

Guidelines for Data Access and Use

May 7, 2009

1. Introduction

In recent years, it has become easier for faculty and staff to access student records, particularly over the web (e.g., via Faculty Services). In addition, the College continues to expand its interest in assessment, leading to an increasing need for analyses of relevant data. As our access to and requests for data have grown, a number of important questions have arisen. These questions include, but are not limited to:

Who should have access to what kinds of data, and for what purposes?

If someone wants access to data on students, faculty, departments, or divisions, whom should he or she ask for permission? What should the criteria be for deciding whether access should be granted or denied?

For data that one already has access to, what do we, as a community, agree are the appropriate uses for that data?

Currently, the Office of Institutional Research, the Registrar's Office, and the Admissions Office decide whether to grant or deny access to most requests for data on a case-by-case basis, often in consultation with the Provost or others. Whether regarding these data, or data that are already available to faculty and staff online, clearly access does not imply permission to use such data for any purpose. Our current practice has been to normally allow access to data and trust people to use it responsibly. We do not propose to change this arrangement; rather, we aim to develop a set of principles these offices can use for guiding such decisions about access, and to develop a consistent definition for responsible use that is agreed upon by the college community. Our goal is to develop guidelines that can consistently be communicated to, consulted by, and applied to each member of the community who has access to, who requests access to, and who grants access to data.

At the beginning of the 2008–09 academic year, the Data Access and Use Subcommittee of the Ad Hoc Assessment Planning Committee was instituted by Provost Connie Hungerford, with the goal of drafting a policy document that would address data access and use issues. The members of the group, which is a subcommittee of the Assessment Committee, include Garikai Campbell (Dean's Office), Ellen Magenheim (Economics), Robin Shores (Institutional Research), Steve Wang (Mathematics and Statistics, committee chair), Martin Warner (Registrar's Office), and Richard Wicentowski (Computer Science). The committee also includes a representative from the Admissions Office, initially Jim Bock, but which has been filled by Ryan Keaton and Sebastian Indacochea.

2. Scope

The committee is charged with creating guidelines for policies relating to data in several contexts:

- 1) Data pertaining to individual students, such as gender, race, grades, and test scores.
- 2) Data pertaining to individual faculty, such as teaching loads, course releases, and grades given.
- 3) Data pertaining to departments and other administrative units, such as course enrollments, staffing information, and budgets.

Currently, there is a culture of openness regarding student records. A faculty member can easily look up any student's transcript online, for instance, without explaining to anyone his or her purpose in doing so, and without leaving any formal record of the access. Although the online interface displays a reminder that "Access to and use of this information is limited to official school business only", faculty are assumed to use the system responsibly and appropriately.

We seek neither to endorse nor to oppose this openness. We emphasize, however, that these guidelines apply to all use of data, including data that are already accessible in the current online system. That is, the fact that access to certain data is possible does not give implicit permission to access or use that data.

Overlap with the Institutional Review Board (IRB): Some of the concerns of this committee overlap with those of the IRB, and for some requests it may be necessary to obtain IRB approval as well. Some requests for data, however, may not constitute "research" per se, and thus would not need to be cleared by the IRB. For instance, a department may want to determine whether students are more likely to major in the department depending on which of two introductory courses they take. This would ordinarily not be considered research because it is not intended to be generalizable knowledge, but instead simply a comparison of the two particular courses. A request for such data would fall under the purview of these guidelines rather than the IRB. In cases for which it is not clear whether or not IRB approval is needed, the chair of the IRB should be consulted.

3. General principles

Some general principles on access and use of college data appear in the college's Information Security Policy:

Faculty and staff granted access to institutional data may [access data] only to conduct College business. In this regard, employees must:

- Respect the confidentiality and privacy of individuals whose records they access
- Observe ethical restrictions that apply to [use of] the data to which they have access
- Abide by applicable laws or policies with respect to access, use, or disclosure of information

Employees may not:

- Disclose data to others, except as required by their job responsibilities
- Use data for their own personal benefit, nor for the gain or profit of others
- Access data to satisfy their personal curiosity

Here we present additional questions and principles that should be considered in decisions on data use and access:

a. Is there an appropriate action or policy decision that could result from analyzing the data requested?

