Privacy and P2PU

Heather Ford and Alex Smolen

Abstract: We apply the results of an ethnographic study of educational communities and a new privacy risk assessment methodology to understand online educational privacy risks and develop a series of design alternatives for P2PU (Peer-to-Peer University). In doing so, we attempt to balance the system’s principles of accessibility and openness while promoting student privacy. The design alternatives are illustrated by a set of prototypes that demonstrate how they could function.

Introduction

On the 22nd of February 2011, P2PU co-founder Stian Haklev made the following proposal to the P2PU research community by:

I’m hereby proposing that P2PU officially decides that all user interaction data are completely public. This includes not only all the material that has been posted, but also when users access the websites, what they click on etc.

Haklev's statement was followed by a lively debate about whether or not to proceed with the suggested functionality. One contributor noted that making every click public by default, in a way no other major site on the web did, could lead to a privacy backlash.

In adopting a ‘completely public’ policy can you guarantee that no personal data could be inadvertently released due to an oversight or lack of a review process?

This kind of question is at the heart of current debates about privacy. Some people may see an act of information disclosure like this one as being acceptable and
necessary for any kind of social web application, while others are may be adamantly opposed. Some may make bold statements about privacy being dead, and suggest that norms have changed, but privacy concerns seem to be omnipresent in the mainstream consciousness of the information age.

In some cases, the argument for disclosure and against privacy has obvious justifications. If a company is in the business of behavioral advertising, they have a vested interested in promoting the disclosure of information that helps inform them of consumer behavior. If a government wants to oversee its citizens, it will argue for the safety benefits of surveillance technology. P2PU, in this case, wishes to enable greater access to its system by researchers in order to improve it, but in doing so, is P2PU favoring the interests of outside parties to the detriment of those central to its mission?

We’ve been interested in P2PU as a kind of experimental laboratory for our privacy research for the last semester. As suggested by the mailing list exchange above, and further discussions described below, there is a considerable tension between the values of openness and privacy when the openness principle is taken to its ultimate extreme. Accessibility and openness are central to P2PU’s founding and development, with nearly all exchanges and data on the site being publicly accessible and available under a Creative Commons license. But is the openness principle making students inadvertently feel as if they are being watched, either by an authority figure or an unknown audience? Is this one of the factors contributing to high dropout rates on P2PU? Is the lack of intimate spaces preventing the community building necessary for successful learning?

We wanted to determine what sort of privacy concerns users may have while engaging with P2PU, and to suggest a set of design alternatives that may help preserve user privacy while incorporating other goals of actors in the system. We started with a contextual analysis of why privacy is important for education and what norms of information disclosures support educational goals. Next, we
performed an iteration of the privacy design analysis methodology using information gleaned from interviews and textual analysis of P2PU. This methodology is a structured way of looking at the different types of information disclosures that occur in the system in order to find potential privacy issues. The result of this process is a set of design alternatives that we propose the P2PU community consider.

Background
As background for our analysis, we describe P2PU stakeholders’ understandings of privacy, and put forward a cumulative definition of how privacy is broadly conceptualized. By rooting the analysis in the understandings of the community, we hope to show that P2PU has the seeds of the solution in their discussions on these issues and that we merely provide them with a framework for thinking systematically about what the solutions might be.

Privacy in P2PU
We asked one of the P2PU co-founders to tell us whether privacy was important to P2PU.

I think privacy is important. Period. Not just for P2PU. I think the individual should be in full control about the information they want to share about themselves and their activities. For us, things like selling off peoples’ email addresses that would be absolutely inconceivable at P2PU because we would feel like we’re taking peoples’ personal information and using it in a way... I mean we could obviously have these very sneaky end user agreements where you never actually know what you’re clicking on and what you’re accepting, but we could do all that stuff. But that would be totally counter to the way that our community works where everything is based on trust and if you start abusing that trust we will lose our community and we will lose P2PU. So for us as a project it’s important that people feel that we’re treating their information with respect and care. At the same time, we are in a fairly good position because there isn’t actually that much information that we have that not everyone else has access to already.
Privacy, because it is a complex concept to understand, is often described in this fairly limited way. In this case, the P2PU co-founder describes privacy in terms of very specific types of information – in this case, email addresses – rather than tied to specific social interactions. Another preconception about privacy is that merely notifying users of what information they intend to make public can solve the problem. In the discussion about whether to enable interaction data to be released to the public, Haklev said that one concern could be that “students have not been properly informed about this.” He said that they could mitigate this by providing “clear information about this in non-legalese (to) be displayed and agreed to when students apply to courses” and provided the example:

I realize that all my interactions with the P2PU site will be logged, and made publicly available. All my contributions will also be licensed under a CC license.

