IDMG Assessment Report

Looking Ahead — Creating a Better Data Environment

Context: Input and Yield

Figure 7: Input-Yield Matrix

	Low Yield	**High Yield**
Low Input		*Best scenario*
High Input	*Worst scenario*	

The above matrix (*Figure 7*) can be a useful tool in evaluating on the front end of a project the utility of proceeding with the endeavor. Every effort should be undertaken to assess the likely costs and benefits of data-related projects prior to their commencement due to their frequently intensive nature and not insignificant rate of failing to produce useful results.

Observation: Importance of Input/Output Analysis of Proposed Data Projects. *The best scenario or ideal project is, of course, one that requires low input but results in high yield. Projects that fall into this category are, however, not particularly common, and much data work that exceeds core business transactional functions (e.g., hiring someone, departmental/unit budgets, course enrollments, etc.) requires high input to derive meaningful results. Obviously, high input/low yield activities represent the worst scenario and should be avoided. Unfortunately, these are not as uncommon as we would like, and it can frequently be difficult to assess whether the end product will be highly beneficial. So, too, it can be difficult to assess how resource intensive a particular project will be.*

Recommendation: Evaluate Likely Benefits and Costs of Proposed Data Projects. *Plotting out the likely level-of-input (and then perhaps doubling it as an upper bound), and then imagining the eventual benefits of successfully completing a project should be in the forefront of participants' minds prior to engaging a particular project.*

High input/low yield endeavors on the Berkeley campus all too commonly result from labor-intensive data-related activities that are driven by day-to-day core business needs or compliance-driven activities and that do not result in the generation of higher-level information used in decision making. Many of these fall into the area of high input because of weaknesses in our current data environment: data stored in multiple systems, data stored in silos, unclear documentation about what data mean, and lack of tools to access necessary data quickly. If correctly identified, many of these projects can be moved from high-input to low-input activities, assuming the necessary infrastructure is put in place to facilitate the endeavor. A number of survey respondents expressed frustration with the current data landscape and its associated inefficiency:

> *Having to go multiple places for information is very annoying and time consuming.*
>
> *[O]ur HR systems are often cumbersome to use, not integrated, and therefore waste our time.*
>
> *Training on [campus] systems is also very limited and you have waste a lot of time playing with the[m].*
>
> *[M]any related systems on Campus do not interface with each other to provide a complete picture of situations, forcing users to run reports out of multiple systems and trying to piece together a picture to form a basis for decisions and problem solving.*
>
> *[F]inding data[on BFS] is usually a long, tedious process. … you, the survey preparers, might not care about this, but BFS is VERY conducive to repetitive motion injuries.*
>
> *We do what we can, but we are so limited in staff capacity that if data is there, we often have a hard time accessing it and using it efficiently. Ergonomically, I have to say that there is a huge cost in all the computer time required. From my own experience and from what I've heard from others in our unit, repetitive stress injuries and resulting worker's comp cases from too much computer work is a problem.*

Context: The High Dispersal of Campus Data Needs and Functional Roles

In recent years, corporate models of data warehousing and decision-support nomenclature have increasingly permeated research university settings. Arguably, the corporate metaphor of well-integrated data systems is useful in providing a yardstick upon which Berkeley can measure itself, but the fulfillment of the mission of Berkeley leads to a diversity and complexity of decision-making areas and accordant data needs that are likely unparalleled in for-profit corporate settings. The following survey responses give a sense of the range of different types of analytical needs that we undertake on the Berkeley campus:

[O]ur questions vary greatly. They may be standard questions on the number and demographic characteristics of students and they may be more complex, such as cohort analyses examining student course taking patterns or financial aid support.

Admissions - tiebreaking in augmented review, assessment of applicants with poor math achievement, review of petitions. Approval of student petitions on multiple matters. Student dismissals and re-instatements. Coordination of external accreditation process. Ongoing course and curriculum modifications.

Decisions involving Continuing Educators [in the Extension] are in the areas of recruitment (we're competing with the private sector), how to support professional development, how to create a climate that leads to job satisfaction and retention.

Overall, data is used for campus planning, analysis, assessment/evaluation, and reports. Analyze data to support a new or proposed change in policy. Establish enrollment targets. Analyze attainment of campus/unit goals. Identify potential areas of concern. Identity campus accomplishments. Respond to external requests (OP, government, and public)--policy recommendations, analysis for resource allocation, or providing basic data for others to analyze.

[L]earning outcomes in the classroom related to instructional technology, webcast & podcast decisions (student demand & satisfaction with program).

Coordinating academic policy across colleges. Assessing effectiveness of American Cultures curriculum and other academic initiatives. Implementing Undergraduate Student Learning Initiative.

Long range compensation vs. market gap analysis, Campus financial information regarding IT investment, Staffing levels, Student and faculty use of technology, cycle times.

We use information about undergraduate satisfaction to help determine the adequacy of [library] collections.

As Dept Chair, I am ultimately responsible for recruitment and retention of faculty, promotions and merit reviews, teaching assignments, mentoring etc.

Observation: High Dispersal of Data Needs Suggests Need for Nuanced, Modular Solutions: *The high dispersal of decision-making activities on the Berkeley campus suggests that a one-size-fits-all solution to our data problems is unlikely.*

Recommendation: Encourage Appropriate Integration of Data Resources and Tools to Meet Local and Campus Needs. *Because our data environment has evolved organically over time—*

largely in a decentralized manner—improvements to it should be made through a concerted iterative effort that preserves some degree of decentralization.

To better understand the dispersal of decision-making activities on the campus, the role of individuals in this process is important to consider, both with regard to how different groups rated particular aspects of the Berkeley climate, and to determine the best types of solutions to existing problems, taking into account the structure of decision-making process.

Figure 8 shows a clear disjuncture between decision-making activities and access to, and average weekly use rates of, campus-level data sets (i.e., data sets residing outside the respondents' unit only[1]). Not surprisingly, the groups at the top of the job structure—members of the Chancellor's Cabinet, deans, chairs, and others with similar ranks—are considerably more likely than other groups to span a large number of major and sub-topical decision-making areas. In contrast, support staff tend to span few. Clearly, a pyramid structure of decision making and support is in place, with certain key individuals charged with taking the forest view, and many others charged with supporting a portion of the decision makers' portfolio—working at the tree level. With regard to data access, the decision makers on the campus are particularly unlikely to have access to a number of the campuswide datasets; and, not surprisingly, they are particularly unlikely to use many campuswide data sets (residing outside their unit) on a weekly basis. Support staff are, however, more likely to have access to a number of data sets and to make use of them on a weekly basis.

Figure 8: Average Number of Major and Minor Areas of Decision Making & Average Number of Data Sets with Access to and Weekly Use of Non-Unit Data Sets by Job Type (based on trumping schema)

	Mean # of Major Areas of Decisions	Mean # of Minor Areas of Decisions	Mean # of Data Sets w. Access	Mean (per week) Times Using Non-Unit Datasets	Total N
Member of the Chancellor's Cabinet	5.6	18.9	2.2	1.3	18
Campus-level decision maker	5.9	18.4	3.8	2.3	17
College/school-level leader (e.g., Dean)	8.3	29.3	2.0	0.5	10
College/school-lev. admin. (e.g. Ass. Dean)	3.6	12.3	2.0	0.4	9
Academic department leader (e.g.,Chair)	7.8	31.8	1.0	0.0	4

[1] Note: The question on frequency of a particular campus-level dataset use was asked only of individuals who did not reside in the unit that controlled the dataset. This likely helps to explain why institutional research analysts appear to have low use of datasets. In all probability, they are spending substantial time using data that reside in their own unit.

Acad. dep. Administ. (e.g., Ass./Vice Chair)	5.3	8.7	0.0	0.0	3
Other acad. Depart. Direct. (e.g., ORU Dir.)	4.1	11.1	5.7	9.9	9
Non-academic department director	3.1	9.9	3.5	4.4	35
Manager of institutional research unit/office	5.9	21.3	9.1	7.3	11
Systems manager	3.0	5.8	3.2	5.0	35
Institutional researcher/analyst	2.8	8.8	5.5	1.8	13
Systems programmer	3.5	6.6	3.3	6.1	8
Staff member who supports campus-level decision maker	2.9	7.9	3.9	4.9	53
Staff member who supports college/school-level decision maker	4.9	16.0	4.4	5.7	30
Staff who supports academic departmental decision maker	5.3	15.8	3.6	6.0	44
Staff who supp. oth. acad. dep. dec. maker	1.9	6.2	4.0	4.1	22
Staff who supp. non-acad. dep. decis. mak.	2.1	4.7	3.5	5.1	41
Policy analyst	2.8	8.0	4.8	9.8	4
General analyst	1.4	2.6	3.8	8.4	9
Other, please specify:	1.7	6.1	4.3	5.4	19

Note: Green Shading indicates Top 5 highest average rate among job groups; Red Shading indicates Top 5 lowest rate.

Source: IDMG Survey, 2008.

Floating in the middle are two important clusters of job groups that are directly involved with the building of the campus data environment and its decision-support capacity: institutional managers/researchers and systems managers/programmers. Managers of institutional researchers are in a unique position in that they both bridge a wide array of decision-making areas, major and minor, and have access to many campuswide data systems. They are the only group that really straddles the two major regimes of decision making and direct data support for decision making. Systems managers and programmers span few areas of decision making, and do not rate unusually high in their average access to campus data sets, though their use rates of data are relatively high.

Observation: Meeting Diverse Needs. *Success in improving Berkeley's data environment will depend on meeting the needs and priorities of all involved functional groups, spanning the complete range of decision making and support functions.*

Recommendation: Maintain ongoing Communication among Different Functional Groups. *In order to bring about successful improvements in Berkeley's data environment, the campus should develop a mechanism that ensures adequate and efficient communication and consultation among all involved functional groups.*

The average number of decision-making areas and relative access to and use of datasets simply provide a general sense of the nature of the workload of a particular respondent. These indicators do not convey, however, the type of decision making or decision support that particular individuals or job groupings are most likely to undertake. *Figure 9* shows the broad patterns in this regard by position type on the Berkeley campus. Specifically, three separate job groups—1) campus-level decision makers, 2) college/school level leaders (deans), and 3) academic department leaders (chairs)—demonstrate the greatest consistent breadth of decision making. Of the major areas of decision making listed in the survey, these three groups fall within the top-five highest rates of decision making/support in 10 of the 12 topical areas (indicated by green shading). Members of the Chancellor's Cabinet, institutional research managers, and staff members who support departmental decision makers (frequently MSOs or CAOs) are the only other groups with high rates of decision making over a broad span of major areas, with all three groups falling in the top-five highest rates in six out of 12 major areas. Other types of niche administrators—assistant deans, associate chairs, Organized Research Unit directors, and so forth—demonstrate breadth of decision-making activities, but not nearly at the level of the above job groups. The remainder of job groups—essentially decision supporters of various types (IR analysts, systems managers and programmers, policy and general analysts, and staff who directly support various types of decision makers)—are much more likely to fall into the bottom-five lowest rates (indicated by red shading) of decision making/support for a particular area than they are likely to fall into the top-five highest rates.

Figure 9: Percent Making or Supporting Decisions in Major Areas by Job Type (check all that apply).

	Under-grads	Grad. Stud.	Faculty	Acad. Staff	Non-Ac. Staff	Fin	Rrsch. Grants	Infra-struct.	Courses	Alumni	Other Popul.	Other Areas	Total N
Member of the Chancellor's Cab.	61%	67%	61%	56%	56%	67%	39%	39%	17%	44%	28%	22%	18
Campus-level decision maker	67%	67%	61%	55%	58%	73%	39%	45%	21%	52%	33%	24%	33
Coll./sch.-level leader (e.g. Dean)	55%	82%	100%	100%	82%	82%	73%	82%	55%	64%	64%	9%	11
Coll./sch.-level adm. (Ass. Dean)	78%	44%	44%	56%	44%	33%	11%	11%	11%	11%	11%	0%	9
Acad. depart. leader (e.g., Chair)	71%	71%	86%	86%	57%	86%	57%	57%	86%	43%	71%	14%	7
Acad. dep. Adm. (Assoc. Chair)	75%	50%	50%	50%	50%	25%	75%	25%	0%	0%	25%	0%	4
Oth. Acad. dep. Direct. (ORU Dir.)	64%	64%	45%	27%	27%	64%	36%	45%	27%	27%	36%	0%	11
Non-acad. department director	38%	33%	33%	28%	51%	38%	21%	28%	10%	26%	21%	18%	39
Manager of institutional research unit/office	27%	64%	55%	55%	73%	82%	64%	55%	18%	27%	45%	27%	11
Systems manager	33%	31%	29%	31%	57%	26%	17%	26%	7%	14%	19%	12%	42
Institutional researcher/analyst	57%	57%	43%	43%	43%	48%	22%	13%	17%	17%	13%	13%	23
Systems programmer	40%	48%	40%	40%	48%	32%	28%	16%	12%	20%	24%	4%	25
Staff member who supp. campus-level decision maker	29%	28%	27%	29%	50%	57%	20%	33%	7%	11%	15%	17%	82
Staff member who supp. coll./sch.-level decision maker	31%	38%	43%	52%	60%	60%	31%	40%	14%	29%	29%	12%	58
Staff who supports academic departmental decision maker	36%	54%	57%	61%	61%	58%	34%	48%	31%	16%	40%	4%	67
Staff member who supp. oth. acad. depart. decision maker	27%	41%	49%	53%	45%	51%	22%	29%	20%	14%	31%	4%	51
Staff member who supports non-acad. depart. decision maker	35%	30%	30%	28%	43%	47%	12%	31%	8%	11%	20%	4%	97
Policy analyst	41%	34%	28%	31%	45%	45%	17%	17%	10%	14%	17%	17%	29
General analyst	30%	29%	35%	30%	52%	49%	17%	25%	11%	10%	22%	3%	63
Data recorder	35%	40%	40%	30%	45%	60%	30%	30%	20%	15%	35%	5%	20

As front-line administrators in colleges, schools, and academic departments, deans and chairs are charged with an almost overwhelming range of duties that span almost all of the major areas of decision making on the campus. The few relatively lower rates of decision making/support in their extensive portfolios are readily explained, however, and in no way diminish the preponderance of their decision-making activities. For example, the relatively lower rate of focus by deans on undergraduate issues is due to the fact that, in many colleges/schools, associate deans are charged with undergraduate issues (see *Figure A-21* in the *Appendix* on undergraduate sub-topical areas of decision making). In fact, members of the Chancellor's Cabinet, campus-level decision makers, other academic department directors (e.g., ORU directors), and institutional research analysts are all more likely than deans and other groups to be involved in a wide-range of sub-topical areas of decision making/support related to undergraduate issues. Deans also appear to focus minimally on enrollment and course content (see *Figures A-21—A-28* in the *Appendix*).

Although these findings point to a meaningful bounding on some deans' functional areas, the other major area of decision making/support where deans fell outside of the top five does not: *other areas of decision making*. This finding simply suggests that the survey design was comprehensive enough to capture the foci of their positions. Among chairs, this residual area also accounted for one of the two areas where they reported a relatively lower rate. The other that fell outside the top-five rate, non-academic staff, was in fact quite high. At a 57% rate of decision making/support in this area, Chairs were the sixth most likely group to be involved in this major area; and the artificial decision to shade the

top-five highest and bottom-five lowest rates in each major areas accounts for this fleeting sense of bounding.

