Supporting Scalable Data Sharing in Online Education

Stephen Cummins
Computer Laboratory
University of Cambridge
Cambridge, UK
Stephen.Cummins@cl.cam.ac.uk

Alastair R. Beresford
Computer Laboratory
University of Cambridge
Cambridge, UK
Alastair.Beresford@cl.cam.ac.uk

Ian Davies
Computer Laboratory
University of Cambridge
Cambridge, UK
Ian.Davies@cl.cam.ac.uk

Andrew Rice
Computer Laboratory
University of Cambridge
Cambridge, UK
Andrew.Rice@cl.cam.ac.uk

L@S 2016, April 25-26, 2016, Edinburgh, Scotland UK
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893376

Abstract
Online educational tools often generate learning data, and sharing such data between tutors and students can often improve learning outcomes. Unfortunately the process of sharing learning data today is not always transparent to students. Our aim is to improve the transparency and user control aspects of sharing data whilst maintaining the educational utility of data sharing between tutors and students. To do so, we start by surveying the possible methods of sharing data, and we use this to design a token-based scheme for facilitating data sharing. We implemented our scheme and observed it in use by 7,798 students over the course of one year. We find that our proposed scheme provides a good balance between transparency, user control, educational utility and scalability.

Author Keywords
Privacy; Authorisation; Review; Data Sharing

ACM Classification Keywords
K.3.1 [COMPUTERS AND EDUCATION]: Computer Uses in Education

Introduction
A key feature of supporting online learning at scale is to ensure safe and reliable data exchange between learners and any collaborators such as tutors or classmates. The is huge

value in the interaction between students and tutors, students and their peers, or indeed for tutors to make informed and targeted interventions during the learning process [1]. This human aspect is amongst the most important for providing motivational support, as well as much needed access to someone with an interest and expertise in the subject.

Learning data captures the personal journey of the student, and often includes metadata such as their name and contact details. Therefore a secure mechanism of sharing learner data is an important consideration when designing online platforms which support informal relationships between students and their tutors and peers.

Existing Models of Sharing Data

We have identified five different models which platforms could use to authorise and orchestrate the sharing of data between students and tutors. Not all these models are commonly used in the education sector today. We label the five models as *Public*, *Reciprocal*, *Anonymous*, *Enrolment*, and *Token*.

The *Public* model makes all data available to any user of the system. An example of this model in practice is the micro-blogging site, Twitter. Posts on Twitter are accessible to all users and non-users equally, with some profile information also publicly visible.

The *Reciprocal* model is used by platforms such as Facebook and LinkedIn. Users of these platforms search a public directory of user profiles to request access to more detailed information. If a request is granted, a bi-directional link is established such that each user may view detailed information of the other.

The *Anonymous* model aims to remove personally-identifiable information from any shared data. For example, previous studies have proposed the peer-wise sharing of marked work and any associated tutor feedback between students [2].

The *Enrolment* model involves users registering for a particular service, or joining an institution and thereby granting access to their data to other members of the service or organisation. Examples in the education sector include EdX, Coursera, or Moodle, where course registration is directly connected to a default data sharing policy.

The *Token* model relies participants sharing authorisation tokens between themselves to grant access to data items. Examples include shareable URLs to enable document collaboration using Google Docs. We have developed a token-based approach for our platform where a tutor can set students work by sending students a digital token generated by the platform. Independently, Khan Academy has recently adopted a token model to share learning data.

The enrolment model and the token model support different approaches to online learning. Under the enrolment model, a centralised approach to course delivery is required because course registration is directly couple with a structured approach to content delivery and an associated sharing policy for learning data. As a result, students trust the course provider to enrol suitable tutors, limiting transparency on what happens to their data. This approach also limits scalability – neither external tutors nor students cannot borrow or reuse course components delivered on these platforms in other contexts. Indeed, a common policy is that if a user wishes to withdraw their consent for data sharing they must deregister from the course and they can no longer access any content. An important benefit of the enrolment model is that students do not need to explicitly authorise access to their learning data – the platform does this automatically.

	Dimensions		
Platforms	*Model of Data Sharing*	*Who is granted access*	*Ability to withdraw*
Coursera	Enrolment	Organisation	Yes[†]
EdX	Enrolment	Organisation	Yes[†]
Isaac Physics	Token	Individual	Yes
Khan Academy	Token	Individual	Yes
Moodle	Enrolment	Organisation	No
SWATT	Anonymous	Group	No

Table 1: Classification of Learning Environments: [†]Indicates no further access to course materials if the user revokes access to their data.

The token model supports a decentralised approach to online learning. Students and tutors can bring together material in any order to build a course or programme of work. This approach is more scaleable because the delivery of content and provision of support is not contingent on any particular party. The disadvantage, when compared with the enrolment model, is the overhead associated with building the course, and configuring the data sharing policy – something which can no longer be automated since there is no centralised register of trustworthy tutors. In the token model, users are able to revoke access to their data at any time without penalty. Furthermore, course content is typically available for free without any connection to a tutor or course timetable or other structure.

Table 1 summarises six online learning platforms. With the important exception of Khan Academy, the popular platforms currently use the enrolment model.

The Isaac Platform & Data Sharing

Isaac Physics is a project funded by the Department for Education and Skills in England which Supports the transition between school and Higher Education. In particular, the project focuses on developing students' problem solving skills.

In order to support a decentralised model of online learning, we need a mechanism to support the sharing of learning data between tutors and students. We therefore implemented an out-of-band token-based model. First the tutor, teacher, or coach (T), asks the platform to generate a token for a specific activity or piece of work. Second, T shares the token with one or more students (S).

When a token is entered by S, the platform asks them if they would like to share their learning data with T. To provide some confidence to S of the provenance of T, the system exposes the name and email address of T. This allows S to be confident that they have entered the correct token.

The token itself is an eight-character alphanumeric code, for example HA12-PG8Y. A short token is important since it can be easily written by T on a whiteboard or copied into a notebook. The consequence of this is that rate limiting is required on the platform to prevent brute-force attacks in which a malicious S attempts to join an activity run by a tutor T who has not shared the token with S.

Once access has been granted by S, then T can view S's learning data, including any answers provided by S. S may revoke access to T, in which case T is informed that they no longer have access to view any of S's learning data.

Results and Discussion

The Isaac platform has been live since September 2014, and after one year, Isaac had 10,565 registered users (4,878 Male, 1,957 Female and 3,730 unspecified) and 105,124 anonymous users generating over 2,197,010 data points.

For the purpose of this paper, we excluded all users who we know are teachers in schools and all users who have not yet made at least one question attempt, either correctly or incorrectly. A total of 7,798 out of our 10,565 users are eligible under these criteria.

In total, 5,455 students (69.95%) engaged with the data sharing mechanism on the platform and 2,343 (30.05%) have not. A total of 27 (1.1%) students used the revocation option to prevent a previously authorised tutor from viewing their data.

We analysed the usage data of students who connect to a tutor (coached users) when compared to those who have not (uncoached users). Coached users attempt a Minimum (m), Median (M), and 90th percentile (P_{90}) of $m = 0$, $M = 55$, $P_{90} = 195$ questions, whereas uncoached attempt $m = 0$, $M = 16$, $P_{90} = 99$.

On average, coached users eventually answer more questions correctly ($m = 0$, $M = 22$, $P_{90} = 74$) compared to uncoached users ($m = 0$, $M = 5$, $P_{90} = 35$), but they also tend to answer more incorrectly ($m = 0$, $M = 9$, $P_{90} = 31$ vs. $m = 0$, $M = 3$, $P_{90} = 16$). This may be because coached users are aware that a tutor will see their end result, and not necessarily how many attempts they made at any particular question. This data does not necessarily suggest that the act of sharing learning data with a tutor directly leads to a difference in interaction; further work is needed to understand this relationship.

Conclusion & future work

We have reviewed a number of different strategies for sharing learning data in online learning platforms and we developed a token-based approach to support controlled sharing of learning data to support a decentralised learning model.

We found a difference in behaviour between users who have shared learning data with a tutor and those who have not. Students supported by a tutor attempt more questions than those who do not, which may be an indication of greater motivation. There are a number of possible explanations. A tutor who assigns work and monitor progress encourages engagement. Furthermore, access to external, tailored and targeted support maintains momentum and promotes student progression.

We intend to investigate the relationship between sharing learning data with tutors further. In addition we intend to enabling any users of Isaac to create data sharing links with each other, with the aim of fostering a peer-learning network.

REFERENCES

1. Michelene T. H. Chi. 2009. Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science* 1, 1 (Jan. 2009), 73–105.
2. Stephen Cummins, Liz Burd, and Andrew Hatch. 2011. Investigating shareable feedback tags for programming assignments. *Computer Science Education* 21, 1 (March 2011), 81–103.