

D R A F T

ELIGIBILITY AND ADMISSIONS STUDY GROUP

Options for Increasing Public Access to Admissions Data

February 6, 2004 PM Version

Office of the President staff have been asked to explore options on creating a relational database containing admissions data that would be available to the public over the web. This exploration included an estimate of the cost of developing and maintaining such a system. During these discussions, some lower-cost alternatives that might serve the policy goal were also explored and are presented in this document.

This document is intended to inform the Study Group of various web-based options for providing UC admissions data and some of the practical, legal, policy and financial issues that each option raises. It does not recommend any solution, but does present the Study Group with preliminary questions, the answers to which might help in evaluating the merits of each option.

Purpose of providing data

A database, or other data reporting, is a solution to a specific problem or needs of a specific audience, and its design should be responsive to those needs. For example, the proposed database may be intended to:

- Provide information that will assist applicants, their families and counselors know what their relative chances of admission to a campus are given certain academic and non-academic characteristics.
- Allow the public (observers, friends, critics, lawmakers, researchers and media) to monitor and analyze UC's admissions decisions.

Or, the Study Group may identify other audiences and needs. This is an essential first step to determining the best method and design for making data available.

General considerations in providing data

Whatever solution is developed, certain issues must be considered, including:

- Availability of data: are the data reported to the UC corporate systems or are they maintained only at campuses? For example, corporate systems do not track course data or certain data elements from applications. Some data elements such as API rank are available only for California public school applicants, but not private school applicants or out-of-state applicants.

DRAFT

- **Timeliness of data:** Data reported to corporate systems (and possibly at campuses as well) are available several months after the events recorded. For example, final admissions data are available in January following the spring in which the decisions were made. However, students file their applications in November, so admissions data to which they would have access as they are completing their applications would be two years old.
- **Confidentiality and protection of privacy:** Student privacy must be protected so that no student can be identified directly (e.g., by name, address or Social Security) or by derivation (e.g., by using sufficient filtering to isolate individual data). An attachment describes some of the common ways of protecting privacy.
- **Appropriateness/relevance of data elements:** there should be a clear purpose for data elements included in a database or report, otherwise it should be excluded.
- **Sophistication of user:** Does the system require (or allow for) special skills, such as SAS or SPSS programming?
- **Development and maintenance costs:** Can the system be produced and maintained by University staff? Or, is a vendor required for development, documentation, explanation of data, upgrades, troubleshooting and maintenance? Will new equipment (e.g., servers) be required?
- **Unknown demand:** The number of users that will avail themselves of these new data cannot be determined and so a significant investment in new systems may not reach a broad audience.

Types of data solutions

Depending on the audience and needs being met, there are several levels of possible response. They are listed below generally in order of increasing flexibility and access to data, with side by side comparison in an attachment. All are web-based.

Option 1: Enhanced Static reports

This option expands access and visibility to existing UC reports related to admissions. Additional reports could contain profiles of admitted students by a number of factors. An example of new reports are those provided to the Study Group. Additional reports could be generated and placed on the web to be accessed by the public.

Option 1A: “Information Digest”

Reports could be accessed by browsing through a list of titles.

Option 1B: “Information Digest” with User-friendly Search Capacity

Reports could be accessed by creating a searchable database of titles and keywords. Users would type in key words and be directed to relevant tables. This would enhance access and assist users in quickly finding the information being sought.

D R A F T

Option 2: Controlled access to individual-level admissions data

This option provides for individual applicant data to be placed in a relational database. Members of the public would be provided some type of controlled access to these data that provided sufficient privacy guarantees. (Privacy issues are discussed separately on the following page.)

Option 2A. Web-based Data Analysis System

This option would allow for ad hoc tables to be created “on the fly” based on a user’s data request. Depending on the level of sophistication desired, tables could contain sophisticated statistical measures or simple percents and averages. Data tables would be restricted so that individual information would not be displayed. The basic difference between this system and Option 1 above is that users would be able to create tables beyond those provided in “canned” reports.

Option 2B. “Upon Demand” Data Analysis Service

This option is a variation on Option 2A above, however the “ad hoc” tables would be created by analysts that would create the tables and provide them on a quick turnaround basis to the user. One of the benefits of this system over the web-based system described in Option 2A is that it would be easier to manage privacy concerns. Another benefit is that a user could provide for more complicated variable definitions that combine data elements and would provide a richer level of analysis.

Option 3. Uncontrolled access to individual-level admissions data

Options discussed above would not allow the user to have access to individual student records. This option would provide direct public access to a data warehouse with raw data at the individual student level, with identifying data deleted (e.g., name, social security number, address). The user would use his/her own software (e.g., Brio) to access and analyze the data through a remote web or dial-up connection.

D R A F T

Methods for protecting privacy

FERPA, recently confirmed by the US Department of Education, prohibits release of information where a student's identity can be easily traceable. California State law uses a similar standard. Thus, any public provision of admissions data needs to consider how individual privacy would be protected.

However, there is a second reason why the University needs to consider privacy issues. We need to assure our applicants and the general public that the University zealously guards student privacy. Meeting this public perception standard may require privacy safeguards that go beyond the “easily traceable” standard in Federal and State law.

The methods listed here may prevent users from identifying individual students from data analysis. They may be used separately or in some cases, in combination. Depending on the level of data access provided, differing methods may be required.

1. Aggregation. Users see only aggregated data and have no access to individual records.
2. Sampling. With large data sets, a data system may query only a sample of data when reporting results. It is not clear that the admissions data bases are sufficiently large to successfully employ this method.
3. Limiting values. This can occur in a number of manners depending on the particular situation. Options include: a) collapsing categories (eg, data would be provided for science majors, but not physics majors); b) data presented in ranges rather than point estimates (eg, SAT scores would be shown in ranges rather than the actual score); and c) restricting levels of queries (eg, results restricted to 3 or 4 levels of analysis such as “Percent admitted by i) campus, ii) SAT, iii) GPA and iv) Major”).
4. Report summary statistics such as percentages, but not counts of students. This helps prevent multiple requests of slightly varying specifications being used in combination to identify individual students.
5. Report “Low-N” instead of a number when the number of students identified falls below a certain threshold. This is one of the methods used by the National Center for Education Statistics to protect individual privacy in their public – access data.
6. Perturb the data: a technique used to change slightly certain attributes of individuals so that in the aggregate they are statistically equivalent, but at the individual level the data are slightly incorrect. (Oversimplified example: Adding five points to one person's SAT score and subtracting five points from another's results in the same average SAT score for the group). This is one of the methods used by the National Center for Education Statistics to protect individual privacy in their public – access data.

D R A F T

OPTIONS FOR INCREASING PUBLIC ACCESS TO UC ADMISSIONS DATA

Option	Example	Privacy considerations	Resources (See Note Below)	Audience and other issues.
Option 1A. “Information Digest”	“Introducing the University” (campus profiles); Info Digest. Digest of Ed Stats (http://nces.ed.gov/programs/digest/d02/)	Privacy inherently protected by design. Methods used include aggregating data and limiting values.	UC: Depends on how many more reports would be created – analyst time to prepare them at OP and campuses. If new campus data required, would involve significant effort to collect consistent data for display. Outside: Some contractor resources required to build new site. Estimate: \$150k first year, \$100k each additional year.	Current audience is researchers and policy analysts. Better access and visibility could target students, counselors and families. Would be easily available but not user-specific
Option 1B. “Information Digest” with User-Friendly Search Capacity	NCES “Quick Tables” (http://nces.ed.gov/quicktables/index.asp)	Same as Option 1A.	UC: Same as Option 1A. Outside: Some contractor resources required to build new site. Estimate: \$250k first year, \$150k each additional year.	A user-friendly search capacity combined with greater visibility would greatly enhance usage by students, counselors and families. Would be easily available but not user-specific
Option 2A: Web-based Data Analysis System	CPEC Online customized report generator: http://www.cpec.ca.gov/OnLineData/FindRpt.asp (moderate sophistication) NCES Data Analysis System (high sophistication)	Sensitive. Would require significant protection. Methods would need to be considered in combination, including aggregation, limiting values, no small-N displays. Perturbing data would provide added protection that might allow for more fine-grained analyses.	UC: Significant analyst time required to provide data upon which system would rely. Significant documentation and data quality improvements required to adapt Corporate data for this purpose. If new campus data required, would involve significant effort to collect consistent data for display. Outside: Contractor required. Estimate: \$800k first year, \$300k each additional year.	If statistical sophistication were kept low or moderate, target audience could be students, counselors, families. Higher sophistication would target researchers and policy analysts. Would be easily available and somewhat user specific; UC determines list of reports that are available.
Option 2B. “Upon Demand” Data Analysis Service.	National Education Data Resource Center.	Sensitive. Would require significant protections such as those described in Option 2A above. Care must be taken to avoid inadvertent release.	UC: Same as Option 2A. Contractor required. Price would depend on estimate of demand. Estimate: could be done for “under \$1M.”	Likely to be used by researchers only, not by the general public. Would not provide for quick access to general questions.
Option 3. Uncontrolled Access to individual-level admissions data.	None found	Extremely sensitive. Not readily apparent that sufficient safeguards could be designed. If this option is to be pursued, further research would be required to answer privacy concerns.	UC: Same as Option 2A. Outside: Not clear how access would be provided. Contractor may be required. Estimate: Could be done for “under \$500k”	Sophisticated researchers and policy analysts only. Would not serve the needs of the general public. This type of access is normally restricted to selected employees.

NOTE: Estimates of contractor costs are “rough” and could vary significantly . Without further specification of details for each option, more accurate estimates cannot be provided. Actual costs would be determined as part of a competitive bidding process, so the estimates provided here should not be used for budgeting purposes. Rather, they are provided to give a rough indication of staff estimates of the level of contractor effort required.