Principle: There should be a plausible action or policy decision that could result from the analysis, even if that action is not actually taken. Data analyses should not be carried out solely to satisfy one's own curiosity or personal benefit.

Example: A Swarthmore alum is running for political office. A professor is curious about how the alum did as a student here and wonders if it is permissible to look up his grades online.

Decision: Access to these data should not be granted for this purpose.

Reason: This is private information, and no appropriate action or policy decision could result from such access.

Example: A department wants to see if incoming female first-year students who express an interest in majoring in the department are more likely to switch to another major than are male students. The department requests a list of intended majors from the admissions office for the purpose of cross-referencing it against the department's own list of majors.

Decision: Access to these data can be granted for this purpose.

Reason: Even if no action is taken, it is conceivable that the department might want to change its advising procedures, increase recruiting, etc.

b. Will any findings be used for an inappropriately exclusive purpose?

Principle: One should not access and analyze data in order to exclude students from some opportunity based on their membership in a group.

Example: A professor wants to see if non-majors from a certain department do poorly in her course. If so, she will no longer allow students from that department to take her course. She asks the Registrar for data on majors and grades in previous offerings of the course.

Decision: Access to these data should not be granted for this purpose.

Reason: Students who meet the stated prerequisites for a course should not be barred from participation based on their membership in some group (e.g., ethnicity, gender, sexual orientation, participation in athletics, etc.). However, there are purposes for which access to these data may be allowed (e.g., to better identify the appropriate prerequisites for a course).

c. Can the work be done without using individual identifiers?

Principle: When possible, analyses should be done without personally identifying information. Even without names or ID numbers, however, at a small college the combination of race, sex, and major may be enough to identify some students, and so data with this level of detail are considered the same as data with identifiers under the Family Educational Rights and Privacy Act (FERPA).

d. Does the work improve the educational experience, or otherwise contribute to the educational mission? Can the results be shared so that others can benefit from the findings, or learn from the methodology?

Principle: Reasons for requesting data and proposed analyses should be consistent with the college's educational mission. The results of such requests should contribute to the quality of teaching and learning at the college.

Example: A faculty member wants to hire a Swarthmore student to be a math tutor for his child, who is currently in middle school. He has a few students in mind, and he wants to use Faculty Services to review their Swarthmore math grades before he contacts them.

Decision: Access to these data should not be granted for this purpose.

Reason: This is private information. A faculty member who wants this information should ask the potential tutors to provide it. Since this is for the benefit of a faculty member's child, who is not a Swarthmore student, access to this information cannot be justified on the grounds of contributing to the educational mission of the college.

e. Should the analysis be done by the faculty/staff member, or would the analysis be more appropriately done by the Director of Institutional Research (IR) or the Registrar's Office?

Principle: There are situations in which it may be appropriate for IR or the Registrar to perform the analysis without disclosing the raw data, especially when dealing with sensitive student information. Normally, this decision should be made when the request for data is made. This protects sensitive data and may avoid creating a situation in which a faculty member could become biased about a student or group of students.

f. How will the analysis and data be disseminated?

Principle: Greater caution may be needed to protect anonymity if the results are to be published (e.g., in a book or journal article). This is especially important now that some journals are requiring that authors deposit datasets in a journal archive as a condition of publication. At a small college, the combination of variables such as race, gender, major, etc., may be enough to uniquely identify individuals.

g. How will the data be secured?

Principles: Confidential data should be stored securely (e.g., password-protected) during the study. Datasets should not be left unsecured, as this increases the chances that confidential information could be lost or otherwise compromised.

h. What will happen to the data after the analysis has been completed?

Principles: Confidential data should be destroyed at the conclusion of the study. Datasets should not be re-used for other purposes after the initial analysis has been completed unless a separate request has been sought for additional analysis.

4. Assessment

Some of these guidelines are motivated by law (e.g., FERPA), others by ethical principles (e.g., that people not be excluded on the basis of group membership). In addition to such guidelines, there are factors that are a matter of cultural norms specific to Swarthmore. For example, at some schools information such as median grade or enrollments for each course or department are publicly available and easily accessible, whereas that has generally not been the case at Swarthmore. Even if access to such data is not prohibited on a legal or ethical basis, our local precedent has not been to make such data widely available. We have tried to account for such cultural norms, as we understand them, in these guidelines. The members of the committee, however, may not be representative of Swarthmore as a whole, and our perception of these cultural norms may not be complete. We expect that feedback over the coming year will help to clarify the Swarthmore community's norms and principles.

