Summary of Phone Conversation with [redacted]
October 21, 2010
From: [redacted]

I called [redacted] to discuss the recent abuse activity at MIT, where someone managed to download approximately 450,000 articles from JSTOR. We spent approximately 40 minutes discussing this breach and its scale. [redacted] apologized on behalf of MIT and said they have been investigating the event with all resources at their disposal. [redacted] emphasized the value of JSTOR both as an asset and as a partner not for profit organization serving the scholarly community. [redacted] said several times that this was horrendous, dismaying, etc. They have also been very concerned about this for their own sake, because the scale of the effort had an impact on their network.

[redacted] outlined the procedures that MIT takes when these kinds of things occur. [redacted] said that these generally fall into one of two categories, what [redacted] called "inside the tent or outside the tent":

1. Inside the tent: A faculty member/graduate student/undergraduate is downloading a ton of data for some kind of computer science project. They want to test out a robot, or they want to do some kind of data mining, or something, and they are just looking for a large corpus. These events are not about getting specific content; they are just about getting data.

Action: when they discover this kind of activity, they "bring the hammer down" on the person responsible, letting them know very clearly that such activity is not within the terms of the license and that they have to destroy all data collected in this way. In most cases, the person did not realize that doing such a thing is not allowed and everything gets taken care of very easily.

2. Outside the tent: Someone is trying to get the content and has hacked into MIT by overcoming a proxy, buying or otherwise stealing someone's ID, etc. In these cases one cannot know the person's intent but it is assumed to be nefarious. If this was the category for our case, [redacted] characterized this as "grand theft". They take this very seriously.

I asked [redacted] what MIT is doing to figure out which situation applies for us.

They contact their IS&T group who analyze logs and work to track the origin of the activity, what seems to have been downloaded, and generally do everything they can to figure out what happened. In our case, to put it succinctly, they don't know who did this. [redacted] said that the approach taken in our case was to exploit dynamic IP addresses. [redacted] said that in their IS&T investigation, it appears that the perpetrator opened a session, got the content, deleted cookies, closed session, etc. [redacted] said they seemed to be downloading the content in a systematic and sequential way. It didn't seem to be aimed at getting specific content (from a particular journal or discipline) but just one after another. This was why they initially thought this might be some kind of computer science project.

They are trying to see if it was someone at MIT by asking people in the departments where such activity were likely to occur. [redacted] has spoken personally with key faculty in those departments to explain what happened and to ask if there were student or other projects underway that might

have led to this kind of activity. [redacted] has not heard back on this but is following up and will keep us posted.

If it is not within the tent, they really have no idea who could have done this or how to track them down. They also therefore obviously have no way of getting the data and destroying it. I specifically confirmed with [redacted] that it is a matter of not being able to find the person, and not a matter of MIT trying to protect the identity of the person (for privacy purposes, etc.) [redacted] said: "absolutely not, it is just not possible for us to identify them." We talked about the fact that this activity seems to have been systematic and sophisticated, and [redacted] wondered if these people were likely just to move on to other colleges or universities to exploit other softer security holes. We agreed that it would make sense to keep an eye out for this content to appear on the web somewhere.

We talked about the steps MIT has taken to prevent such an event from recurring. Essentially it amounts to driving everyone through an authentication step to make them verify their identity before they can get access to JSTOR. This would make it much harder to pull off an approach like this without leaving a trail, but it is not absolutely failsafe, as people can buy or steal identities. But it would make it much more difficult.

The problem with this approach is that it penalizes everyone for the actions of the few (or even one). With this in place all JSTOR users at MIT will have to submit to a ID/password challenge. This will make us look more protective and proprietary than other resources available on the MIT campus, since this approach will not be taken for all e-resources, but for a select few. We should try to find out what other resources are handled in this way.

I told [redacted] that I would consult internally and get back to [redacted] if there were next steps we needed to take. In the meantime [redacted] and [redacted] are working together on figuring out an implementation plan to address the situation in a mutually acceptable way.