The discussion that followed indicated that the community collectively understood the privacy problem much more broadly. Some pointed out the problem was that the data at stake was not data that is currently available but clickstream data, the traces that a user leaves when interacting with the system. It would be difficult to conceptualize this data, particularly because it is invisible to most users. Only administrators see how long a user interacts with the P2PU website, which pages they visits and where they are physically located when they log into the system, for example. When some community members indicated concern, others kept returning to the current open data on the site, believing that what was being proposed was little more than developing an accessible data file of what was currently available. According to this line of thought, if some people were concerned with this, then they should be concerned with all the other interactional data that was being released. As one community member wrote:

These are open forums already. Everything is already being tracked (try and search Google for a phrase from the forum) and can be analyzed if anyone cares.
Haklev pointed out a second concern in the forum. He said that that doing may “stop people from participating, it would negatively affect learning processes, it would give us a bad reputation.” Responding to a request for exactly what “unintended consequences” might result from aggregating user data and making clickstream data available to the public, one community member wrote that:

The unintentional consequences I think largely involve social factors like privacy, trust, encouraging small group learning and formation, and simple risk mitigation.

This got community members thinking – not just about the proposed initiative but the effects of the current open system and ways to mitigate some of the harmful effects that might be taking place:

I would say that, quite arguably, any chilling effect from "publicity" is already fully in effect. However, people are free to sign up with a pseudonym, and most appear to do that. P2PU could offer a sign-up service that offered to more fully "mask" users online, and I think this would be a good idea. It could also offer a sign-up service that found a way to associate real-world identity to someone (e.g. send them a thumbdrive with a given cryptographic key on it) and I think that could also be a good idea.

This exchange showed that the discussion was driving towards a sophisticated understanding of privacy. Starting off with a very limited view of privacy and relating to specific types of information, the community was starting to under privacy as tied to specific roles and relationships. We approach privacy from this viewpoint – not as a feature of certain types of information, but as the ability to control information, or who knows what about you (Livingstone, 2006). In doing so, we understand that students must and do disclose personal information in order to sustain intimacy in the learning environment, but they wish to be in control of how they manage this disclosure. From this vantage point, P2PU-ers started to question the open-everything perspective, asking themselves:
Personally I think the question shouldn't be "should the data be open?", but rather, "what data, precisely, should be open, and why?"

The community was clearly grappling with the question of how they can remain true to their identity as an organization committed to openness and accessibility while still respecting the privacy needs of their users. But they needed to know what were the negative effects of making all data open to the public and how to mitigate the effects of such openness. A balance needed to be struck but they seemed to not have the tools to strike that balance.

**Negative effects**

We wanted to know whether students believed the current system to be adequate for their privacy needs. In looking at the system, we identified features that may indicate a need for greater privacy. In one course that we looked at, the course organizer asked students to keep a learning journal to reflect on what they were learning:

> One of the main expectations for this course is that participants will keep a journal throughout the 8-week period. This journal can be a blog, or it can be a private journal - the purpose is to encourage you to begin and maintain a reflective practice, and it is up to you whether or not you would like to make your writing public.

Only four out of the forty students responded to the course organizer’s post. One student responded with a link to a blog that he had set up specifically for the course, another linked to his personal blog in which he had made a category specifically for the course, yet another wrote: “I have a private journal that I keep and if it is okay I will use that” with a fourth (the only woman among the four) saying that she would “use a private journal until I catch up and can figure out how to blog.” This interaction indicates the need for private, or at least more intimate spaces
controlled by students outside of what was being offered by the open P2PU platform. As the course organizer indicated in an interview:

Overall, I think that the more open people are, the more they are able to contribute and the more valuable discussion and exchange of ideas becomes. However, personal (private) reflections help participants to make connections and understand the material, and I think this is critical.

**Privacy in Education**

These personal reflections that the course organizer highlighted also reflected a need for a variety of interaction spaces. These interactions could be more useful and common if they were shared with a few trusted peers, or only with the course organizer. In Goffman’s language, personal reflection enables us to “try on” a diverse range of “costumes” or identities in order to choose who we want to be and which communities we want to join (Lave and Wenger). In the formative stages of the learning process it is important that “audiences” are limited in order for the student to feel safe enough to experiment.

The need for safe spaces for experimentation in education is echoed by our ethnographic research of a graduate school in the United States to understand their relevance to learning and privacy. In this study, we find that the “spaces between” – the classroom before and after class, the student lounge and computer laboratory – play an important role in the learning experience because it is where students can construct knowledge with their peers and practice the performance of new identities. The fact that these spaces are located outside the purview of those in authority and that they enable students to choose who they can be intimate with is consequently central to the success of these spaces. When private spaces are unavailable, students “hack” the system by opening up their own private channels resulting in harms including exclusion, identity crises and self-censorship.
In a similar way, we can see the example of the learning journal as a way in which P2PU users need to hack the system for privacy reasons. Without the technology to provide more intimate spaces, users turn to blogs, systems that enable only completely public viewing or completely private accessibility, thus excluding the opportunity for more intimate social spaces. In addition, the use of private blogs means that those who don’t have the know-how about how to set up those blogs are excluded, forcing them to use analogue technology when what they want to do is be able to be part of the learning community. As the female student wrote about her personal diary, “It is only temporary until I figure this out and get caught up. I do want to share and be a part of it all.”

We wanted to enable a system on P2PU that didn’t exclude students in this way. We wanted them to be able to meet the needs of more intimate spaces while retaining their goals of openness and accessibility. To do this, we needed a more systematic, transparent approach.

