



Yale College Council
Open Data Policy Recommendations

January 2014

Introduction

In January 2013, the administration decided to shut down the websites Bluebook+ (now CourseTable) and CourseRoulette. Despite the sites being online for several semesters, the administration argued that the sites violated Yale's acceptable use policy by taking Yale's course data and repackaging it on their own servers. YCC met with administrators to discuss the shutdown, and soon after the administration agreed to reach out to site developers Harry Yu and Peter Xu.

We support the efforts of student developers and would like to applaud Harry and Peter for working to improve the shopping experience. It became evident, however, that legal regulations in place to protect private University information and trademarks were violated. Yet, given the fact that the site was in use for three semesters and other students created applications in a similar way, the manner and timing of the shutdown was inappropriate and unnecessarily disrupted shopping period for thousands of students. While the concerns raised by the administration were justified, we fault administrators for neglecting to communicate with YCC or other students to better understand the impact of their actions before the sudden shutdown.

YCC also met with Harry, Peter, and other developers to discuss what steps should be taken to ensure that University policies foster student innovation and are clearly communicated such that student-designed websites, including CourseTable, can benefit students. This report recommends changes to increase transparency and revise Yale's acceptable use policy to allow for both data security and student innovation.

Recommendations on Open Data Policy

Yale's current Acceptable Use policy as outlined in the Information Technology Appropriate Use Policy (ITAUP)¹ is potentially damaging to student innovation. YCC recommends revising the ITAUP in order to protect both Yale data integrity and student innovation.

ITAUP 1607.1 C.5.e reads "Use damaging the integrity of University IT Systems or non-Yale systems. This category includes, but is not limited to, the following... e) Unauthorized modification or removal of data or equipment."

An interpretation of the current rules could lead to a situation in which a student developed application, which repackages data without explicit approval from the administration, could be shut down for violating acceptable use policies. This leads to potential situations in which the administration can shut down any site that repackages data even if it meets the rest of the

¹ <http://policy.yale.edu/policy/1607-information-technology-appropriate-use-policy#1>

acceptable use standards.

We recommend rewording this section to clarify Yale's acceptable use policy and to account for the below recommendations in favor of open data.

1. Students should be able to repackage and use any Yale data available to them as long as the following conditions are met:
 - a. The developer does not infringe on Yale's trademarks. Yale should authorize use of certain copyrighted data such as course descriptions assuming the rest of acceptable use conditions are met. Use of the Yale name or other trademarks should be guided by existing regulations.
 - b. The developer does not cause damage to or unreasonably burden Yale's network or servers.
 - c. The data is not available to people who do not originally have access to it.
 - d. The data displayed by an application must have integrity. For example, one should not be permitted to post a site with deliberately altered course reviews or descriptions, while presenting the data as unaltered.
2. No non-malware site is blocked with IP blocking and packet inspection such as how CourseTable was blocked. The university firewall should be used only for the bona fide blocking of malware or to shield the university from outside legal liability (e.g. blocking of torrents). Any use for reasons of content (even when university trademarks or copyrights have been infringed upon) creates a dangerous precedent toward censorship as the tool is being used for purposes other than its original application.
3. A clear procedure should be created to address violations of the published data use policy. Such a procedure should treat developers with respect, while at the same time recognize the need for rapid compliance.

Recommendations to Increase Transparency

A lack of communication about data and appropriate use policy was a significant factor in the CourseTable shutdown. The administration should make clear to the developer community what acceptable use of Yale data is. YCC recommends the following to ensure developers are aware of policies before creating new applications.

1. Designate a member of the administration who can serve as a liaison to the developer community for any questions about data use. This person should be familiar with Yale's Acceptable Use policy and should be able to answer developer's questions or tell them

whom to contact if they have additional questions. This person should have easily accessible office hours, and preferably experience working with student developers.

2. Create a website that can serve as a resource for student developers. This site should clearly list the guidelines for Yale data usage, and answer frequently asked questions on developing applications with Yale data. The site should do more than link to the Acceptable Use Policy to be a useful resource. If Yale has API's then this site should list them. A successful example of this concept is Berkeley's API central (<https://developer.berkeley.edu>).
3. Publish a list of data sets on the website that may be used without restriction. Course descriptions should probably be in this category, since it is public; data sets that may be used with some restriction (e.g. the developer must place the data behind CAS authentication and ensure that the user is a Yale undergraduate, although the developer may be held disciplinarily accountable for failure to do so); and data sets that may not be used such as databases containing personally identifiable information or graded final exams for each class. These data set categories can be called Unrestricted, Restricted, and Prohibited.
 - a. To the extent technically possible, Yale should create APIs that enable access to Unrestricted Data, and enable access to Restricted data while at the same time serving as a means of security. Where an API does not exist, scraping should be permitted to enable the authorized use of data sets.
 - b. Other data sets that should be made available in either unrestricted or restricted form include Yale dining information, Course selections and evaluations, UCS calendar and employment data, room reservation interfaces, and Bass media data for reserving items.
4. Communicate this information to the student body. Potential methods of communication should include an email from the YCDO and YCC, notices in the Computer Science department newsletter, and distribution to the Student Technology Collaborative.

Contact Information

YCC Communications Director Andrew Grass manages this project. You can contact him at andrew.grass@yale.edu.