Spark: UAL Creative Teaching and Learning Journal

Exploring student control over data privacy through speculative design

Michael Madaio, Ph.D. student, Human-Computer Interaction, Carnegie Mellon University

Abstract
Higher education institutions face a number of questions about the data they gather from students and how it will be used in the future. Between commercial incentives for collecting student data and anxieties about the privacy and security of that data, lies the intentional sharing exemplified by the 'Quantified Self' movement, when users voluntarily provide information for personal development. Though there are risks to making student data available to third-party service providers, students could benefit from advanced learning analytics that personalise their educational experience. This research paper describes how designers of educational technologies can use 'speculative design' as one approach to exploring possible futures, in this case, a future in which students are able to control the data they share.

Keywords
speculative design; student data privacy; learning analytics

Introduction
In recent years, as digital learning technologies become more ubiquitous in the classroom as well as in more informal learning environments, a wealth of student data has been collected. In an online class for instance, these data sets may comprise everything from the viewing behaviour of students watching lecture videos, to their participation in a discussion forum. Burgeoning areas of research include analyses of multimodal video data, gathered from the students' own cameras, to study their facial expressions as they learn. It is argued that this information is gathered in order to benefit students and provide them with a more personalised learning experience, tailored to their ability levels and occurring at a pace that suits their learning. However, in order to provide personal learning adaptations, schools, or more realistically, the educational software companies contracted to provide these services, must first aggregate a large amount of data from many students in order to develop and deploy the 'machine learning algorithms' that power 'predictive learning analytics'.

Predictive analytics are increasingly commonplace in the consumer sector. Companies such as Netflix and Amazon aggregate data from their users in order to provide a more personalised media viewing or shopping experience in which users are recommended items. Other technology companies such as Google, Facebook and Twitter harvest vast quantities of data, extracting information about their users' preferences in order to, among other things, provide them with targeted advertisements that they are more likely to 'click on' and thus purchase. The World Economic Forum refers to personal data as, alternatively, 'a new asset class', 'the new oil' or 'a new currency' (WEF, 2011).

However, it is not always clear who the owners of that new asset class are. Is it the users who produce data? Or is it, in fact, the companies that 'aggregate' such data, extracting insights and

delivering product recommendations that users may or may not need or want? This is an important question that needs to be asked, as excitement about the potential insights derived from data builds. We should consider the 'value' of student data, as their desires and perspectives are often left out of the conversation – particularly their desire for privacy and control over what is made available to third-party service providers. This research paper explores how a 'speculative design' approach can help designers develop educational technologies that consider students, their relationship to their data, and their attitudes towards privacy, and that can encourage students to think about how this information will be used in the future.

Using a speculatively designed student data privacy system entitled 'KnowledgeTrees' as an example, this paper discusses the speculative design process as well as the particular future that this work envisions. This fictional project probes the ways in which a learning management system (LMS), which uses student data to support instruction, could provide students with more granular control over their data privacy, allowing them to be the primary beneficiaries of its 'value'. Designed to raise more questions than answers, this work presents a vision of a future world in which the metaphor of data as an asset class or currency is taken seriously, as a way of encouraging designers to think more deeply and ethically about the values 'encoded' into learning technologies.

Economic context

In the current landscape of increasingly ubiquitous learning technologies, there is a set of contrasting forces at work. On one hand, economic incentives are driving an increase in student data collection, while at the same time, other forces are working to inhibit ease of access that third-party service providers have to data. This section describes the economic incentives that motivate venture capital investment in ever more complex platforms for learning, within which students are generating vast quantities of data. Investments in US-based 'ed tech companies' (a slippery, often ill-defined category label) increased from less than $200 million in the first half of 2010 to almost $1.1 billion in the first half of 2015 (Wan and McNally, 2015). For the return on that investment, a report by the McKinsey Institute estimates the value of global open education data at $800 billion over the next five years, but this is dependent upon whether data is made 'open' to third party companies (Manyika *et al*, 2013).

Anxieties

And yet, there is pushback. Producers of data are resistant to becoming free labour for the benefit of the technology companies that collect and refine their data, the harvesters of this 'new oil'. In particular the 'Parent Coalition for Student Privacy' (PCSP) have recently been advocating for increased federal regulations to protect student data privacy, culminating in the testimony of co-founder and co-chair Rachael Stickland before the House of Representatives' Committee on Education and the Workforce in March 2016 (U.S. Education and the Workforce Committee, 2016). Though the Family Educational Rights and Provisions Act (FERPA) does protect some aspects of student data, after 40 years from its inception, it has not been adequately updated to reflect the ease, granularity, and scale of data collection facilitated by digital learning platforms.

The anxieties of students and their parents are attested to by the fact that there were over 100 state-level bills about student data privacy introduced in 2014 alone, 28 of which were enacted into law, evidence of widespread concerns about student data privacy (Trainor, 2015). With these public anxieties come commercial repercussions for education technology companies. For example, the private non-profit organisation InBloom – described as a 'data warehouse' – was shut down in 2014 (after less than three years of operation), amid serious concerns about how they were aggregating and using student data (MacCarthy, 2011).

Learning benefits
Despite anxiety over the extent to which third-party companies can access educational data, such data represent an opportunity to provide real educational 'value' for students. Where the predictive analytics of a company like Netflix or Amazon might be used to recommend to customers the next movie they should watch or book they should buy, an educational technology that collects fine-grained data about student learning could use analytics to predict the kinds of materials students should learn, their level of difficulty or what type of instruction would best support student learning (Siemens, 2013). In a traditional analogue classroom, the teacher performs this 'personalisation' as they review students' assessment performances and modify their teaching to respond appropriately. However, analogue approaches to personalised learning are difficult to scale and implement for students at the university scale, or even on the massive scale adopted by "Massive Open Online Courses" (MOOCs), which can have upwards of 10,000 students at a time (Siemens, 2013).