We intend these guidelines to be a work in progress. We have not attempted to address all situations that may ever arise, nor can we possibly do so. We expect that these guidelines will be used over the coming year by the Office of Institutional Research, the Registrar's Office, the Admissions Office, the Dean's Office and the Provost's Office. Those offices should keep track of the requests for data they receive, what decisions are made, and what role these guidelines play in reaching those decisions (see the Data Request Form at the end of this document).

In Spring 2010, the Registrar and Director of Institutional Research should meet with representatives of the offices listed above to determine how to proceed. This meeting should be held before May 2010, when COFP makes committee appointments for the upcoming year, in case it is necessary to convene a new committee.

Appendix: Vignettes

In meetings of the Data Use Committee, we created and discussed a series of vignettes, describing hypothetical situations in which access to data was requested for possibly controversial purposes. The goal of these vignettes was to help us come to a consensus about appropriate guidelines and principles for data access and use. These vignettes are given below.

Student data

101. A student asks Professor X for a recommendation letter, which the professor agrees to write. The professor uses Banner to look up the student's transcript.

102. The math/stat department wants to explore whether student SAT scores predict success in Stat 11. They collect SAT scores, as well as ethnicity and gender, from admissions data, and then collect student grades. They fit a regression model using these variables to predict Stat 11 grades.

104. A faculty member does a study on academic dishonesty. He requests a variety of academic data include gpa. He then compares whether students with higher gpas are more or less likely to be charged with academic dishonesty. He plans to publish his results in a journal.

105. A department wants to determine whether its "best" students are going into honors so it collects data on SATs and Swarthmore grades to analyze whether there are significant differences between the students who go honors and those who do not and what those differences are. It might use the results to do more outreach to get certain students to enter the honors program.

106. A department wants to determine what follow-on rates are for various classes, requiring them to download information about specific students to see if they took course A and then if they took course B. They then need to determine the importance of students continuing to course B if they were freshman when they took course A vs if they were seniors.

107. The foreign study advisor wants to determine whether there is a correlation between student academic performance and the likelihood of studying abroad. The advisor collects data on Swarthmore grades and demographic data to analyze whether there are significant differences between the students who go abroad and those who do not.

108. The Bio dept wants to study success in intro bio and requests a variety of data including demographic data including ethnicity to make sure they are reaching all ethnic groups fairly.

109a. Professor X wants to hire a summer research student. In order to carry out this research, candidates must have taken Basketweaving 101 (which is not taught by Professor X). Student Y applies for the position and gives Professor X her transcript. Basketweaving 101 is not listed on her transcript, but the professor vaguely remembers the student saying that she is taking the course this semester. Professor uses Banner to check if the student is currently enrolled in the class.

109b. Professor X wants to hire a summer research student. In order to carry out this research, candidates must have taken Basketweaving 101 (which is not taught by Professor X).

Student Y applies for the position. Professor X uses Banner to look up whether Student Y has taken Basketweaving 101, and if so, what grade she received.

109c. Professor X wants to hire a summer research student. In order to carry out this research, candidates must have taken Basketweaving 101 (which is not taught by Professor X). Professor X uses Banner to look up the grades of students taking Basketweaving 101, and offers the position to the two students in the class who got A's.

109d. Professor X wants to hire a summer research student. In order to carry out this research, candidates must have taken Basketweaving 101 (which is not taught by Professor X). Professor X asks Professor Y, who taught Basketweaving 101, which students in the class got A's, and he offers the position to the two of these students.

110. To determine if the College as a whole has a problem with grade inflation, a faculty member wants to compare the number of A's given in, say, 2007-2008 vs, say, 1997-1998. In order to provide some context for her findings, she asks to have the SAT scores of the students in her study.

Faculty and departmental data

201. The CS department wants to know how many of its majors were satisfying the CS degree by taking cross-listed courses in Engineering. They download class rosters from classes in CS and Engineering and determine the average number of courses that majors and minors are taking in CS vs. Engineering.