**Approach**

There are a variety of approaches for identifying and describing privacy concerns in a system. They range from the privacy impact assessment used by the federal government to sophisticated privacy modeling languages. As part of our research, we developed a methodology and tool for eliciting privacy design issues and mitigations. For P2PU, we were wanted to ensure our analysis met three criteria: that it was a) contextual b) holistic and c) practical.

**Contextual**

Although innovative online systems like P2PU are new and change the power dynamics of learning in fundamental ways, they still fall within the context of education. Helen Nissenbaum’s theory of “contextual integrity” suggests that privacy is about preserving norms of information disclosure. This proved useful in understanding the role of contexts – what Nissenbaum calls “structured social
settings” – in designing for privacy. Nissenbaum believes that we need to look to formally established spheres (in this case, the tertiary educational institution) in order to understand the key characteristics of its corresponding socio-technical systems and practices (in this case, the P2PU system) (Nissenbaum, 2010). In addition to being a site based on educational goals, P2PU also has functions similar to a social networking site, and more broadly is a site on the World Wide Web. Understanding these multiple contexts of P2PU helps us establish the norms governing information flow, a central tenet of our understanding of privacy.

Holistic

The predominant approach for dealing with privacy has, until recently, been the “notice and consent” model. Companies tell users what they do with their information and the user makes a choice about whether to use the system or not. But this is not always an effective approach since people rarely read these privacy policies. If P2PU were simply to update its privacy policies to include information disclosures that its users may find objectionable, it would not be serving its users, since these statements may go unnoticed.

Our methodology looks at a comprehensive set of disclosures and potential design alternatives. It considers information disclosures that are problematic not just from a legal viewpoint, or from a security viewpoint, but from a broader social perspective. Finally, and perhaps most importantly, our analysis is not solely focused on achieving privacy goals. It also designed to respect the other goals of P2PU stakeholders. When a design alternative represents a tradeoff in these goals, then at the very least the costs should be well understood and transparent. This is critical in building privacy solutions that are grounded and relevant to the system’s priorities.

Practical

We believe that much debate about privacy occurs in silos. Privacy advocates talk in their circles about their dissatisfaction with current systems in terms of privacy needs, while system designers often find their recommendations irrelevant and out
of touch. Within organizations, the person responsible for privacy may reside in the legal department and not share a common language with the person responsible for implementing the technology. We wanted to develop a way to bridge the gap between theoretical privacy discussions and real-world system implementation. This tool attempts to bridge that gap with policy-focused input and technical design outputs.

Privacy Design Analysis

To analyze P2PU privacy and generate a set of design alternatives, we turn to the aforementioned privacy design analysis methodology. This section discusses the results of that methodology applied to P2PU.

Step One: Determine Actors

The first challenge in determining potential privacy issues in P2PU is to understand the actors and their roles in the system. Understanding the different actors helps elicit who may be disclosing data, and to whom they may be disclosing it. Correctly aggregating groups of actors can be somewhat difficult because there may be slight differences between actor types who may or may not be material in the analysis. Generally, if groups have significantly different goals, or there are power differentials between two groups, then they should be considered separately.

In the context of P2PU, the following groups of actors were identified:

Users

Users in P2PU are people who register for the site and contribute, both as course facilitators and as students. In initial runs of this analysis, these two groups were considered separately, but it soon became apparent that their privacy concerns and goals largely overlapped. Since this is a peer-to-peer model of education, there isn’t the same power differential between educators and learners. People who participate in classes are encouraged to lead classes, emphasizing the fluidity of the educational roles and the unity of purpose between the two. Some users may
participate and lead a large number of courses, while others may register but not actively participate in courses. However, since they have identified themselves to the system, they are no longer considered the “public”. Additionally, some P2PU organizers may be users, but their goals in that context are different.

Users are the group whose privacy concerns are explored in our analysis. Generally, users are the ones who disclose data to the system and other actors are those who receive data from the system.

**Organizers**

Organizers are the administrators, developers, and backers of the P2PU project. These are people who are stakeholders in P2PU’s success and contribute to the project in some material way. They are distinct from users because they are responsible for defining the P2PU architecture and practices, and are in many cases privy to the data submitted by users. Similar to users, there are many different “types” of organizational actors, such as developers, administrators, and evangelists, but they generally share the same set of concerns about P2PU.

**Researchers**

In the context of P2PU, researchers are the people who intend to use data from P2PU in order to create and publish research about online education. Since P2PU is an “open” educational environment, it represents a fertile ground for data about how students engage in online education. This data is useful for researchers who hope to understand and improve online education system including P2PU. Researchers tend to belong to educational institutions, and may interact with organizers through the P2PU researcher mailing list.
Public

The public actor is defined as the people browsing the Web who visit P2PU. They may want to browse a course anonymously, or they may be looking for information about a user. Anything disclosed on the Web is available to the public, and may be available to entities such as search engines and web archives that store and index publicly available material.

Others actors not included

In the context of defining the relevant P2PU actors, several actors were considered but ultimately excluded because they didn't directly affect the educational disclosure relationships that we were primarily concerned with. These included third parties like governments and ISPs as well as local network users who may be able to view users traffic. While a more extensive privacy analysis may consider these actors, and the potential privacy threats they pose, these concerns are not especially specific to P2PU and there judged to be less important in the analysis.

Step Two: Identify data disclosed by actors

The next step in the process is to identify the sensitive data that is disclosed by the actors in the system. In the case of P2PU, the set of sensitive data is composed only of data that is disclosed by users. While presumably organizers, researchers, and the public disclose some form of data to the P2PU system, none of this data has the same sensitivity as user data. This is primarily due to the structure of the power relationships – organizers create the rules of the system and have access to all (or most) of the data, researchers don’t necessarily disclose any information directly to the system, and the public do not directly identify themselves to the system, so any data they leave behind is mostly anonymous. Users, on the other hand, disclose significant amounts of data to the system. This data is grouped into a set of categories to ease the analysis process – if data has the same general set of concerns, then it makes more sense to treat it as a single type of data. The following categories of data were identified:
Profile information
Profile information includes user name, full name, email address, location, bio, and profile image. All of these pieces of data may be used to uniquely identify or otherwise describe a user in real life. One piece of data not included in this category is the password that should never be disclosed to anyone, under any circumstances, and its disclosure represents a pure security concern as opposed to a privacy concern.

Social connections
Social connections are tracked in P2PU similarly to other social networking and media services, where users can follow and be followed by other users. This information includes both the number of followers and followees, as well as the list of users who are following and followed. Users in P2PU are enrolled in courses that they “follow” so the social connections includes information about courses that users in P2PU are involved in (either as course organizer or student).

Posts
Posts are messages that students publish as members of a course. They may be in response to a task, and are typically a response to some sort of question. In the previous architecture of P2PU, this functionality was structured using a forum-based interface, while in the current interface it is based on a threaded comment format. This is the primary place for learners to interact as a group on the P2PU platform.

Private messages
In addition to posting on course pages, users can send private messages to each other. These messages work similarly to Facebook messages or Twitter direct
messages, which are understood to be private except to those whom they are addressed to.

Clickstream activity
Clickstream activity is the set of data that users disclose as they use the P2PU site. This includes pages visited, time that pages were visited, login and logout activity, and other information about a user’s session. Anything that ends up in a web server log could potentially be part of the clickstream activity data set.

Step Three: Identify actor goals
The next step is to identify the goals of the users of the P2PU system. In many cases accessing the publicly available documentation from the P2PU project websites and mailing lists enabled us to fairly easily identify these goals. In other cases, our understanding was shaped by interviews that we conducted with P2PU staff and organizers. The following set of goals attempts to list the major concerns, even when these goals are somewhat abstract. In the case where goals are conflicting, they can be broken down into sub-goals. This decomposition proved unnecessary in this case.

Users
Learn about subjects from peer contributions
The primary goal for P2PU students and coordinators is to gain competency in particular subject areas, and secondarily to broaden their social network. According to P2PU co-founder, Phillip Schmidt, users were initially equally interested in the idea of peer to peer learning as they were in learning, whereas now, ‘people are coming to the site specifically because they are much more interested in actually learning something’. Users therefore see P2PU as a way to get an education and broaden their qualifications.
Organize courses that are engaging and enable relationships to be formed
In addition to learning, some users want to organize courses for other users. Because the majority of course organizers are not paid by P2PU, their motivations are largely philanthropic. For this reason, a key goal for course organizers is to run a course that is engaging and that elicits responses from students. Many course organizers are doing this work because they are passionate about a particular topic and want that passion to transfer to students. When students drop out or don’t respond they may become frustrated or disconcerted.

Avoid revealing embarrassing or otherwise harmful information online
Privacy concerns are prevalent among users of social networking services. Students have specific concerns regarding how potential employers and respected peers view their competency in educational settings. Also, since P2PU has a privacy policy that dictates what it can and can’t do with users information, users will certainly have an expectation that their information will be used in a way that doesn’t violate their privacy. This goal is central to the analysis: if no one had the goal of protecting their privacy, then the analysis wouldn’t yield any results.

Organizers
Create an open environment for learning in order to enable participation, accessibility, innovation and accountability.
P2PU describes openness as it primary core value, followed by community and peer learning. Openness is not an end in itself for P2PU; openness is seen as a way to enable participation, accessibility, innovation and accountability and it is these sub-goals that we need to be aware of when assessing the impact of information sharing on these goals.

Bring learning materials to as many people and to as diverse of a range of people as possible.
P2PU wants to enable broad access to learning. This is demonstrated in its “learning for everyone, by everyone” tag line. According to Phillip Schmidt, P2PU should become useful to a more diverse audience. “I think we need to be much more useful for people who don’t live on the web already; who don’t use these tools everyday; who are not geeks already (and geeks here in the broad sense – not just technology, you know: learning geeks). People who would choose p2pu as an opportunity to learn as an alternative existing formal education. So people who would usually go to formal education, we should be able to give value to them so that they can learn with other people in an open way.” Thus, diversity is as important a goal as increased numbers of students for P2PU.