Figure 10: Access to datasets by job group

	BAIRS	Berkeley Information System (BIS)	Cal Profiles	Planning and Analysis Databases	Admissions Database	BearFacts	Departmental Student Award System	Financial Aid (SAMS)	Graduate Student Information Systems
Member of the Chancellor's Cabinet	39%	6%	39%	11%	6%	6%	0%	6%	11%
Campus-level decision maker	45%	12%	39%	6%	3%	12%	3%	3%	6%
College/school-level leader (e.g., Dean)	27%	9%	18%	0%	0%	18%	9%	0%	9%
College/school-level administrator (e.g., Associate Dean)	0%	0%	33%	0%	0%	33%	0%	0%	11%
Academic department leader (e.g.,Chair)	29%	14%	14%	0%	0%	0%	0%	0%	0%
Academic dept administrator (e.g., Associate/Vice Chair)	0%	0%	25%	0%	0%	25%	0%	0%	0%
Other academic dept directors (e.g., ORU Director)	55%	18%	18%	0%	9%	27%	27%	9%	9%
Non-academic dept director	46%	15%	26%	3%	5%	8%	3%	0%	8%
Mgr of institutional rsch unit/office	82%	45%	55%	18%	9%	18%	27%	27%	18%
Systems manager	57%	19%	19%	0%	14%	5%	0%	7%	2%
Institutional rschr/analyst	57%	35%	70%	22%	13%	9%	9%	13%	13%
Systems programmer	52%	16%	32%	12%	16%	16%	0%	8%	4%
Staff who supports campus-level decision maker	63%	27%	37%	5%	5%	10%	6%	5%	5%
Staff who supports college/school-level decision maker	66%	21%	43%	2%	7%	21%	14%	3%	9%
Staff who supports academic dept decision maker	43%	28%	27%	1%	10%	19%	18%	1%	9%
Staff who supports other academic dept decision maker	41%	20%	27%	2%	10%	24%	14%	0%	12%
Staff who supports non-academic dept decision maker	49%	20%	27%	1%	9%	19%	8%	4%	4%
Policy analyst	45%	38%	55%	10%	10%	21%	10%	10%	10%
General analyst	54%	19%	30%	0%	5%	22%	6%	2%	8%
Data recorder	65%	35%	25%	0%	5%	15%	15%	5%	5%

Although deans and chairs and their designates, associate deans and associate chairs, are collectively involved in the greatest number of major and minor decision-making areas, they show the lowest rates of both data access and data use (see the preceding table, *Average Number of Major and Minor Areas of Decision Making & Average Number of Datasets with Access to and Weekly Use of Non-Unit Datasets by Job Type*). *Figure 10* demonstrates that even across the most commonly used campuswide data sets, deans and chairs and their designates are particularly unlikely to have access to these systems (though they would very likely be granted access upon request). **Thus, deans and chairs represent an extreme among campus data consumers: they demonstrate the greatest array of decision-making activities but the least direct access to campuswide systems.**

> ***Observation: Importance of Deans and Chairs.*** *The unique role of deans and chairs and various groups of niche administrators (e.g., assistant deans, associate chairs, et al), their support staff, and their data needs is important to consider in future improvements to the campus data landscape.*

Recommendation: Consistently Consult with Deans and Chairs*. Although it may be challenging to solicit deans' and chairs' feedback because of the unrestricted nature of their positions and accordant work commitments, undertaking the effort to do so appears essential given the specific nature of their needs and their critical functional role at the University.*

Key Task: Mapping Common Areas of Data System Access/Use

Given the fact that a large number of campuswide data systems already exist (41 separate systems were listed on the survey; see http://gradresearch.berkeley.edu/data/Major IT Systems.htm for a full list of campuswide systems), an examination of patterns of access to data systems can help to identify possible areas to examine for future data integration. For example, if the same individuals are routinely accessing the same data systems on a regular basis, an initiative in support of consolidating these various systems might be beneficial (since routinely pulling data from multiple systems is probably not as efficient as pulling data from a single or smaller number of interfaces).

Based on principal component analysis (a type of factor analysis), *Figures A-29--A-30* in the *Appendix* show patterns of common access to major campus data systems (as seen in each of the factors). The first factor includes many of the campuswide systems that are used in performing the core business practices of the University (hiring and paying employees, transferring funds, reimbursing travel, reserving equipment, and so forth). The second factor is focused on OSR data systems (including their survey system) and the Student Data Warehouse. Many of the institutional research analysts and policy analysts have access to OSR data systems and the Student Data Warehouse so it is therefore not surprising that they emerged as a separate factor. The third factor is associated with issues relating to undergraduates and graduate students. In all likelihood, the CARS system is included in this factor because most student charges are billed through this system. The fourth factor is a spatial data one, and the fifth factor includes additional data systems related to campus infrastructure and facilities.

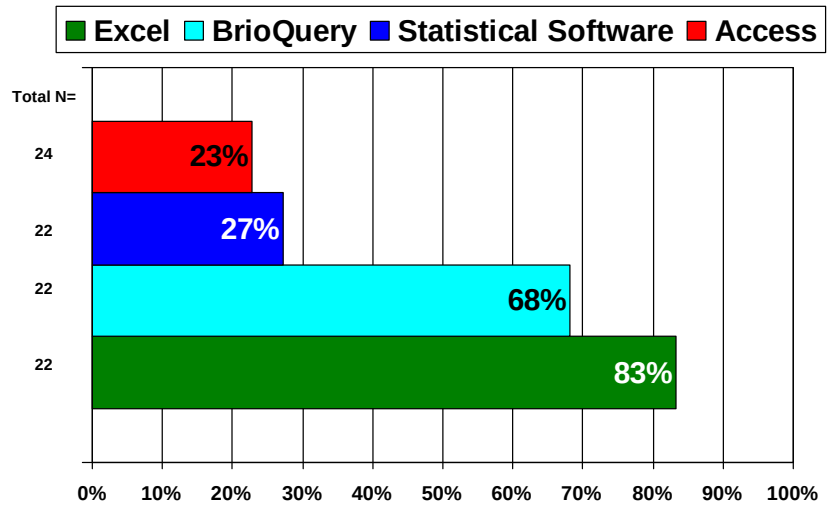
The remaining factors, 6 through 13, are either stand-alone factors or factors with only two highly correlated systems. These include the course-related data (factor 6); Office of Planning and Analysis systems, including Cal Profiles (factor 7); academic staff related data (factor 8); grant and contract related financing (factor 9); undergraduate admissions and financial aid offers (factor 10); development-related systems (factor 11); UNEX student systems (factor 12); and the library systems (factor 13). Based on this analysis, factors 1–3 seem to be the most likely areas of potential synergy with regard to future larger-scale integration efforts (the smaller factors could be addressed on an ad

hoc, more modular level). Of course, there may be areas where no current factors exist simply because our campus systems are not up to the task of supporting some of the more human-centric or longer-term planning issues that currently appear to be under-supported.

Observation: Importance of Existing Data Use Patterns: *The general patterns of use of existing campus systems lead to the following conclusions. Core business functions, student-related systems, and human resource systems hold the most promise for improvement through larger-scale integration. Other systems, such as those that deal with infrastructure/facilities, course-related data, or development, hold the least promise for this type of integration.*

Recommendation: Consider Data Use Patterns in Future Efforts. *Based on use patterns, larger-scale integration should be explored for core business functions, student-related systems, and human resource systems. Smaller, more modular approaches should be explored for systems dealing with such issues as infrastructure/facilities, course-related data, and development. Existing data use patterns should be one factor in the consideration of future design efforts, not the sole determinant.*

Figure 11: Percent of Institutional Research Analysts Using the Following Types of Applications a Great Deal or Much*



*vs. Somewhat, Little, or Never.

Another important consideration in any future initiatives to integrate data is the types of applications that data supporters use in their effort to provide decision makers with useful data-derived information. *Figure 11* shows the percentage of institutional research analysts who use the four most commonly used applications or types of applications on the campus, Microsoft Access, Microsoft Excel, BrioQuery, and various statistical packages (including SPSS, SAS, STATA, and so forth). Clearly, even among the presumably most data savvy on the campus, Excel is most commonly used, followed by BrioQuery (generally used to access data in the larger more integrated databases, e.g., HRMS, BIBS, etc.). The low rate of Access use is probably just due to the increasing integration of relational databases and growth of data warehouses on the campus that use Oracle and other similar large-scale data storage products. In contrast, the low use of statistical packages that can be used to restructure data, merge multiple datasets, and conduct sophisticated analysis is notable, however.

> ***Observation: High Use of Excel.*** *Excel is used much more frequently than statistical packages by institutional research analysts on the Berkeley campus, our resident data experts.*

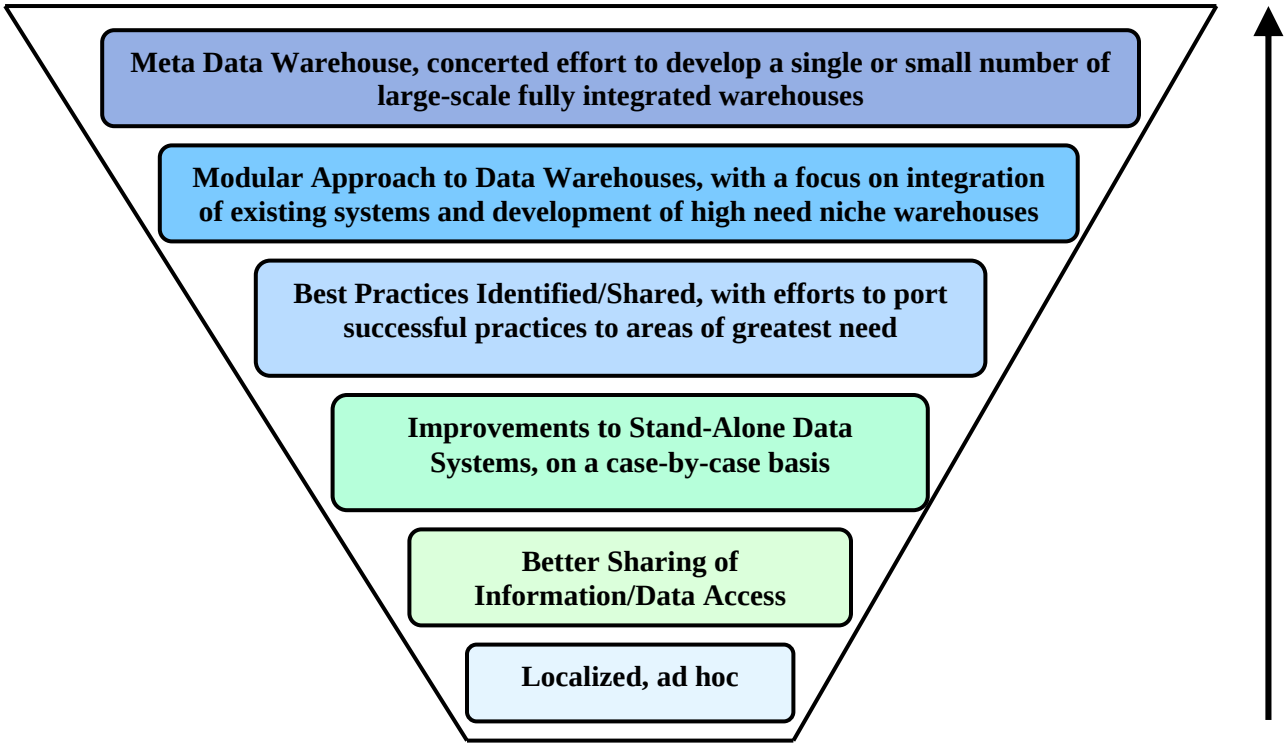
> ***Recommendation: Provide Seamless Data Interfaces to Excel/BriQuery and Encourage Greater Use of Statistical Packages.*** *Future efforts to improve the campus data environment need to provide seamless interfaces to the applications of choice, Excel and BrioQuery. Furthermore, efforts to increase the types of applications used by campus data supporters should be considered, particularly with regard to encouraging the use of statistical packages that can be employed to overcome some of the weaknesses of our existing data environment.*

Future Opportunities: Range of Possible Solutions

Figure 12 represents a possible way of contemplating various approaches to improving UC Berkeley's data landscape through data integration. As it stands, our collective efforts seem to typically fall somewhere in the lowest-three tiers of possible approaches. Based on our analysis of data from the IDMG survey, the campus is hampered by its lack of data integration. Currently, we are neither efficiently nor effectively able to support many campus decision-making areas with useful data-derived information. Given our current weaknesses and current fiscal climate, however, it is not reasonable to expect that we could seamlessly create a new one-size-fits-all solution to address our current inadequacies. A multi-level approach appears necessary, with new initiatives carefully chosen based on their likely campus impact and likelihood of success. In general, projects that are higher up the pyramid

are likely to be more costly and may well have a greater likelihood of failure (though they may also offer greater potential benefits if successful). Perhaps the best approach for now is to look for likely improvement at all the different levels of the data integration pyramid, starting with smaller scale initiatives with clearly established guidelines to allow for measures of short-term and longer-term efficacy. This proof-of-concept approach appears to be both the best way to begin to tackle some our current challenges and to lay a solid foundation upon which future initiatives can be built.

Figure 12: The Range of Possible Approaches to UCB Data Integration



Report Summary and Planning Questions

This section contains the following two parts: 1) A statement of the report's rationale followed by a list of its key observations and recommendations for improving the data landscape at UC Berkeley; and 2) A list of primary and follow-up questions intended to provide guidance to decision makers when considering making particular improvements to that landscape. It is the hope of the committee that this section, as well as the report as a whole, will serve as a frame for productive dialogue and positive institutional change with respect to data use, management, and analysis at the University.

I. Rationale of the Report, and Key Observations and Recommendations

A. Rationale

Data and UC Berkeley's Mission. The survey that this report addresses was undertaken due to widespread opinion within the campus community of 1) the increasingly crucial need for data-informed decision making at many institutional levels; and 2) an uneven data landscape that often hinders appropriate data use, management, and analysis—adversely affecting the University's mission of excellence in teaching, research, and public service. This report seeks to take a substantive first step in helping the campus move efficiently toward needed solutions to a variety of significant data-related challenges faced by administrators, faculty, and staff across the campus. (See p. 9)

B. Observations and Recommendations Regarding Respondents and Response Rates

Observation: High Dispersal of Data Use and Needs. The large number of respondents [to the report] suggested to some members of the Advisory Group that data use and needs on the campus may be more dispersed than they initially anticipated. (12)

Observation: Lower Response Rates among Academic Units/Departments. When compared to the general campus control unit headcount, it appears that the respondents from administrative units are somewhat overrepresented among our respondent population, whereas respondents from academic units/departments are somewhat underrepresented. (17)

Recommendation: Seek Additional Input from Academic Units/Departments. The campus data community should interface more intensively with academic units/departments to ascertain their data use patterns and needs. (17)

Observation: Diversity of Functional Roles. The large group of respondents was characterized by a great diversity of functional roles (with many individuals playing multiple roles) in regard to data and decision making. (17)

Recommendation: Consult Full Range of Data Consumer and Producers in Future Efforts. Future data management/governance efforts need to pay careful attention to the range of potential data consumers and data producers. Vetting future efforts in light of data from this survey can be a first step in seeking to build solutions that meet the needs of the larger campus community. In many cases, additional in-depth analysis will be necessary. (17)

C. Assessment of Campus Data Environment—General

Observation: Low Campus Ratings. Taken as a whole, our campus data environment received discouragingly low ratings. (18)

Observation: Importance of High-Quality Analytical Work. Clearly, the existence of high-quality data and its security are of critical importance; but without high-quality analysis-related work, Berkeley may be severely limiting its ability to profit from that data. (21)

D. Assessment of Campus Data Environment—Specific

Recommendation: Prioritize Cost-Effective Projects that Increase Data Efficiency. Careful analysis of 1) the likely input (resources) necessary to succeed at developing useful information, and 2) the eventual yield in doing so should be undertaken early on in the process. Furthermore, the development of resources or technologies that make [the data portion of the process] more efficient for the broadest feasible array of mission-critical decision-making areas should be prioritized. (22-23)

Observation: Decisions without Data Support. The failure to use data or lack of clarity regarding possible use of data is of concern. (25).

Recommendation: Investigate Why Data Is Not Used. As much as possible, decision making should be supported with data-derived information. A concerted effort to investigate further what accounts for this (decisions made without strong data support) should be pursued, starting with a review of verbatim survey comments. (25)

Observation: Highs and Lows of Campus Ratings. Individuals working on financial issues display either the highest or second highest overall rating for each item evaluated in the survey. Individuals working on course-related data seem particularly distressed by the current Berkeley data environment. (27)

Observation: Necessity of Detailed Understanding. Preserving and enhancing excellence in the University's myriad data efforts requires a detailed understanding of the campus's overall data landscape. (28)

Recommendation: Continue Mapping and Analyzing Campus Data Landscape. As the University's organizational structures, priorities, and data needs evolve, ongoing mapping and analysis of the campus data landscape will be necessary if we are to preserve and enhance excellence in our overall data efforts. (28)

Observation: Importance of Clear Procedures to Request Data Access. Clear procedures for requesting data access appears to be of particular concern across a number of the major decision-making areas, including the following major areas: undergraduate, graduate student, faculty, academic staff, research, and other populations. (28)

Observation: Importance of Access to Data. Access to high-quality data is essential to supporting informed decision making on the campus. (29)

Recommendation: Port Successful Procedures to Problem Areas *(if appropriate)*. Since the financial sector of campus decision making has demonstrated the viability of having relatively successful procedures regarding data access issues, the possibility of porting these methods over to other areas of decision making should be explored. (29)

Observation: Existence of Data Is Not Our Primary Problem. Because the job groups who are arguably among the best positioned to assess the existence of necessary data on the campus rate this item more favorably than most other job groups, it is reasonable to conclude that in general the existence of data is not a major bottleneck on the campus in terms of supporting informed decision making. Rather, the subsequent portions of *Data-Informed Decision Making Flow* (22) appear to be of greater concern, including gaining access to data, understanding the meaning of it, conducting methodologically sound analysis, and converting it to meaningful information that can inform decision making. (30)

Recommendation: Focus Improvements on Increasing Data Access and Consistency. Because the middle portion of the *Data-Informed Decision Making Flow* (22) is overall most encumbered, future efforts should prioritize access to and consistency of data to allow for increased production of high-quality analysis. (30)

Observation: Inadequate Access to Data for Those Who Support Academic Chairs. Gaining access to necessary data appears to be particularly problematic for staff who support departmental chairs. If the staff of departmental chairs are blocked from access to necessary data, departmental-level decision-making activities are likely to be compromised. (31)

Recommendation: Address Accessibility Issues on the Departmental Level. Further investigation should be undertaken to alleviate any potential bottleneck in this regard, particularly in light of the fact that departmental chairs are involved in such a large

number of campus decision-making areas but appear in general to rate the campus situation less favorably than most other job groups. (31)

Observation: Core Business Functions vs. Planning, Analytical, and Assessment Functions. Although the integration of financial data across the campus appears to be associated with more-favorable ratings with regard to release of up-to-date data, topical areas of decision making that are complex and either human-centric (mentoring, climate, productivity, etc.) or require longer-term planning (e.g., staff succession planning, hiring policies, proposal trends, etc.) are associated with more-negative ratings. Certainly, core daily business needs (transferring funds, budget accounting, hiring employees) are of critical importance to the campus; so too, however, are broader-scale planning issues, and human-centric areas that directly relate to recruitment, retention, and productivity of employees and students. In general, the campus appears currently stronger with regard to meeting the immediate needs of core business functionality, and weaker with regard to planning, analytical, and assessment functions, including those involving human-centric climate issues. (31-32)

Recommendation: Improve Deans' Access to High-Level Data. Since deans are frequently charged with dealing with non-business planning, analytical, and assessment issues, their tendency to rate the campus poorly with regard to release of up-to-date data should be addressed. (32)

Observation: Possible Danger to Sound Decision Making. Although the campus in general received relatively more-favorable marks regarding accuracy and quality of data than many other items, one or more sub-topical areas are associated with a lower rating (e.g., below 50% excellent/good). These present a danger or perception of poor-quality or inaccurate data compromising the decision-making process. (33)