202a. To determine if the College as a whole has a problem with grade inflation, a faculty member wants to compare the number of A's given in, say, 2007-2008 vs, say, 1997-1998. The faculty member looks up grades for each department in Banner.

202b. To determine if the College as a whole has a problem with grade inflation, a faculty member wants to compare the number of A's given in, say, 2007-2008 vs, say, 1997-1998. The faculty member looks up grades for individual professors in Banner.

203. A faculty member from a department with relatively lower enrollments per faculty member wants to show that in that department, faculty teach more different preparations than in departments that have higher enrollments. The faculty member goes into Banner to collect data on how many different courses are taught by faculty in other departments per year.

204. As departments are crafting policies for what a 4-course load would look like, it might be helpful to see how other departments are implementing the current 5-course load. A department chair looks up the average number of courses taught by faculty in particular departments.

205. A faculty member wants to determine whether taking course releases for maternity or paternity leave affects the probability of getting tenure or being promoted to full professor. The faculty member collects data on faculty course releases and plans to ask the provost to confirm which releases are related to maternity/paternity. Is the faculty member justified in collecting such data?

206a. A department undergoing review asks the Registrar for student lists with name, majors, minors, grad year, and honors. The department also asks for course lists including enrollment by course and professor, for its courses.

206b. The above department decides to compare its enrollment data with that of another department with whom it competes for majors. It asks the Registrar for student lists with name, majors, minors, grad year, and honors, along with course lists including enrollment by course and professor, for the competing department.

207a. A department asks Institutional Research to administer a special end of course evaluation survey, part of which will be used in an assessment of educational practices for a grant. The survey asks for the student ID. The survey data goes directly to Institutional Research (not through the department), and IR will analyze the portions of the survey relevant to the grant. IR then strips out the student ID's and returns the course evaluation responses to the department, so that they can work with it as they would any course evaluation data without knowing the students' identities. Using the Student ID, IR will link the survey data to academic (e.g. SAT scores) and demographic (e.g. race and gender) data. (IR will also link it to data provided by the department about grades and attendance.) That will be necessary for IR to analyze the data so that we can draw conclusions about the participation in and effectiveness of various features of the course by students with different backgrounds for the grant. Some interesting summary charts are prepared. Question: Can summaries showing data by various student groupings be shared with the department? (e.g. ratings of whether interest in the subject improved as a result of taking the course by gender)

207b. The department would also like to have the academic and demographic data for their internal analyses, with similar (good) concerns - are they effectively reaching all students? Question: May the department have the administrative data (the academic and demographic stuff that IR linked to) along with the survey data at the individual student level?

207c. If we decide in 207b that it is not appropriate for the department to receive that information at the student level, they could simply ask these questions directly on the course survey. Question: If we didn't allow access to administrative data, does that mean we think it's inappropriate for them to have it at all, or is it alright as long as the student voluntarily provided it?

208a. A faculty member is acting as a consultant on an assessment of the Lang Center. He wants access to student data reflecting courses taken which have a community based learning component, as well as the students' majors, minors, and honors participation. He asks IR for data on demographics, first generation, SAT's etc.

208b. A faculty member is conducting research on outcomes relating to service learning. He wants access to student data reflecting courses taken which have a community based learning component, as well as the students' majors, minors, and honors participation. He asks IR for data on demographics, first generation, SAT's etc.

209. A department is preparing for an external review, and develops a survey in which it asks students to provide satisfaction ratings not only of the department, but also of another department in which its majors take many of their courses.

210. A student wants to know his chances of getting into a popular class. His advisor uses Banner to look up enrollments in this class to see if it has been full the past few years. Based on this information, the advisor counsels the student to select a backup class.

Data Request Form

Please provide the following information on a separate sheet.

Date:

Name:

Department or Office:

1. What data are requested?
2. What will the data be used for?
3. Who will have access to the data?
4. Who will have access to the results of the analysis?
5. Will the results be publicly disseminated (e.g., in a presentation, book, article, etc.)?
6. When will the data be destroyed?
7. Does this request require IRB approval? If so, have you received approval already?

FOR OFFICE USE ONLY:

Decision:

Reason: