
Learning analytics and giving students access to their data

- Moderator** Welcome to the JISC podcast. Learning analytics specialist Neil Slater has been working with a number of universities to help them with the legal and ethical issues they encounter as they implement learning analytics. Here he talks about giving students access to their data.
- Niall Sclater** Data is at the very heart of learning analytics. As much of it relates to the students themselves, there are ethical and legal considerations we need to be thinking about when dealing with it. One of these is the legal obligation to give students access to their data if they request it, but I wondered how important is it to provide access to all the data an institution has about you if you ask to see it. I asked Susan Graham, University Records Manager at the University of Edinburgh.
- Susan Graham** This is essential, I would say. There's two reasons for that: one is the Data Protection Act says that students have a right to obtain a copy of all information that institutions hold about them, so within the law if they ask they have a right to receive this. The other reason, I'd say it's [inaudible 0:01:12] essentially slightly too far, but I'd say it's very important, is that our students are adults, we should be empowering them as independent learners and making the data we hold about them transparent not only helps with data protection compliance but actually is an effective way of supporting the student and enabling them to work independently.
- Niall Sclater** Andrew Cormack, chief regulatory adviser for JISC Technologies, had an interesting qualification to this approach when I asked him whether you need to provide students with all the data you hold about them if they request it.
- Andrew Cormack** I think in principle that should be the aim. There will be some things, I think in the current education processes there are some things where there are exemptions around examinations and things like that, but I think I'd expect access to be the norm and you'd
-

be looking for exemptions, rather than the other way around and saying that secrecy was the norm. It should be collaborative and helping one another. There's no interest in the organisation processing inaccurate data, it'll either have no effect in which case it's a waste of time processing it, or it'll actually skew the results in which case it's harmful to everybody.

- Niall Sclater To put it bluntly, institutions don't always know all the data that they have about students, is there a chance that a student could sue an institution if they don't provide them with all that data?
- Andrew Cormack Sue is unlikely; we have law students, there's always a chance that they will decide to exercise what they've learned in class. Whether you'd be successful, for a suit you'd have to show damage and it has to be actual monetary damage. There is a suggestion around distress as well but it's a pretty high bar, and you'd have to demonstrate that the organisation had acted unlawfully or unreasonably. I struggle to see how that would come out of failing to provide access. Typically what would happen is the student will ask for access, if the university says, I haven't got it the student can then first appeal to the Information Commissioner who can then say to the organisation, justify your lack of provision of this, so it goes through a whole host of processes before you get anywhere near a civil case for damages.
- Niall Sclater So if a student does ask to see everything you know about them, would you need to provide all the low level transactional data or would some kind of summary of it suffice?
- Andrew Cormack In law there's limits on the cost and effort. In the new Data Protection Regulations it has been noted that with online stuff it should be easier to provide access because the data's already in a database and it's being generated by online, so why can't you get access to it? So I suspect there's going to be more of a push on that threshold. A summary might well be more helpful. It is a right to access information, it's not a right to access documents, so that might well play into that question, because it's the personal information of the individual, it's not the documents that contain that personal information.
- Niall Sclater Wilma Alexander and Anne -Marie Scott from the Learning Teaching and Web Services Division at Edinburgh University are inclined to agree.
- Wilma Alexander I think if at all it should be summarised. If we have a responsibility in that area it's to give people information that's actually understandable and useful, so a summary is going to be much more helpful because it gives us a chance to contextualise it. We know already that learning analytics data in the context of an on-campus course which really doesn't make much use of online activity, the logs in that context just don't mean anything. We're not doing anybody any favours by providing that kind of information, it's meaningless, it's just data, it's not information.