Ensure learning material is high quality

P2PU organizers want to ensure that the learning material on P2PU is high quality. This is specifically mentioned on the web page, which suggests, “P2PU enables high-quality low-cost education opportunities”.

Support research that leads to better understanding of open educational resources and P2PU specifically

P2PU “strongly supports both research activities under the auspices of P2PU as well as research activities about P2PU ” and priorities research that “drive(s) understanding and lead(s) to action”. According to the organization, “we want research to improve and increase open social learning opportunities. We support research that leads to better understanding of the overall OER ecosystem, but we have a strong interest, and can commit more resources to research that also directly helps P2PU.”

Research community

Conduct research and experiments that provide insight and guidance
Researchers who use P2PU for research material are intending to create and publish research about online education that leads to increased insight and advancements in their field.

*Follow ethical guidelines*

According to the P2PU research guidelines, “[r]esearchers agree to be respectful of the privacy and time of everyone involved, especially individual users who are approached for data collection.”

*Public*

*Benefit from courses without directly participating*

Members of the public may visit and use P2PU in order to learn about subjects. The assumption here is that others learn from watching the interaction as well as reading the syllabus and materials. According to Phillip Schmidt, “[W]e don’t know if they’re learning anything but we do know that a lot of people click on the courses who are not in the courses. If you look at the course participation even within the people who sign up for the course many more sign up than actually participate but still some of them log into the site occasionally.”

*Step Four: Flag problematic disclosures*

After identifying the set actors, goals, and data, the next step is to cross-reference all of the possible disclosures and determine which disclosure may potentially violate expectations of privacy.

The process of determining which disclosures are problematic isn’t concretely defined. First, we considered if the disclosure occurred in the system. So, for instance, while it would be a privacy problem if users’ passwords were published on the Internet, P2PU didn’t support this disclosure. Second, we considered whether or not a disclosure was consistent with information disclosure norms. To determine if
it was consistent, we contrasted disclosures with existing social networks like Facebook and Twitter as well as thinking about norms of information flow in the educational context. So, for instance, when people are on Twitter, they are publishing information to be consumed by others. However, when they interact on P2PU, they may be more interested in figuring out the answer to a question. Their goals and expectations about how that data will be disclosed will differ.

Out of the set of possible disclosures, the following four types of disclosures were identified as potentially problematic:

**Users disclosing profile information to the public**
Profile information includes identifying information like email address and full name along with biographical information, a profile picture, and links to user’s websites (along with automatically polled content from any RSS feeds). While most of this information is optional, a user may disclose this information without realizing that it will be disclosed publicly, and if archived or indexed may exist permanently.

A similar disclosure that demonstrates why this may be a privacy concern is when Facebook decided to make certain aspects of a profile public. People assumed that their detailed profile information would only be available to friends, but it was now available to anyone online. Users were upset, and Facebook changed some defaults to make profiles more private by default.

**Users disclosing social connections to the public**
In addition to profile information, the user’s social connections are displayed publicly. This is not optional and a user has no notice that this information is made public. In the educational context, social connections could be seen as potentially sensitive if students have few friends, or if they are friends with those in authority. There have also been cases of using social graphs to predict sensitive information
about users (link to studies). Additionally, Google Buzz revealed users’ social connections by making followers out of people that users emailed. This violated people’s sense of privacy, even though no message contents were disclosed.

**Users disclosing posts to public**

The fact that users disclose their posts to the public is a central tenet to P2PU’s open architecture. However, users may feel uncomfortable posting in an educational context if all of the posts are publicly available, linked to their public profile, and archived permanently. Clearly there is some form of tension here, as users may not necessarily feel a privacy violation when their posts are publicized, but may not feel comfortable posting in the first place. Even if they feel they must post, they may not include any sensitive information (for example, if they’re struggling with material) that would help to bring about social cohesion necessary for the growth of community and trust.

**Users disclosing clickstream activity to research community**

P2PU has discussed the concept of releasing clickstream data for all users to the P2PU research community. This would allow the research community to perform a variety of experiments with the data, but it may also lead to problematic disclosures since users are unlikely to realize that their data will be used in this way. Privacy debacles have occurred in both the AOL query research and the Netflix recommendation research when user data was made available to a broad research community, even when that data had been anonymized. The fact that this “invisible” data is being released to third parties may also produce a chilling effect on social interactions among peers who feel they cannot use the platform for social exchanges. This may decrease intimacy and sense of solidarity that motivates students to continue the learning process in order to engage in a process of “becoming” (Lave and Wenger).
**Step Five: List potential mitigations**

The disclosures that lead to a sense of violated privacy in P2PU can be augmented with controls that either change user expectations or reduce the impact of the disclosure. In this step, each disclosure was considered against a set of potential design alternatives that would mitigate the harm of the disclosure. These potential disclosures include preventing the information from being disclosed at all (don’t disclose), removing the identifying information from the data so that it can’t be traced back to an individual user (anonymize data), allowing users to provide different identifying information that may or may not be tied to their real identity (allow pseudonyms), aggregating data so that a single user’s data is not separable (aggregate data), providing notice to the user about the disclosure before it occurs (provide notice), and allowing the user to choose whether or not the data is disclosed (offer choice).

This list of mitigations is simply a common set of design choices for preserving user privacy. Certainly, there are design choices outside of this list that may be appropriate for certain problematic disclosures. There was no hard and fast rule that made us consider only these mitigations. For instance, there could be contractual obligations between actors to use data ethically, or more exotic technical concepts like differential privacy could be used. However, we were able to find some commonly used design alternatives that effectively addressed these problematic disclosures.

Each disclosure was examined against these alternatives, and the set of potentially applicable alternatives are discussed below:

**Users disclosing profile information to public**

*Don’t disclose:* Profile information could be forever hidden from the public. This is similar to some forums and online communities, where you must be a member to view other users’ data.
Provide notice: P2PU could provide more explicit information about what data will be shared publicly.

Offer choice: Users could have some form of control over what profile information is shared publicly.

Users disclosing social connections to public

Don’t disclose: Social connections could be hidden from the public.

Aggregate data: Rather than displaying a user’s lists of friends, P2PU could only display the total numbers of connections.

Provide notice: P2PU could let users know that their social connections will be displayed to the public.

Offer choice: P2PU could offer users the option to show or hide their social connections from the public Web.

Users disclosing posts to public

Don’t disclose: All posts could be hidden from the public until they signed up for P2PU.

Anonymize data: Posts could be published anonymously, allowing users to comment without disclosing their identity.

Allow pseudonyms: Posts could be published pseudonymously, allowing users to selectively disclose their identity.

Offer choice: Some posts could be made private or published to a selected audience, rather than public only.

Users disclosing clickstream activity to research community

Don’t disclose: Clickstream data could be withheld from the research community.

Anonymize data: Anonymized clickstream data (stripped of identifying information) could be made available to the research community.
**Aggregate data:** Aggregated clickstream data could be made available to the research community.

**Provide notice:** Users could be given notice that their clickstream data is used for research purposes.

**Offer choice:** Users could be given the opportunity to choose whether or not their clickstream data is captured and used for research purposes.

**Step Six: Determine impacts of mitigations**

In this step, we list the mitigation, then the list of goals that are supported or harmed by the mitigation.

**1. Profile information**

**Only display users' profile information to other users**

*Supported:* (Users) Avoid revealing embarrassing or otherwise harmful information online

*Harmed:* (Organizers) Create an open environment for learning in order to enable participation, accessibility, innovation and accountability

*Harmed:* (Public) Benefit from courses without directly participating

If users don’t disclose their profile information to the public, then they will be less likely to reveal harmful information. But if they want to share their learning profile with others who aren’t registered P2PU users, then they may be less motivated to contribute. Also, if users are unable to disclose their profile information publicly, even if they want to, it impacts the ability of P2PU to foster openness. Learning about users may help the public understand how a user’s identity affects their comments but it is unlikely that (unless they are a researcher, covered separately below) they will be able to make anything more than superficial judgments about the content or the community from reading profile pages.
Offer choice when users disclose profile information to public

Supported: (Users) Avoid revealing embarrassing or otherwise harmful information online

If users are given the choice of whether to reveal their profile publicly, they are given more control over their information, potentially lessening the privacy harms. Users may be more willing to share more sensitive information (for example, their learning styles, what they struggle with, what they excel at), which could advance social cohesion.

Provide notice when users disclose profile information to public

Supported: (Users) Avoid revealing embarrassing or otherwise harmful information online

If users are made aware that their profile information will be made public, they will be more likely to avoid revealing embarrassing or otherwise harmful information on P2PU.

2. Social connections

Only disclose social connections to other users

Supported: (Users) Avoid revealing embarrassing or otherwise harmful information online

Harmed: (Public) Benefit from courses without directly participating

Harmed: (Organizers) Create an open environment for learning in order to enable participation, accessibility, innovation and accountability

If this information isn’t disclosed, it lessens the choice of a problematic disclosure. However, like preventing disclosure of profile information, not revealing the social graph impacts the public’s ability to access a learner’s social connections.
Aggregate data when users disclose social connections to public

**Supported:** (Users) Avoid revealing embarrassing or otherwise harmful information online

**Harmed:** (Public) Benefit from courses without directly participating

**Harmed:** (Organizers) Create an open environment for learning

Instead of displaying complete “connections” lists, P2PU could only display the number of connections. If social connection data is aggregated, it has the same (although somewhat lessened) impacts as not disclosing the data at all.

Provide notice when users disclose social connections to public

**Supported:** (Users) Avoid revealing embarrassing or otherwise harmful information online

If users are made aware that their social connection data will be made public, they won’t be surprised to find it available. The data will still be available to the public.

Offer choice when users disclose social connections to public

**Supported:** (Users) Avoid revealing embarrassing or otherwise harmful information online

**Harmed:** (P2PU) Create an open environment for learning

If users are given the choice as to whether or not to disclose social connections, their ability to control their data is supported. However, some may feel that this inhibits the community since it is useful to know who is connected to whom in the group. If a user chooses to hide their social connections, then no one will see the data, which would suggest that the public’s goal of learning from others’ material, is neither supported nor harmed.
3. Posts

**Don't disclose when users disclose posts to public**

*Supported:* (Users) Avoid revealing embarrassing or otherwise harmful information online

*Supported:* (Users) Learn about subjects from peer contributions

*Harmed:* (Public) Benefit from courses without directly participating

*Harmed:* (Organizers) Create an open environment for learning

*Harmed:* (Organizers) Bring learning material to as many people as possible

If users’ posts are only displayed to “members” of P2PU, then privacy goals are supported, and users may feel more comfortable contributing to courses, which supports their learning goals. However, the ability of the public to learn from P2PU, as well as the organizational goals of openness and accessibility, may be harmed by the lack of disclosure since not everyone can access users’ posts.

**Anonymize data when users disclose posts to public**

*Supported:* (Users) Avoid revealing embarrassing or otherwise harmful information online

*Harmed:* (Users) Organize courses that are compelling and informative

*Harmed:* (Organizers) Ensure learning material is high quality

If users are allowed to comment anonymously, then they may be more likely to provide truthful comments. However, the communal learning goals may be harmed since anonymous comments may have the effect of devaluing the comments, leading to “trolling”. The ability to maintain high quality content, important for both course facilitators and P2PU organizers, may thus be harmed if anonymous comments are enabled.

**Allow pseudonyms when users disclose posts to public**
**Supported:** (Users) Avoid revealing embarrassing or otherwise harmful information online

**Supported:** (Users) Learn about subjects from peer contributions

If users can employ pseudonyms in addition to their primary identity, their privacy will be protected. This may also help them learn by encouraging them to contribute without making them feel like their contributions will be permanently associated with a single official identity. Since pseudonyms are semi-persistent, they can still be used to accrue a reputation and have been effective in many online communities as a compromise between unique persistent identities and complete anonymity.

**Offer choice when users disclose posts to public**

**Supported:** Learn about subjects from peer contributions

**Supported:** Avoid revealing embarrassing or otherwise harmful information online

**Harmed:** Benefit from courses without directly participating

Users who are able to choose whether or not their posts are publicly available may be able to learn more effectively since they will feel more comfortable contributing. They may be able to ask questions or post information that would be embarrassing if posted publicly. However, the public’s ability to access courses may be harmed, since some interactions would take place in a back channel.

4. **Clickstream data**

**Don’t disclose when users disclose clickstream activity to research community**

**Supported:** (Users) Avoid revealing embarrassing or otherwise harmful information online

**Harmed:** (Users) Conduct research and experiments that provide insight and guidance
If users’ clickstream data is never disclosed, then they are at less risk of problematic disclosure. However, the researchers’ capabilities for providing insight on the community are harmed since clickstream data could prove useful in analyzing learning patterns that could improve learning.

**Anonymize data when users disclose clickstream activity to research community**

*Supported:* (Users) Avoid revealing embarrassing or otherwise harmful information online

*Supported:* Follow ethical guidelines

*Harmed:* (Researchers) Conduct research and experiments that provide insight and guidance

If clickstream data is anonymized, it may help to protect users’ privacy and researchers follow ethical guidelines, although it is important to note that anonymized data may, in many cases, be re-identifiable. Additionally, the anonymized data may not be as useful for researchers, since it may not be as accurate.

**Aggregate data when users disclose clickstream activity to research community**

*Supported:* (Users) Avoid revealing embarrassing or otherwise harmful information online

*Supported:* Follow ethical guidelines

*Harmed:* (Researchers) Conduct research and experiments that provide insight and guidance

Similarly, if clickstream data is aggregated then users’ privacy is protected and researchers ethical goals are preserved, but it may impact the researchers’ ability to effectively analyze the data and reach conclusions about user behavior.
Provide notice when users disclose clickstream activity to research community

Supported: Conduct research and experiments that provide insight and guidance
Supported: Follow ethical guidelines
Harmed: Avoid revealing embarrassing or otherwise harmful information online
Harmed: Create an open environment for learning in order to enable participation, accessibility, innovation and accountability

If users are given notice that their clickstream data is being used for research purposes, but are obligated to provide it, then researchers’ goals are served but users may either feel as though sensitive information is being shared inappropriately or that they are unable to effectively participate in P2PU because they feel uncomfortable disclosing the data.