Among other applications, many universities are attempting to use the data they collect to improve the quality of student academic advising. Students at risk of failing a course or dropping out are targeted for advice and support using predictive analytics (Fritz, 2011; Slade and Prinsloo, 2013). To do this well, requires large amounts of student data, aggregated across students and across courses, meaning that students are encouraged to make the data (from their learning behaviours) accessible on a learning platform for third-party service providers to develop analytics that are purportedly for the students' own benefit.

Risks
Even with the potential benefits of actions such as proactive, 'intrusive' advising or personalised learning and so forth, there are risks to sharing student data with third-party providers. First and foremost, even when personally identifiable information (PII) is removed from student data records, it is possible to reverse this and 'de-anonymise' the data by aggregating across many data points from the same individual (Prinsloo and Slade, 2015). Furthermore, the meaning of even the most granular data on student learning behaviours may not be readily apparent, as they always require an act of interpretation on the part of designers or users of learning analytics algorithms. For instance a long pause in watching patterns may indicate that students do not understand, or it might mean that they went for a snack (Siemens, 2013). Other learning analytics systems work by using multimodal audio and video data of students' facial expressions to infer affect (emotions) from the movement of small muscle groups in their faces (Taub *et al*, 2016). Although algorithms' ability to detect the movements of such 'facial action units' is indisputable, the accuracy of their ability to map those expressions to emotions is less clear. For instance, a single facial movement (a brow furrowing) could be indicative of a student's frustration, confusion, or focused concentration.

Beyond the risks of de-anonymisation and interpretation issues, there is also the risk that predictive learning analytics will reinforce existing patterns of student behaviour, 'keeping them prisoner to their past choices' (Slade and Prinsloo, 2013, p. 1517). The dangers of this can be seen in the current media landscape, where personalized, news content serves to reinforce readers' existing beliefs, leading to a so-called 'echo chamber' in which people see only the news they want to see. For students, if their learning patterns show that they learn best in a particular medium (video lectures, for instance) they might be recommended content presented entirely on video formats, denying them the opportunity to be challenged by the format or the content of the material. At its most dire, this sort of self-reinforcement could lead to the disenfranchisement of vulnerable populations, for example when students who have difficulties with a subject are tracked into classes of 'underperforming' students, a technique that precludes them the benefit of working in and being challenged by heterogeneous groups of varied abilities (Prinsloo and Slade, 2015; Wen, 2015; Zurita, Nussbaum and Salinas, 2005).

Quantified self

Despite the aforementioned anxiety, many people already share their personal data to a significant extent. Between the various incentives to collect and aggregate personal data and anxieties about personal data privacy, lies the voluntary release of the so-called 'Quantified Self' (Wolf, 2010). The 'Quantified Self' describes people who for example, wear wrist-based technology in exchange for easily visualised information about exercise patterns or sleep habits. Products such as the Fitbit, the Jawbone, and the Microsoft Band are designed for just such a market, using gyroscopes and accelerometers to track their users' movement and sleep. This trend adds to the enormous quantity of other personal data people now willingly share on social media, including location check-ins on Foursquare, restaurant reviews on Yelp, and Tweets and Facebook posts galore.

The insouciance with which we have come to freely give away details about our location, our purchasing preferences and personal opinions, in exchange for free services is rivaled only by the average consumer's lack of knowledge about how that data might be used by the corporations that collect it (Slade and Prinsloo, 2013). The difference between how people feel about the data they willingly share on social media or via fitness tracking devices, and wider anxieties about students' learning data is that the personal benefits of sharing exercise data are more immediately visible. Personal exercise informatics systems, display patterns of walking, running or sleeping in colourful charts and graphs. On learning platforms, however, the benefits of sharing personal learning data may be less immediately obvious.

Opt-in or opt-out?

The way forward for student data seems clear. Why not allow students to make individual decisions about the data collected from and made public about them, as many already do on social media? Some have proposed allowing students to opt-out of data collection (with data collection being the default, of course) – an option overwhelmingly favoured by students in one study by Slade and Prinsloo (2014). On the other hand, in another study by those same researchers, students preferred having the option to choose whether certain data was made available, though that was for the purpose of marketing analytics (Prinsloo and Slade, 2015).

The challenge of an individually managed 'opt-in' approach is that students may not always be fully aware of the risks of having their data collected and made available to third-party providers, which is necessary for them to make an informed decision. Alternatively, when weighing the potential benefits that sharing their data may provide them, students may be informed about the short-term privacy risks, but unaware of the longer-term consequences of a data security breach (Prinsloo and Slade, 2015). There is of course, the issue of to what extent an End-User License Agreement (EULA) actually informs an end user about the potential range of uses to which their data might be put. This is an important field of research in itself and well beyond the scope of this paper (see Grossklags and Good, 2007, for a review).

With consumer technology, if the user would prefer to have a reduced scope of service, in exchange for allowing the app or wearable device to collect less data on them, that is their choice. However, the university as an educational institution has a responsibility to ensure that each student is provided with the best possible education (DeAngelis, 2014). Unlike the consumer technology sector, higher education institutions have a 'fiduciary duty' and responsibility to educate students. Thus, those institutions face challenges when it comes to balancing the inherent paternalism of a decision that may benefit students despite their wishes, with the fact that respecting students' agency might involve making decisions that are not pedagogically beneficial for them.

Speculative design

Speculative design can be used to explore the possible futures arising from design decisions made in the design of novel technologies. Specifically, this article demonstrates how a set of guidelines can inform the design of a (speculative) learning analytics platform that allows students to have more control over the privacy of their student data. This fictional artefact can be used as a 'design probe' to drive conversations about how such a platform might balance the 'soft paternalism', that nudges students to share data for their own good and on the other, the agency of students to control their personal learning data.

Speculative design is an exploratory design methodology that allows researchers and designers to probe possible futures without being beholden to the economic incentives of more traditional design methods. As articulated by Dunne and Raby in their book, *Speculative Everything* (2013), speculative design attempts to interrogate the normative assumptions embedded into the design of systems that are both technical and social. These products are not designed for mass consumption, but for 'mass conversation' instead, and though grounded in reality, need not be tied to practical engineering considerations (DiSalvo, 2012; Gray and Boling, 2016). The artefact designed using this approach is not the product of a realistic system of production, but instead arises from an extrapolation of the values, norms, and practices of our current social system.

Educational technologies may be shaped by a variety of factors, including the design of previous systems, users' expectations and how those tools fit into the existing ecology of learning technologies in a socio-technical educational system. By contrast, speculative designs do not have these constraints and are thus better able to tease apart the values and assumptions that are encoded into the design of technologies. Through the critical conversations that arise around artefacts from a 'possible future', speculative design reveals the values instantiated in current tools and systems.

One common method employed in speculative design is to use metaphors or tropes to highlight an aspect of the phenomena being critiqued, taking them seriously during the design of an artefact. For instance, the 'Audio Tooth', developed by James Auger in 2001, is a classic example of the successful use of metaphor in speculative design. Developed at a time when personal, handheld electronics were becoming widespread, it was designed to critique the idea of personal augmentation through technology (Auger, 2013). As he argued, people were at this time already augmenting their bodies and senses through personal computers and handheld phones and cameras, and he wanted to probe the threshold of this acceptance. Grounded in reality – through the appearance of a professional product design and the use of technical jargon describing its internal mechanisms – it was taken seriously by large numbers of people, as evidenced by being featured in *Wired* magazine and included on the cover of *Time* magazine, which described it as one of 'the coolest inventions of 2002' (Auger, 2013). By taking the metaphor of personal augmentation through technology and pushing it to its limit in a technological tooth implant, Auger was able to highlight the absurdity of the desire for self-augmentation.

Another common method for speculative design is to use ambiguity in the design to provoke the viewer to fill in the gaps themselves, encouraging them to make explicit their assumptions that might otherwise have been hidden. This can be done through a low-fidelity prototype, where the unfinished aesthetic allows the viewer to focus on the mechanisms at work, instead of on the professional veneer of a seemingly more 'finished' product. The 'Metadating' project can be seen as one example of this. Designers created a paper prototype of a fictitious online dating platform, where users could share their personal data as part of a face-to-face or speed-dating event (Elsden *et al*, 2016). The low fidelity design allowed each user to choose different aspects of their identity to represent through data. Participants in the Metadating study did not simply choose to represent themselves to potential

partners by answering questions and allowing a matching algorithm (like that used by dating sites such as OkCupid) to recommend a potential partner to them. Instead – although they still shared their data with other participants – they manually drew representations of their dating data out on paper, which was woven into a narrative they constructed with their partners throughout the 'Metadating' event. Their personal data became not just grist for the mill of a dating recommendation algorithm, but it instead became a social artefact, something discussed, interpreted, downplayed, and built on in subsequent interactions (Elsden *et al*, 2016). This low fidelity data representation and the lo-fi nature of the paper artefact enabled participants to insert more of themselves into the speculation, building upon what was there, as opposed to being given a complete product.

Design guidelines
With those speculative design methods in mind, the KnowledgeTrees LMS was created. By first developing a set of design guidelines based on the methods and problem space described above and considering the challenge of ensuring student data privacy, this project aimed to explore how students might control their data privacy, turning this so-called 'new oil' to students' own benefit. Below are some examples of the design guidelines and the potential questions inspired by them, used to guide the development of KnowledgeTrees.

Guideline 1: Take seriously the metaphor of data as a currency.

- What is the 'economic' value of a data point?
- How will students keep track of the data 'points' that they have accrued?
- Will different kinds of data be valued at the same rate?
- What will determine the price of various insights to be purchased with the data points?
- Once the data is 'spent', who will be able to access or view it?
- In an economy built on the earning and spending of personal data, how will economic phenomena such as inflation be prevented?

Guideline 2: Consider the impact on the social system of the class or university from a data economy.

- Will students be able to see what data the other students have shared?
- Will students know how many data points the other students have earned?
- Will students know what insights their classmates have spent their data points on?
- Should the system nudge more students to share more data?
- Or should it prompt students to provide social pressure on their peers to share more data?

Guideline 3: The design should be grounded in real LMS data management, as well as the aesthetic design style of a traditional LMS.

- How much detail should be included about the data being made public?
- If there is data that is 'transformed', should the lowest level of detail be displayed, such as the timestamps of clicks on an LMS? Or should it be displayed at a more abstract level, such as the patterns of usage?
- What type of information visualization, if any, should be used for the data point management?

System design
With these guidelines and questions in mind, I set out to design a (fictitious) system architecture to explore how the 'KnowledgeTrees' system might operate. As part of this process, I created process models and system architecture diagrams, which, in addition to fulfilling the software engineering purpose of coordinating computational system modules, are as much cultural artefacts as the system itself is (Shaw, 1995; Jansen and Bosch, 2005). The speculatively designed system for

KnowledgeTrees was then built as a prototype, using a low-fi Javascript-based website, following the design of the system architecture diagram (Figure 1). This two-step process aimed to make the system's possible use case and application more visible, to help articulate the design guidelines for such a system, were it to actually exist.
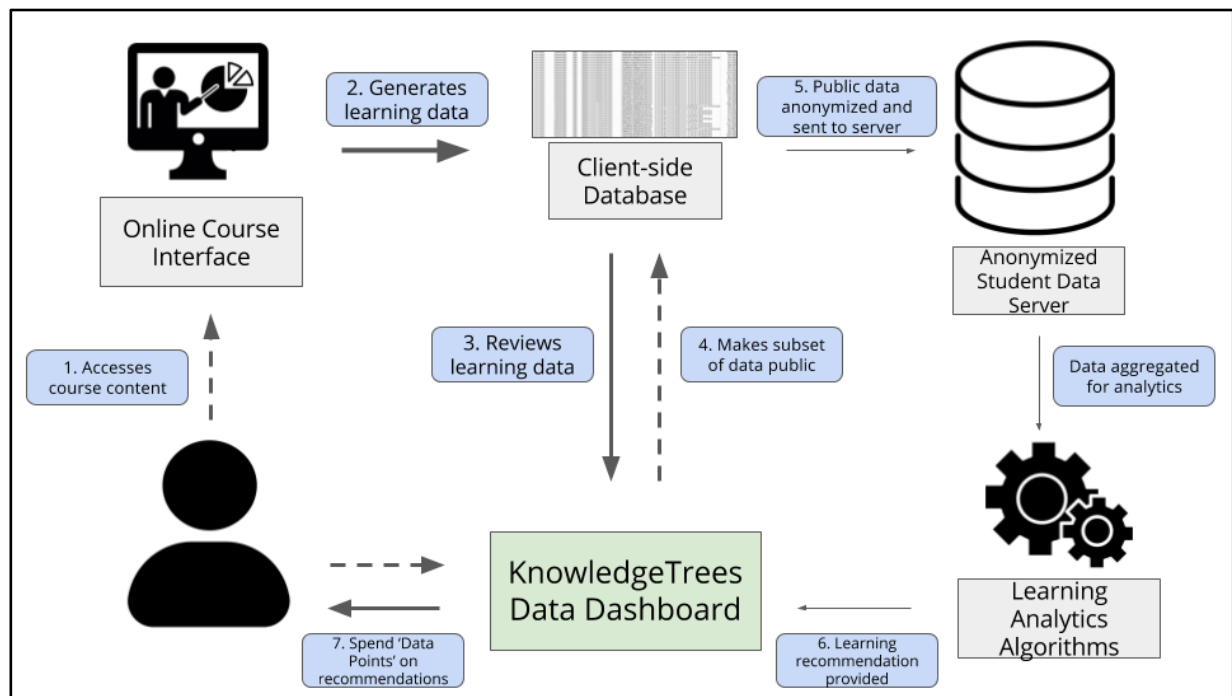

Figure 1: KnowledgeTrees – system architecture diagram (Madaio, 2016).

The KnowledgeTrees System Architecture Diagram outlines a number of steps:

      Step 1: The student interacts with the online course interface.
      Step 2: Data is stored in a client-side, private database.
      Step 3: Using the KnowledgeTrees Data Dashboard, the student reviews the data they generated.
      Step 4: The student can choose to make some subset of that data public, earning 'data points'.
      Step 5: Publicly released data is anonymised and sent to a student data server and aggregated with data from other students in the class and university.
      Step 6: Aggregated public data is used by machine learning algorithms to generate predictions.
      Step 7: The student spends their 'data points' to access their learning recommendations.

At Step 3 in the process, students can see how 'much' data was generated by each of the items in each of the three categories. Students generate 3 main types of data:

Data type 1
Learning Object data, such as lecture videos watched, homework, quizzes and practice problems attempted (Figure 2).
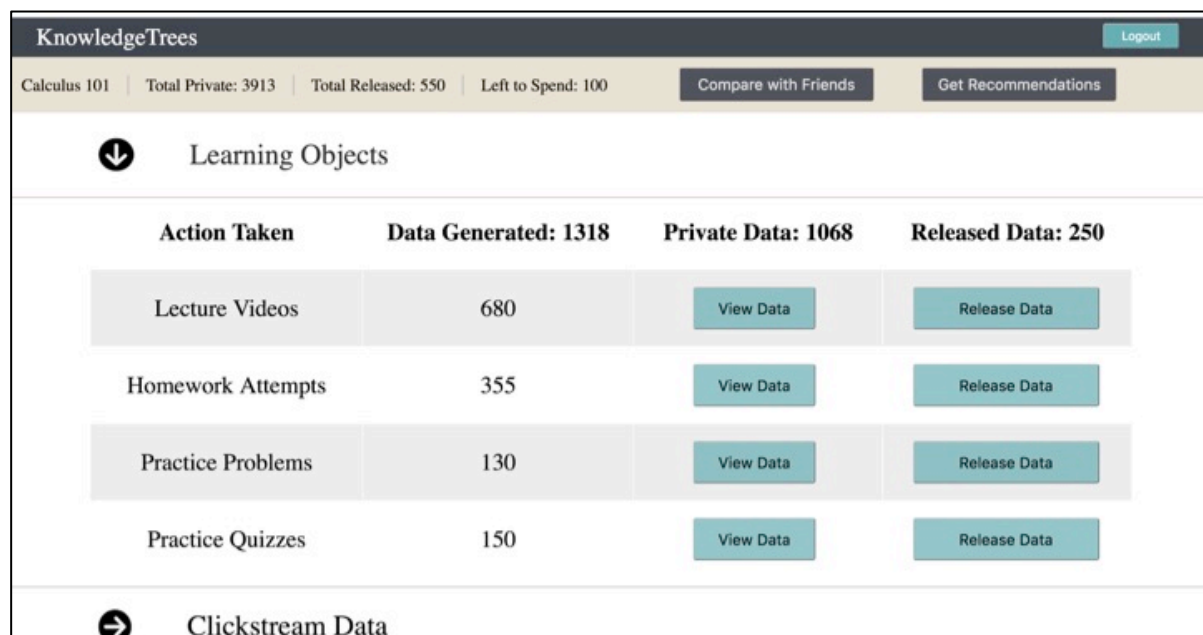

Figure 2: Data type 1, learning object data details (Madaio, 2016).

Data type 2
Clickstream data, which provides more detail about materials accessed and patterns of video watching behaviours, such as pause, rewind, fast-forward and their participation in the discussion forum (Figure 3).
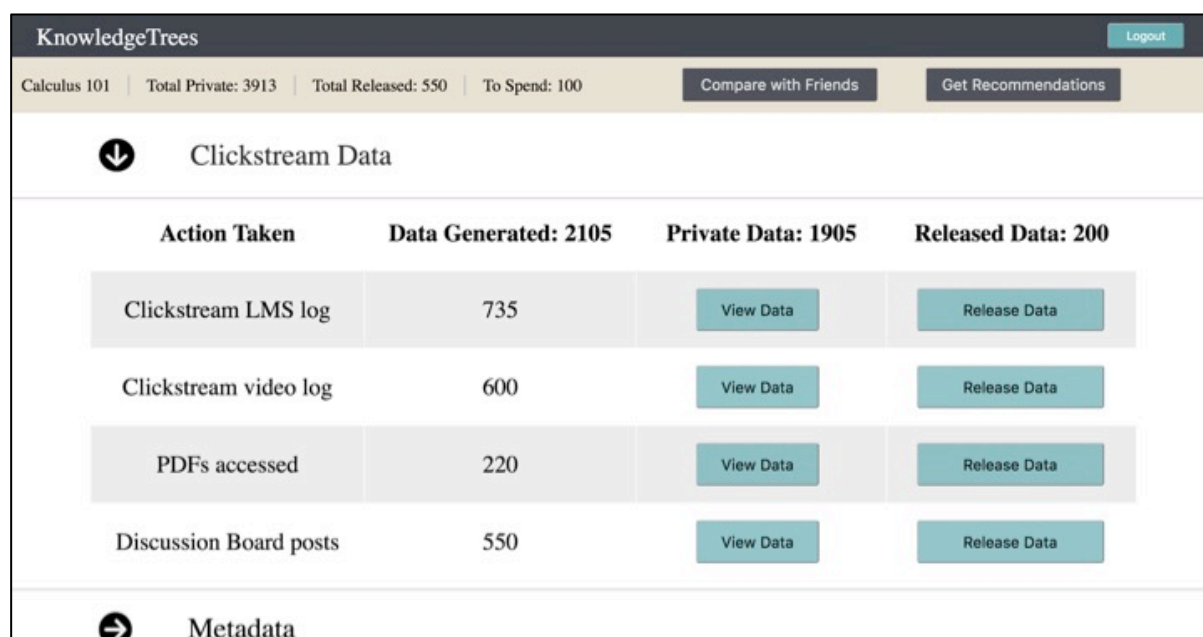

Figure 3: Data type 2, clickstream data details (Madaio, 2016).

Data type 3
Metadata provides more abstract or high-level representations than the clickstream or learning object data. As seen in Figure 4, this includes categories such as location and temporal usage data, learning style and collaboration style.
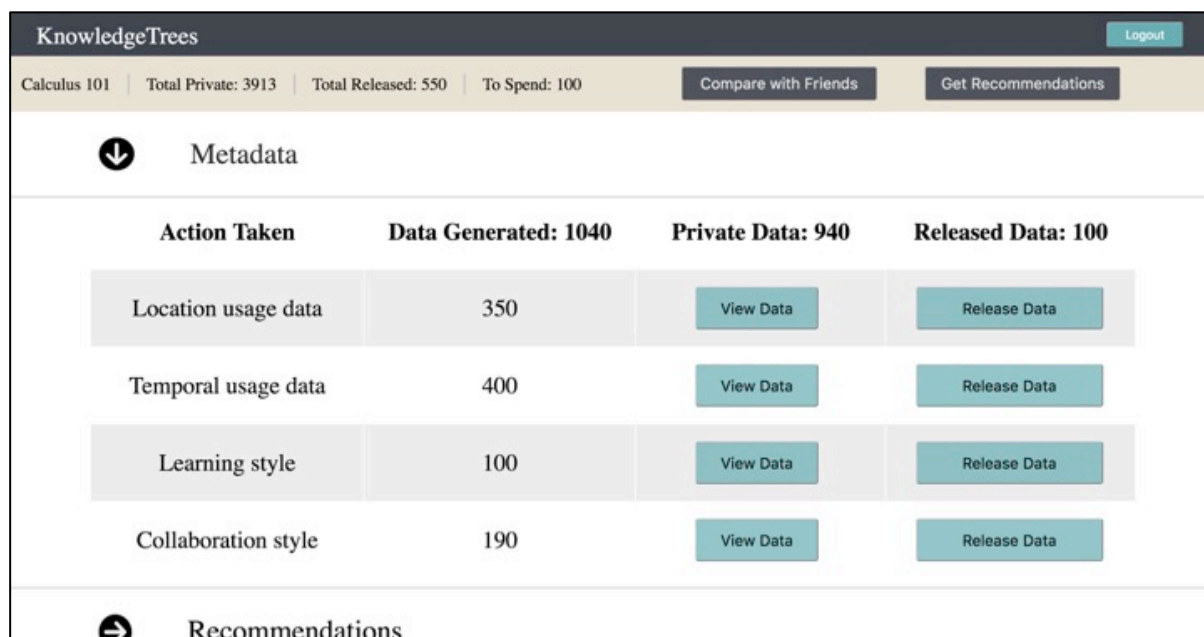

Figure 4: Data type 3, metadata details (Madaio, 2016).

At Step 4, students can choose to make some subset of that data public, earning 'data points'. By 'releasing' the data, they earn data points, which can be spent on recommendations and insights about their learning behaviours (Figure 5).
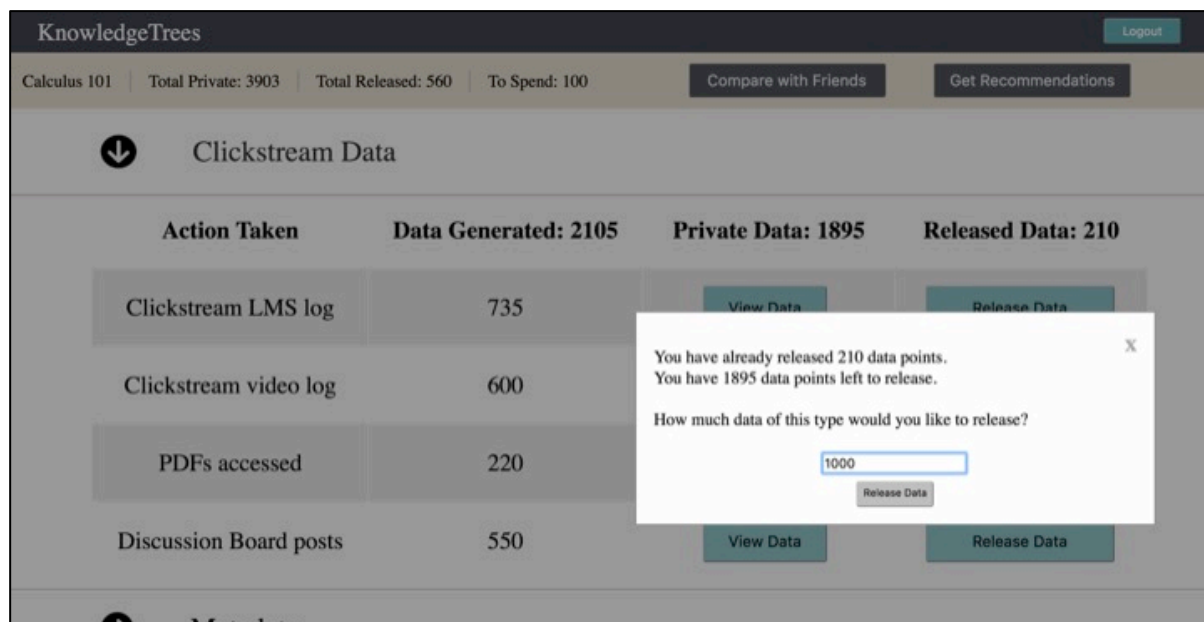

Figure 5: Making data publicly visible to earn data points (Madaio, 2016).

In Step 5, the data they make public is anonymised and sent to a student data server, aggregated with data from other students in the class and university. Then, in Step 6, the aggregated, public data

is used by machine learning algorithms to generate predictions about when, how and even where each student learns best (Figure 6).
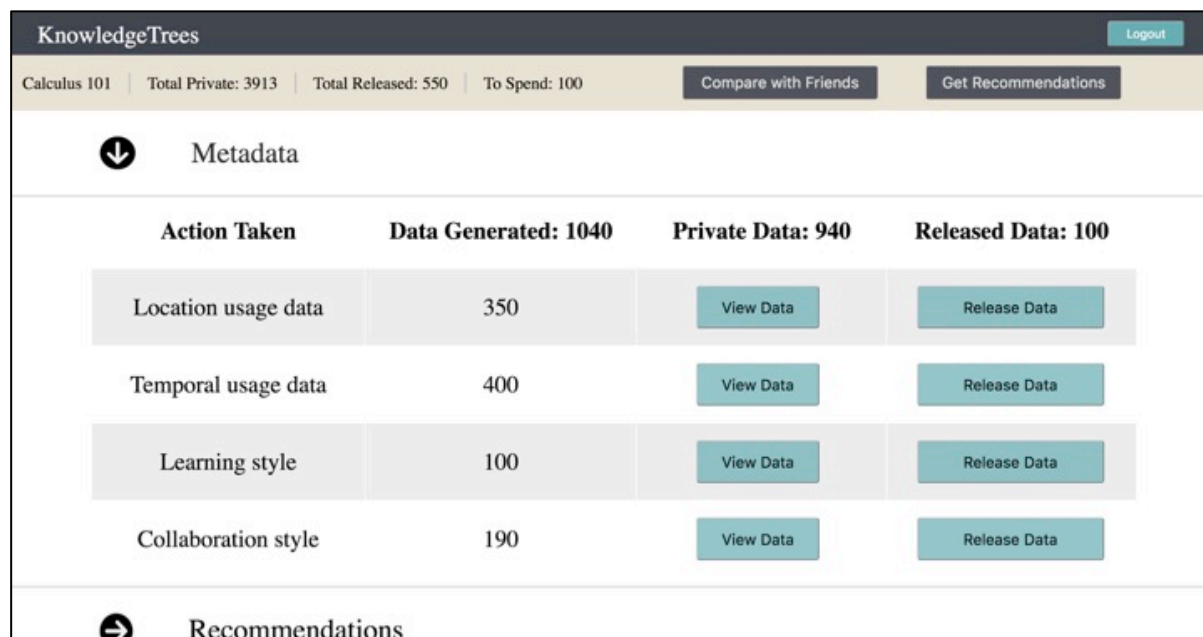

Figure 6: Recommendations to be purchased by students, using data points (Madaio, 2016).

Finally, in Step 7 the student spends their 'data points' to access learning recommendations. By 'spending' data points – earned through making certain data public – they are able to 'purchase' recommendations or insights about their learning behaviour. This includes such recommendations as the optimal lecture video for them to watch next (if multiple options are available), the specific exercises they would be best served by practicing, or even the optimal time of day to take quizzes or the best student to collaborate with, all based on their performance in the class as inferred from the specific data they released.

Students are then encouraged to make more of their data public, as the accuracy for the predictions will improve, due to having more data on which to train the machine-learning model. This encouragement is made explicitly in the form of a message displayed after 'purchasing' a recommendation, and implicitly from the ability to view the amount of data released by other students, for each of the three data types. Students can then see how their data sharing behaviours compare to their classmates', which may function as a wider social incentive for peers to share more, for the good of the class.

Discussion
Speculative design is one of many differing methods used by designers to make possible futures more visible, in order to stimulate conversations about the future we want to live in. Designers of educational technologies, as is true for many other technologies, may often have economic and commercial incentives for creating those tools. A speculative design approach, however, allows designers to circumvent those incentives and probe the values embedded in designed artefacts in a more exploratory way. The 'KnowledgeTrees' design seeks to explore one potential (perhaps undesirable) future, in which personal data collection will continue unabated, anticipating that it will likely increase in granularity and frequency, driven largely by economic incentives. As we move towards that (possible) world, designers can use a speculative design approach to foster a larger conversation about the values that are incorporated into the technologies we design and use.

Many people are already using data about aspects of their lives to change their behaviour – from exercising to driving or even dating – and students could likewise be actively involved in a similar process of positive behavior change, using the data generated from their learning activities to improve their learning. KnowledgeTrees was created to probe the values embedded in one particular learning technology from that potential future. This speculative design artefact would ideally serve as a design probe that real students, faculty or educational technology designers could interact with as part of a set of structured conversations about the benefits, risks and anxieties surrounding data privacy. The readers of this research paper, be they design instructors, instructional designers or educational technology designers, can play an important role in inventing this future, through participating in a spirited conversation around the values of privacy and control embedded in the design of learning analytics and management systems.

The act of creating a speculative design artefact such as KnowledgeTrees gave rise to a number of questions, which may prove fruitful in thinking about possible futures for student data privacy and control. When designing KnowledgeTrees, the metaphor of treating data as an asset class or a currency was taken seriously, but during development of the prototype, the limits of this metaphor became clear. Traditionally, 'currency' itself does not have an intrinsic meaning or worth, other than the value it signifies in an economic system. In KnowledgeTrees, the hypothetical students may place an intrinsic value on the privacy of their data (e.g. clickstream logs, discussion forums, and learning traces), which may be different from the economic value third-party service providers may place on that data. In this speculative system, the exchange rates between the student data and the data 'points' (or, 'currency') which they earn by making their data public were deliberately left ambiguous. Using data to earn an economic unit of exchange may also subject students' learning behaviours to other second-order economic incentives, such as students limiting their data sharing practices to avoid inflation, or taking particular (perhaps pedagogically unproductive) learning actions if the data generated by those actions is 'worth' more data points than others. For instance, students who might otherwise not be inclined to share their location data might be encouraged to do so if few location-based data were made public, and thus the value of location data were temporarily increased.

Along with the 'economic' incentives for students to share more data, this model incorporated social incentives for making personal learning data public. Learning analytics and machine learning algorithms are better able to make predictions and more accurate recommendations if they have more data points on which to train and 'learn'. When students are 'spending' their data points on learning recommendations, the accuracy will improve not only if they share more data, but also if their classmates do so too. Conversely, the self-management of learning data might be seen as wasted effort, but the act of reflecting on one's digital traces and learning behaviours could motivate positive behavioural change, much like the idea behind the 'Quantified Self'. Finally, for learners who regularly review their learning data, that 'meta-cognitive' self-reflection could prove extremely beneficial to learning.

Conclusion

In this paper, speculative design has been used as a method to interrogate the ways in which the values embedded in our designed objects create a possible future. KnowledgeTrees, as one such speculative design artefact, attempts to stimulate conversation around the implications of providing students with more control over their learning data privacy. The anxieties that students and parents feel about their control over data privacy mirrors a larger public anxiety about the need for control over personal data. Though the technological infrastructure for large-scale data collection is often motivated by substantial economic incentives, there may also be individual incentives for the collection and use of personal data, as in the positive behavioural change of sleep and exercise

patterns of the 'Quantified Self'. At the heart of discussions about students' self-management of data privacy, there is a tension between the, perhaps paternalistic, responsibility of higher education institutions to provide the best possible education and a student's right to have control over their own education and data privacy. Speculative designs may probe the future, but these tensions exist in the present, as the details of students' learning behaviours are currently being made accessible to the benefit of third-party learning analytics service providers. This question of students' control of their data privacy, then, should ideally become a larger debate involving students, teachers, instructional designers, designers of education technology and many other stakeholders in the education process.

References

Auger, J. (2013) 'Speculative design: crafting the speculation', *Digital Creativity*, 24(1). http://doi.org/10.1080/14626268.2013.767276.

DeAngelis, W. (2014) 'Academic Deans, codes of ethics, and......fiduciary duties?' *Journal of Academic Ethics*, 12(3), pp.209-225. http://doi.org/10.1007/s10805-014-9212-4.

DiSalvo, C. (2012) 'Spectacles and tropes: speculative design and contemporary food cultures', *Fibreculture Journal*, (20), pp.109-122. Available at: http://twenty.fibreculturejournal.org/2012/06/19/fcj-142-spectacles-and-tropes-speculative-design-and-contemporary-food-cultures/ (Accessed: 30 March 2017).

Dunne, A. and Raby, F. (2013) *Speculative everything: design, fiction, and social dreaming*. Cambridge, Mass: MIT Press.