- Anne-Marie Scott We did some analysis with students on a distance based course, so as part of a research methods course inside a digital education MSc we introduced learning analytics as a research topic, and one of the things we do in that course is give them a slightly prettied-up Excel dump of their activity log, and the feedback has been generally that it's not actually that enlightening. It's a bit interesting but it doesn't actually tell them an awful lot, so I think that backs up the need to contextualise and summarise, having tested in the few students.
- Niall Sclater It's a similar situation at the Open University. Sharon Slade, Senior Lecturer in the Business School.
- Sharon Slade I think it would be more usefully summarised if they were to make sense of it, so one of the things that we're trying to do at the OU is looking at whether or not something like a student dashboard would be a sensible way to communicate what we feel is relevant to students, not as a way of withholding information but as a way of sense-making, I think. So that's one of the work streams that's going on at the moment: what data would be most useful to students, what would they find helpful in making their next study choices, etc? So that's certainly something that we're investigating with students at the moment.
- Niall Sclater Susan Graham also gave me her opinion on whether a summary of the data would be enough.
- Susan Graham Within the strict interpretation of Data Protection, if a student asked for everything the university held about them that would include the lower level transactional data. Within the learning environment, hopefully a lot of that should be accessible to the student in any case. They do have the right to ask for it, although in my experience I've yet to encounter a student who's actually interested in that information.
- Niall Sclater Would you need to make it comprehensible to the students or would a dump of the data be enough?
- Susan Graham It is a requirement that it be comprehensible, so it does actually have to be comprehensible to the student. This includes things like if there are codes, if there's some form of coding in the data that you need to provide the student with a key so that they can understand what the code is. A dump that's utterly incomprehensible would not be acceptable.
- Niall Sclater So it appears that you need to be able to give students access to pretty much anything they ask for, and ideally make it understandable. But is it ever acceptable to withhold any data from a student?
- Susan Graham It can be. Again there are two levels to this. The Data Protection Act says people have a right to ask to see information, so if they haven't asked you don't have an obligation to give it to them. If they have asked there are some exemptions within the legislation; I

don't think many of those would apply to learning analytics though. Some of the more common ones are things around issues like third party information where you can only answer the student's question by giving out information about somebody else as well, and the other one is around areas like crime, if for some reason the police were investigating a student it might not be appropriate, it may jeopardise the investigation if you were to tell the student that. However, in the context of just pure learning analytics I'm not sure what exemptions we might have, or even actually what issues we would have about sharing the information with the student anywhere. I would generally recommend that we make analytics as transparent as possible to students because the whole point of it is to support them.

- Niall Sclater Sophie Stalla-Bourdillon from the University of Southampton pointed out a couple of situations where it might be acceptable to withhold data from a student.
- Sophie Stalla-Bourdillon I suppose here you might find cases like, if by disclosing the data to the students you are also disclosing the data of other students, therefore you would have a good case to justify the fact that you can't give the data to the students, and withhold the data. Or if by disclosing the data you are also infringing IP laws, then there would be an argument that you can withhold the data.
- Niall Sclater Andrew Cormack from JISC also comes up with some scenarios where you might withhold data from a student.
- Andrew Cormack Yes, there are one or two exemptions to subject access right; I can't do the full list of them but things like if disclosure, if were in the process of investigating a crime, disclosing to the individual that you were in the process of investigating whether they'd committed a crime might well harm that investigation, so there are some exemptions there. So is it ever? Yes, but the conditions in which it is permissible are fairly tightly drawn.
- Niall Sclater Presumably it's a good idea for an institution to make it clear to everyone what kind of situations they would potentially be withholding that data from individual students?
- Andrew Cormack Absolutely, try and write the processes before you have somebody jumping up and down saying, the Information Commissioner demands an answer in 20 working days. That's not a good situation to be writing a process or a policy in. Work out as much as you can in advance so that these things are all clear and that a dispute doesn't arise.
- Niall Sclater Sharon Slade at the Open University does think we need to be careful about what we provide to students from a pedagogical or at least a motivational point of view.
- Sharon Slade I think we need to be careful when we're talking about predictive model output for example, because these are only probabilities, they're not definite outcomes. So we need to take huge care, both in the confidence levels placed around any model and how we communicate those probabilities to students, if we do at all actually, and I'm not a

big advocate for doing that. I think we should be using models like that to inform us and to inform students and to provide advice in how to best move forward, but I don't think we should ever say to any student, we think you've got a 10% chance of passing your module.

- Niall Sclater Obviously it's important for a number of reasons to keep your data accurate and up-to-date. I asked Susan Graham what happens if a student discovers something wrong with their own data and the institution doesn't have the mechanisms in place for them to correct it.
- Susan Graham The very simple answer is that there has to be a mechanism to correct it. If information is inaccurate we should not be holding it, we should not be storing it and we should not be using it. There are areas of interpretation, shall we say, which can affect this issue, so for instance you might have data in contention where a student says it's inaccurate but the institution believes it's correct, and in those kinds of situations best practice is to note that the student actually disputes this information. But if it's incontrovertibly inaccurate then we have to have a means to correct that.
- Niall Sclater Andrew Cormack has a similar viewpoint.
- Andrew Cormack If it's something that's observed then actually can you correct an observation? You can say, don't use that, but I think actually, if logs of the Wi-Fi network say that the student was in place X, maybe the student could say, I wasn't, please delete that, I find it hard to see that the student could say, actually correct that to a statement I was in location Y. That seems highly questionable, so there might be some cases where there's justification for suppression rather than correction. But I think, as I said, it's not in the organisation's interest to keep carrying, keep processing data that's known to be inaccurate; either don't process it or correct it. So I'd expect there would be a mechanism to correct it.
- Niall Sclater I wondered how Edinburgh University was making any data and analytics available to the students themselves, and I asked Wilma Alexander how students were being given access to information about themselves.
- Wilma Alexander We already do a great deal in that the student record is really very open and transparent for the students themselves, so they have access to that just as a matter of routine, it's been pretty much automated and where appropriate they can edit that data themselves. Obviously they can't change their grades but within reason that's already all there for them. Indeed, the main thrust of the data from VLEs project that is just drawing to a close has been to provide student facing data from the VLEs where that has in general, and I'm generalising right across the board, but generally it's been available to staff for ages but not available to students. So we've got quite a strong strand and quite a good record in making things student facing, and I think we'd continue to do that wherever we can and wherever we feel it's appropriate.

That's true also of our personal tutoring system which our students also have access to through the student record, which is the kind of log of their meetings with their personal tutor. They can add notes to it, their personal tutor can add notes to it, and each can see it as well, so if there were interventions triggered by the data they might be recorded in places like this, so not just the data but some of the outcomes of the data would be visible to them.

Niall Sclater I also asked Sharon Slade if the Open University was providing access to the analytics to the students directly.

Sharon Slade That's a tricky question. I think at the moment we're bound by the existing rights the students have to access any data that's stored on them. I don't think that existing legislation includes the models themselves or the predictions that drive the model output. So how we would manage that in a practical way is a big question moving forward. It's certainly being discussed; I'm not sure whether we've bottomed it out.

Niall Sclater I suppose the issue is that the analytics themselves are a new form of data about students which they could theoretically demand access to as well under the Data Protection Act?

Sharon Slade It's whether or not it's stored and held, I guess. That leads to a whole other set of questions as to whenever we run a model do we store that data against a student record, and at the moment I don't think that we do. So if we were required or if we thought it were a good idea to start doing that and to be able to show that to students we would need to rethink how we're managing outputs from existing models. Things that are based on not predicted models but basic tracking of the students, do they live in a certain area and do they have a particular disability and have they failed their last module sort of tracking, all of that data is held and any interventions which are triggered as a result of those sorts of combinations of binary data matching, they're all stored against the student record anyway. So if we did make a direct contact with the student based on they hadn't submitted their last assignment, for example, that's stored against that student on their record anyway, so they would be able to see that and why they had received a particular contact. But the predictive modelling is a different issue and I don't think we've thought that through at this stage.

Niall Sclater But if the analytics themselves are just temporarily presented to personal tutors, for example, and not subsequently stored anywhere, there presumably would be no obligation to provide that subsequently to students, because you can't provide something that doesn't exist. I put that legal conundrum to Susan Graham.

Susan Graham This is an interesting question. I'm thinking there's both a technical and a practical answer. Technical answer is twofold. The first point is data protection only applies to recorded information, so if the information isn't recorded data protection doesn't apply to it. However data protection does talk about automated decision-making and I would

suggest that if you're running an algorithm that is showing a personal tutor information about a student, then you've got automated processing going on and data protection does require us to explain to a student what kind of automated processing goes on and how that is used. So there are legal rights in that area. My practical question is that having information that flashes up for a tutor and isn't stored and isn't re-accessible and isn't available to the student, I'm questioning how helpful that is in the long term, because, for instance, in terms of exploring what the content of the discussion was, in the future if you need to refer back to it, you can't refer back to it. The student hasn't seen it, so the student doesn't know what it is and can't discuss it, so practically I'm wondering how helpful that kind of unrecorded information would be.

Niall Sclater

Finally, I wondered if it was important to explain how the analytics work to the students themselves. Sharon Slade.

Sharon Slade

I think realistically not in huge detail. I think we need to be able to explain the purposes and we need to be able to demonstrate, probably in a peer review way, that the approaches that we're taking are robust. I don't think we anticipate being able to explain in detail to students and staff how those models work, I'm not even sure I would understand how most of those models work, only a very cursory understanding, but we do need to be able to explain what we're doing and why, and be able to, I think, demonstrate that the approaches are robust.

Niall Sclater

Anne-Marie Scott and Wilma Alexander from Edinburgh are also sceptical that you could explain all the processes around learning analytics in full to staff and students.

Anne-Marie Scott

I think it's useful for transparency. I think we always hit interesting questions when perhaps it's commercial software around explaining algorithms. I think for some of our staff, particularly those who are active in the field or in related fields, a good portion of establishing the credibility of what we're doing can be done by being very thorough in our explanations of these things. I think that there's a difference between explaining them and being clear about them and people having to understand them, to actually use whatever it is we produce or offer to them, and I certainly think that at scale there is no way that you could train up our entire staff and student body in such a way that they could get to grips with the nitty-gritty detail of it. So I think the explanations have to be available but they mustn't be a mandatory part of using whatever it is that's provided.

Wilma Alexander

I think the trick is going to be to find the right level of particularly staff development and training that's going to help staff understand what these tools that we offer are now not good for, but at the same time, as Anne-Marie says, not to overdo it. Again, it's clear from the information that we've collected so far that there's an enormous range of information literacy across the piece for both staff and students, so I think there is some responsibility resting with us to be as transparent as possible and to do our best to remove the possibility of misunderstanding what this data may be telling people.

Niall Sclater

So we've heard from quite a few different experts on the issues around giving students access to their data. The Data Protection Act makes it clear that if they ask for information held about them, we do need to provide it to them, and that probably means pretty much everything we know about them. However, the low level transactional data is unlikely to be of any use to students and we need to be able to summarise it for them in a useful format.

There are theoretically a few situations where it may be possible or advisable to withhold data from students, but these are limited. There's some uncertainty about whether we should provide all the analytics though to the students, it might damage their motivation for example to be labelled as likely to fail.

I think the conclusion here is that transparency probably trumps most of the other issues. These subtleties are still being worked out, but some institutions are taking the position that whatever staff say about individual students on a dashboard should also be presented to those students. Certainly though we should be explaining as clearly as we can to students what is being done with their data, how the analytics processes work and what interventions are being carried out on the basis of those analytics.

[ENDS]