Offer choice when users disclose clickstream activity to research community

Supported: Avoid revealing embarrassing or otherwise harmful information online
Supported: Create an open environment for learning in order to enable participation, accessibility, innovation and accountability
Supported: Follow ethical guidelines
Harmed: (Researchers) Conduct research and experiments that provide insight and guidance

If users are simply given the choice as to whether or not to contribute their clickstream data, then users’ privacy is protected and the “openness” of the community is upheld. Researchers goals may be somewhat harmed because of the bias introduced by self-selecting participants, but subject choice is commonly a requirement of research ethics guidelines.
**Design Alternatives**

After we analyzed how each mitigation impacts goals, we chose which ones represented the best compromise between stakeholders’ interests. The goal of the previous steps is not to arrive at a conclusion for what to build. Instead, it is meant to guide the analyst through the process of enumerating potential privacy concerns and to understand the ramifications of design choices. There are several factors that could influence the choice of design alternatives. For example, actors who are most vulnerable in the system should have their goals prioritized. This helps preserve the ultimate goal of the initiative, which is to serve its users. This is encompassed by the broader ethical goal of user empowerment. Also, not all of the alternatives are exclusive – some may be useful to consider in tandem. For instance, notice and choice are commonly used for privacy settings in applications.

**Offer choice to protect users disclosing profile information to public**

A user should be able to choose what information in their profile is disclosed to the public, since this information may be indexed by search engines, archived permanently, and otherwise available to an extremely large invisible audience.

**Provide notice to protect users disclosing social connections to public**

A user should be made aware that their social connections, including enrolled courses, are publicly archived and accessible.

**Allow pseudonyms to protect users disclosing posts to the public**

Users should be able to use pseudonyms while interacting on the P2PU system. This will allow more freedom in the learning environment and support users’ privacy goals while upholding the ultimate goals of an open system.

**Offer choice to protect users disclosing posts to the public**
Users should be able to choose whether or not their posts are accessible publicly or only available to a select group of users. This would facilitate private conversations among users as well as public conversations.

**Offer choice to protect users disclosing clickstream activity to research community**

Clickstream activity disclosure should be a configurable feature. This would allow users to choose whether or not they want their clickstream activity disclosed, and make it apparent to users when their clickstream activity is being recorded, and what data is in their clickstream activity.

**Prototypes of Design Alternatives**

After identifying the set of design alternatives, and considering them in the light of education privacy research, we designed the following set of prototypes:

**Identity Selector**

The “Identity Selector” is a feature that supports pseudonyms in interactions on P2PU. Rather than having a single identity that is used for all interactions, users can create multiple identities, or pseudonyms, and choose which identity to use for each interaction. This is a balance between a single, persistent identity and a user being anonymous. A user can explore different identities freely without worrying about a single “mask”, but each persona is tractable and can accrue reputation.
In the first screenshot, there is a new dropdown list in the upper right corner where a user can select an identity. In this case, the identity selected is “alsmola”.
By selecting the identity list, the user can choose from a list of “masks”, including the possibility of being anonymous. The masks described here are helpful for showing why someone might want different identities. Someone who is learning or teaching material that could be considered “pirating” may want to keep their behavior disassociated with their other identity, while someone who is learning a new subject (e.g. “python n00b”) may also want to keep a separate face to the world. The option of anonymity may or may not be a desirable feature. It could be excluded if it was seen to encourage bad behavior.
Once the user switches to a new identity, s/he has a new set of social connections, a new status feed, and so on. From the outside, this is no different than a new user, but the fact that it is supported by the interface encourages students to create new identities and use them to take on new challenges that they may otherwise be intimated by.

**Clickstream recording**

Rather than silently disclosing all of the clickstream data to researchers, P2PU should allow users to “opt-in” to clickstream recording and notify them when their clickstream is being recorded. Furthermore, users should be allowed to see what data is being submitted in a format that is readable and comprehensible.
In the first screen, a new privacy section is available in the profile editor. One of these options is the ability to opt-in (or opt-out, depending on the configuration) to the clickstream data program. Another option is to hide profile information from the public. This is not part of this prototype, but a simple design alternative identified in the process above that would fit here well. An additional option could be used to hide social connections.
If the user chooses to enable clickstream recording, they will see a recording option in the upper right corner of their screen. This serves two purposes: first, users will be given notice that their actions are being recorded, and second, it will give them the option to enable and disable recording.
The user can disable the clickstream recording, similar to going into browser “incognito” mode, through the interface.
The user can also view all of the data recorded about them in a clickstream data viewer, which would show them in a clear format what data would be available to researchers. This is similar to a browser history mechanism. The user could also delete data if they didn't want it to be submitted for research.

**Private discussion**

Users should have the option of participating in discussions that do not take place in public. This allows a greater sense of freedom while preserving some of the openness of the community by also allowing public interactions.
This prototype supplants the existing private message feature, which is one-to-one, with a multi-party private conversation feature. The user could select a group of students to invite to a back channel conversation.
The user will then be able to engage in conversation with a select audience, out of the view of the public and other students.

**Conclusion**

We believe the result of this analysis is a thoroughly considered set of recommended features for P2PU to support user privacy and broader goals of delivering education to a wide audience. The privacy design analysis methodology is a structured way of identifying privacy concerns and design alternatives that offers a glimpse into the effects of data disclosure that may not have been apparent at first glance. The resulting design prototypes represent a justified set of features that help protect user privacy while supporting the goals of the other actors in the system.
While we realize that not all of the features are likely to be acceptable in their current form, but hope that at the least they will inspire thinking about how privacy concerns could be addressed and that aspects of them find their way to P2PU.