Elsden, C., Nissen, B., Garbett, A., Chatting, D., Kirk, D. and Vines, J. (2016) 'Metadating: exploring the romance and future of personal data', *Conference on Human Factors in Computing Systems, CHI'16,* San Jose, CA, 7-12 May, pp. 685-698. http://dx.doi.org/10.1145/2858036.2858173.

Fritz, J. (2011) 'Classroom walls that talk: using online course activity data of successful students to raise self-awareness of underperforming peers', *Internet and Higher Education*, 14(2), pp.89-97.

Gray, C. and Boling, E. (2016) 'Inscribing ethics and values in designs for learning: a problematic', *Educational Technology Research and Development*, pp.969-100. http://doi.org/10.1007/s11423-016-9478-x.

Grossklags, J. and Good, N. (2007) 'Empirical studies on software notices to inform policy makers and usability designers', *Financial Cryptography and Data Security: 11th International Conference, FC 2007, and 1st International Workshop on Usable Security, USEC 2007: revised selected papers*, Scarborough, Trinidad and Tobago, 12-16 February, pp.341-355.

Jansen, A. and Bosch, J. (2005) 'Software architecture as a set of architectural design decisions', *5th Working IEEE/IFIP Conference on Software Architecture (WICSA'05)*, 6-10 November, pp.109-120. https://doi.org/10.1109/WICSA.2005.61.

MacCarthy, M. (2011) 'New directions in privacy: disclosure, unfairness and externalities', *I/S: A Journal of Law and Policy for the Information Society*, 6(3), pp.425-512. Available at: http://explore.georgetown.edu/publications/index.cfm?Action=View&DocumentID=66520 (Accessed: 30 March 2017).

Manyika, J., Chui, M., Farrell, D., Van Kuiken, S. Groves, P. and Almasi Doshi, E. (2013) *Open data: Unlocking innovation and performance with liquid information*. McKinsky Global Institute. Available at: http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information (Accessed: 30 March 2017).

Prinsloo, P. and Slade, S. (2015) 'Student privacy self-management: implications for learning analytics', *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, LAK'15*, Poughkeepsie, New York, 16-20 March, pp.83-92. https://doi.org/10.1145/2723576.2723585.

Shaw, M. (1995) 'Comparing architectural design styles', *IEEE Software*, 12(6), pp.27-41. https://doi.org/10.1109/52.469758.

Siemens, G. (2013) 'Learning analytics: the emergence of a discipline', *American Behavioral Scientist*, 57(10), pp.1380-1400. https://doi.org/10.1177/0002764213498851.

Slade, S. and Prinsloo, P. (2013) 'Learning analytics: ethical issues and dilemmas', American Behavioral Scientist, 57(10), pp.1510-1529. https://doi.org/10.1177/0002764213479366.

Slade, S. and Prinsloo, P. (2014) 'Student perspectives on the use of their data: between intrusion, surveillance and care', *European Journal of Open, Distance and E-learning*, 18(1). Available at: http://www.eurodl.org/index.php?p=special&sp=articles&inum=6&abstract=672&article=679 (Accessed: 30 March 2017).

Taub, M., Azevedo, R., Martin, S. A., Millar, G. C. and Wortha, F. (2016) 'Aligning log-file and facial expression data to validate assumptions linking SRL, metacognitive monitoring, and emotions during learning with a multi-agent hypermedia-learning environment', *Annual Meeting – American Educational Research Association, AERA'16*. Washington, DC, 8-12 April.

Trainor, S. (2015) 'Student data privacy is cloudy today, clearer tomorrow', *Phi Delta Kappan*, 96(5), pp.13-18. https://doi.org/10.1177/0031721715569463.

U.S. Education and the Workforce Committee (2016) 'Testimony of Rachael Stickland: co-founder, co-chair Parent Coalition for Student Privacy' in *Strengthening Education Research and Privacy Protections to Better Serve Students* [Printed Hearing], 22 March. Serial No.114-43. Available at: http://edworkforce.house.gov/calendar/eventsingle.aspx?EventID=400420 (Accessed: 30 March 2017).

Wan, T. and McNally, T. (2015) 'Education technology deals reach $1.6 billion in first half of 2015', *EdSurge News*. Available at: https://www.edsurge.com/news/2015-07-29-education-technology-deals-reach-1-6-billion-in-first-half-of-2015 (Accessed: 30 March 2017).

Wen, M. (2015) *Investigating virtual teams in massive open online courses: deliberation-based virtual team formation, discussion mining and support*. Carnegie Mellon University. Available at: http://www.cs.cmu.edu/~mwen/papers/thesis.pdf (Accessed: 30 March 2017).

Wolf, G. (2010) 'The data-driven life', *The New York Times* [Sunday Magazine, 2 May, p.MM38], 28 April [online]. Available at: http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?_r=1 (Accessed: 30 March 2017).

World Economic Forum, WEF (2011) *Personal data: the emergence of a new asset class.* Switzerland: Bain & Company Inc/World Economic Forum. Available at: http://www.weforum.org/reports/personal-data-emergence-new-asset-class (Accessed: 30 March 2017).

Zurita, G., Nussbaum, M. and Salinas, R. (2005) 'Dynamic grouping in collaborative learning supported by wireless handhelds', *Educational Technology & Society*, 8(3), pp.149-161. Available at: http://www.ifets.info/journals/8_3/14.pdf (Accessed: 30 March 2017).

Biography
Michael Madaio is a Ph.D. student in Human-Computer Interaction at Carnegie Mellon University, where he studies how educational technologies can support students' interpersonal social bonds, effective collaboration practices, and learners' agency over their own learning. A former public school teacher, he received his Masters in Education from the University of Maryland and a Masters of Science in Digital Media from Georgia Tech. He uses mixed-methods, computational, and exploratory design research methods to design and understand the impact of novel educational technologies, in contexts as varied as public secondary schools and universities in the U.S. and community learning centres in West Africa.