–633. doi: 10.1080/13504630.2010.509565.

Caulfield, M. “Why the Course Signals Math Does Not Add Up,” *Hapgood* (blog), September 26, 2013a, https://hapgood.us/2013/09/26/why-the-course-signals-math-does-not-add-up/.

Caulfield, M. 2013b. “What the Course Signals ‘Kerfuffle’ is About, and What it Means to You.” *EDUCAUSE Review Online*, November 13. https://er.educause.edu/blogs/2013/11/what-the-course-signals-kerfuffle-is-about-and-what-it-means-to-you.

Crooks, R. (2017). “Representationalism at Work: Dashboards and Data Analytics in Urban Education.” *Educational Media International* 54 (4): 289–303. doi: 10.1080/09523987.2017.1408267

Crow, M. M. 2012. “‘No More Excuses’: Michael M. Crow on Analytics.” *EDUCAUSE Review* 47 (4): 14–22. https://er.educause.edu/~/media/files/article-downloads/erm1241p.pdf.

Day, R. 2015. *Indexing It All: The Subject in the Age of Documentation, Information, and Data*. Cambridge, MA: MIT Press.

Denley, T. 2013. “Degree Compass: A Course Recommendation System.” *EDUCAUSE Review Online*, September 4. http://er.educause.edu/articles/2013/9/degree-compass-a-course-recommendation-system. Accessed 15 March 2018.

Ekbia, H., M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V. Ratandeep Suri, et al. 2015. “Big Data, Bigger Dilemmas: A Critical Review.” *Journal of the Association for Information Science and Technology* 66 (8): 1523–1545. doi: 10.1002/asi.23294.

Ekowo, M., and I. Palmar. 2016. “The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis. *New America*, October 24. https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/. Accessed 15 March 2018.

Espeland, W. N., and M. L. Stevens. 2008. “A Sociology of Quantification.” *European Journal of Sociology* 49 (3): 401–436. doi: 10.1017/S0003975609000150

Essa, A. “Can We Improve Retention Rates by Giving Students Chocolates?,” *Data Entropy* (blog), October 14, 2013, http://alfredessa.com/2013/10/can-we-improve-retention-rates-by-giving-students-chocolates/.

Feldstein, M. “Course Signals Effectiveness Data Appears to be Meaningless (and Why You Should Care),” *e-Literate* (blog), November 3, 2013a, https://mfeldstein.com/course-signals-effectiveness-data-appears-meaningless-care/.

Feldstein, M. “Purdue’s Non-Answer on Course Signals,” *e-Literate* (blog), November 6, 2013b. https://mfeldstein.com/purdues-non-answer-course-signals/.

Feldstein, M. 2016. “Understanding the Origins of Ed-Tech Snake Oil.” *The Chronicle of Higher Education*, April 6. https://www.chronicle.com/article/Understanding-the-Origins-of/235977.

Floridi, L. 2012. “Big Data and Their Epistemological Challenge.” *Philosophy & Technology* 4 (25): 435–437. doi: 10.1007/s13347-012-0093-4.

Floridi, L. and M. Taddeo. 2016. “What is data ethics?” *Philosophical Transactions of the Royal Society of London: A Mathematical, Physical and Engineering Sciences* 374: 1–5. doi: 10.1098/rsta.2016.0360

Fountain Hopper. 2015a. “FoHo: How to Get Your Internal Stanford Admissions File (Or: What They Really Thought of You).” *The Fountain Hopper*, January 15. https://us9.campaign-archive.com/?u=c9d7a555374df02a66219b578&id=3a69d6c439.

Fountain Hopper. 2015b. “FoHo #20: KA Under Investigation, FERPA Followups, CAPS Funding and Other Stanford News You Can Use.” *The Fountain Hopper*, March 17. https://us9.campaign-archive.com/?e=%5BUNIQID%5D&u=c9d7a555374df02a66219b578&id=c113f01f39.

Frické, M. 2015. “Big Data and Its Epistemology.” *Journal of the Association for Information Science and Technology* 66 (4): 651–661. doi: 10.1002/asi.23212.

Gioia, M. 2015. “First Students Gain Access to their Admissions Files Through FERPA Provision.” *The Stanford Daily*, March 9. https://www.stanforddaily.com/2015/03/09/first-students-gain-access-to-their-admissions-files-through-ferpa-provision/.

Haggerty, K. D., and R. V. Ericson. 2000. “The Surveillant Assemblage.” *The British Journal of Sociology* 51 (4): 605–622. doi: 10.1080/00071310020015280.

Hardy, I. 2015. “Data, Numbers and Accountability: The Complexity, Nature and Effects of Data use in Schools.” *British Journal of Educational Studies* 63 (4): 467–486. doi: 10.1080/00071005.2015.1066489

Illiadis, A., and F. Russo. 2016. “Critical Data Studies: An Introduction.” *Big Data & Society* 3 (2): 1–7. doi: 10.1177/2053951716674238

Johnson, J. A. 2017. “Structural Justice in Student Analytics, or, the Silence of the Bunnies.” Paper presented at the annual meeting for the Eastern Sociological Society: 2017 Digital Sociology Mini-Conference, Philadelphia, February 24.

https://the-other-jeff.com/wp-content/uploads/2017/02/Structural-Justice-in-Learning-Analytics-ESS.pdf.

Jones, K. M. L. 2018. “Advising the Whole Student: eAdvising Analytics and the Contextual Suppression of Advisor Values.” Education and Information Technologies: 1–22. doi:10.1007/s10639-018-9781-8.

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks, CA: SAGE Publications.

Kitchin, R., and T. Lauriault. 2014. “Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work.” *The Programmable City Working Paper 2*. https://ssrn.com/abstract=2474112.

Kraft-Terry, S., and C. Kau. 2016. “Manageable Steps to Implementing Data-Informed Advising.” *NACADA*. http://www.nacada.ksu.edu/Resources/Clearinghouse/View-Articles/Manageable-Steps-to-Implementing-Data-Informed-Advising.aspx. Accessed 15 March 2018.

Latour, B. and S. Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. London, England: Sage.

Lazer, D., A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, et al. 2009. “Life in the Network: The Coming Age of Computational Social Science.” *Science* 323 (5915): 721–723. doi: 10.1126/science.1167742.

Lifshits, A. 2015. “FERPA Requests Yield Limited Access to Files.” *The Brown Daily Herald*, April 22. http://www.browndailyherald.com/2015/04/22/ferpa-requests-yield-limited-access-files/.

Lund, N. W. 2009. “Document theory.” *Annual Review of Information Science and Technology* 43 (1): 1–55. doi: 10.1002/aris.2009.1440430116.

Lupton, D. 2014. “Self-Tracking Cultures: Towards a Sociology of Personal Informatics.” In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, 77–86. New York, NY: ACM. doi: 10.1145/2686612.2686623.

Manovich, L. 2011. “Trending: The Promises and the Challenges of Big Social Data.” In *Debates in the Digital Humanities*, edited by M. K. Gold, 460–475. Minneapolis, MN: University of Minnesota Press. doi: 10.5749/minnesota/9780816677948.003.0047.

Otlet, P. 1934. Traité de Documentation: Le Livre sur le Livre. Brussels, Belgium: Editiones Mundaneum.

Pérez-Peña, R. 2015. “Students Gain Access to Files on Admission to Stanford.” *The New York Times*, January 16. https://www.nytimes.com/2015/01/17/us/students-gain-access-to-files-on-admission-to-stanford.html.

Phillips, E. D. 2013. “Improving Advising Using Technology and Data Analytics.” *Change: The Magazine of Higher Learning* 45 (1): 48–55. doi: 10.1080/00091383.2013.749151.

Pistilli, M. D., and K. E. Arnold. 2010. “In Practice: Purdue Signals: Mining Real-Time Academic Data to Enhance Student Success.” *About Campus* 15 (3): 22–24. doi: 10.1002/abc.20025.

Polonetsky, J., and O. Tene. 2015. “Who is Reading Whom Now: Privacy in Education from Books to MOOCs.” *Vanderbilt Journal of Entertainment and Technology Law* 17 (4): 927–990. http://www.jetlaw.org/wp-content/uploads/2015/09/PolonetskyTene_Final.pdf.

Pomianowski, J. 2015. “Yale Law School Is Deleting Its Admissions Records, and There's Nothing Students Can Do About It.” *The New Republic*, March 15. https://newrepublic.com/article/121297/yale-law-deletes-admissions-records-congress-must-fix-ferpa.

Prinsloo, P., E. Archer, G. Barnes, Y. Chetty, and D. van Zyl. 2015. “Big (ger) Data as Better Data in Open Distance Learning.” *The International Review of Research in Open and Distributed Learning* 16 (1): 284–306. doi: 10.19173/irrodl.v16i1.1948.

Prinsloo, P., and S. Slade. (2017). “Ethics and Learning Analytics: Charting the (Un)Charted.” In *Handbook of Learning Analytics,* edited by C. Lang, G. Siemens, A. Wise, and D. Gašević, 49–57. doi: 10.18608/hla17

Rubel, A. and K. M. L. Jones. 2016. "Student Privacy in Learning Analytics: An Information Ethics Perspective. *The Information Society* 32 (2): 143–159. doi: 10.1080/01972243.2016.1130502.

Ruckenstein, M. 2014. “Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles.” *Societies* 4 (1): 64–84. doi: 10.3390/soc4010068.

Ruppert, E. 2013. “Rethinking Empirical Social Sciences.” *Dialogues in Human Geography* 3 (3): 268–273. doi: 10.1177/2043820613514321.

Schürmeyer, W. 1935. “Aufgaben und Methoden der Dokumentation.” *Zentralblatt für Bibliothekswesen*, 52: 533–543.

Selwyn, N. 2015. “Data Entry: Towards the Critical Study of Digital Data and Education.” *Learning, Media and Technology* 40 (1): 64–82. doi: 10.1080/17439884.2014.921628.

Siemens, G. 2012. “Learning Analytics: Envisioning a Research Discipline and a Domain of Practice.” In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, edited by S. Buckingham Shum, D. Gašević, and R. Ferguson, 4–8. New York, NY: ACM. doi: 10.1145/2330601.2330605.

Siemens, G. 2013. “Learning Analytics: The Emergence of a Discipline.” *American Behavioral Scientist* 57 (10): 1380–1400. doi: 10.1177/0002764213498851.

Straumsheim, C. 2013. “Researchers Cast Doubt about Early Warning Systems’ Effect on Retention.” *Inside Higher Ed*, November 6. http://www.insidehighered.com/news/2013/11/06/researchers-cast-doubt-about-early-warning-systems-effect-retention.

Svrluga, S. 2016. “University President Allegedly Says Struggling Freshmen are Bunnies that should be Drowned.” *The Washington Post*, January 12. https://www.washingtonpost.com/news/grade-point/wp/2016/01/19/university-president-allegedly-says-struggling-freshmen-are-bunnies-that-should-be-drowned-that-a-glock-should-be-put-to-their-heads/.

Tally, S. 2013. “Purdue Software Boosts Graduation Rate 21 Percent.” *Purdue University*, September 25. https://web.archive.org/web/20131102055008/http://www.purdue.edu/newsroom/releases/2013/Q3/purdue-software-boosts-graduation-rate-21-percent-by-data-mining-academic-behviors.html.

Tsesis, A. 2014. “The Right to Erasure: Privacy, Data Brokers, and the Indefinite Retention of Data.” *Wake Forest Law Review* 49: 433–484. https://lawecommons.luc.edu/cgi/viewcontent.cgi?article=1502&context=facpubs.

University Innovation Alliance. n.d.. “Collaborative Project Goal: Predictive Analytics.” http://www.theuia.org/sites/default/files/UIA_predictive_onepagers.pdf.

van Barneveld, A., K. E. Arnold, and J. P. Campbell. 2012. “Analytics in Higher Education: Establishing a Common Language.” *EDUCAUSE Learning Initiative*, ELI Paper no. 1. https://library.educause.edu/~/media/files/library/2012/1/eli3026-pdf.pdf.

Williamson, B. 2018. “The Hidden Architecture of Higher Education: Building a Big Data Infrastructure for the ‘Smarter University.’” *International Journal of Educational Technology in Higher Education* 15 (12): 1–26. doi: 10.1186/s41239-018-0094-1