Recommendation: Investigate Perceptions of Data Inaccuracy and Mitigate Any Identified Problem. Any sub-topical area that is associated with a lower rating (e.g., below 50% excellent/good) should receive further investigation to determine whether there are inaccuracies in the data with an eye to improvements. If data are accurate but a perception of inaccuracy exists, investigate what accounts for this perception and seek to mitigate it. (33)

Observation: Poor Consistency of Data Fields across Systems. The fact that systems managers and policy analysts—who likely possess significant expertise in this area—are particularly likely to rate consistency of data fields across systems in the negative suggests that this is a particular area of concern for the campus. (34)

Recommendation: Make Consistency of Data Fields across Systems a Campus Priority. The lack of consistent data fields across campus systems and clear definitions undermines our ability to conduct high-level analysis, support well-informed decisions, and represent ourselves in a consistent and clear manner. The campus as a whole needs to prioritize consistency of data fields across existing systems and in future efforts to improve the data landscape. (34)

Observation: Unequal Knowledge and Access. Institutional research and policy analysts rate existence of necessary data, access to user-friendly reporting tools, and access to analytical tools to help with data more favorably than many other groups. Indeed, some respondents not in these analyst groups noted in the survey that they had no idea that so many data systems existed on the campus. (35)

Recommendation: Disseminate Inventories of Data and Analytical Resources. The campus should develop and effectively disseminate clear and easily digestible inventories of existing data and reporting tools. (36)

Observation: Access to Data vs. Security of Data. Though there is an inevitable tension between strong data security and ready access to data, both are essential to furthering Berkeley's mission. At present, the campus is rated more favorably for securing data than providing access to data. (36)

Recommendation: Improve Access to Data While Maintaining Security. As we move forward, efforts to increase access to data should be prioritized while security of data maintained. (36)

E. Looking Ahead—Creating a Better Data Environment

Observation: Importance of Input/Output Analysis of Proposed Data Projects. The *best scenario* or ideal project is, of course, one that requires low input but results in high yield. Projects that fall into this category are, however, not particularly common; and much data work that exceeds core business transactional functions (e.g., hiring someone, departmental/unit budgets, course enrollments, etc.) requires high input to derive meaningful results. Obviously, high input/low yield activities represent the *worst scenario* and should be avoided. Unfortunately, these are not as uncommon as we would like, and it can frequently be difficult to assess whether the end product will be highly beneficial. So, too, it can be difficult to assess how resource intensive a particular project will be. (37)

Recommendation: Evaluate Likely Benefits and Costs of Proposed Data Projects. Plotting out the likely level-of-input (and then perhaps doubling it as an upper bound), and then imagining the eventual benefits of successfully completing project should be in the forefront of participants' minds prior to engaging a particular project. (37)

Observation: High Dispersal of Data Needs Suggests Need for Nuanced, Modular Solutions: The high dispersal of decision-making activities on the Berkeley campus suggests that a one-size-fits-all solution to our data problems is unlikely. (40)

Recommendation: Encourage Appropriate Integration of Data Resources and Tools to Meet Local and Campus Needs. Because our data environment has evolved organically over time—largely in a decentralized manner—improvements to it should be made through a concerted iterative effort that preserves some degree of decentralization. (40)

Observation: Meeting Diverse Needs. Success in improving Berkeley's data environment will depend on meeting the needs and priorities of all involved functional groups, spanning the complete range of decision making and support functions. (42)

Recommendation: Maintain Ongoing Communication among Different Functional Groups. In order to bring about successful improvements in Berkeley's data environment, the campus should develop a mechanism that ensures adequate and efficient communication and consultation among all involved functional groups. (42)

Observation: Importance of Deans and Chairs. The unique role of deans and chairs and various groups of niche administrators (e.g., assistant deans, associate chairs, et al), their support staff, and their data needs is important to consider in future improvements to the campus data landscape. (45)

Recommendation: Consistently Consult with Deans and Chairs. Although it may be challenging to solicit deans' and chairs' feedback because of the unrestricted nature of their positions and accordant work commitments, undertaking the effort to do so appears essential given the specific nature of their needs and their critical functional role at the University. (45)

Observation: Importance of Existing Data Use Patterns: The general patterns of use of existing campus systems lead to the following conclusions. Core business functions, student-related systems, and human resource systems hold the most promise for improvement through larger-scale integration. Other systems, such as those that deal with infrastructure/facilities, course-related data, or development hold the least promise for this type of integration. (46)

Recommendation: Consider Data Use Patterns in Future Efforts. Based on use patterns, larger-scale integration should be explored for core business functions, student-related systems, and human resource systems. Smaller, more modular approaches should be explored for systems dealing with such issues as infrastructure/facilities, course-related data, and development. Existing data use patterns should be one factor in the consideration of future design efforts, not the sole determinant. (47)

Observation: High Use of Excel. Excel is used much more frequently than statistical packages by institutional research analysts on the Berkeley campus, our resident data experts. (48)

Recommendation: Provide Seamless Data Interfaces to Excel/BriQuery and Encourage Greater Use of Statistical Packages. Future efforts to improve the campus data environment need to provide seamless interfaces to the applications of choice, Excel and BrioQuery. Furthermore, efforts to increase the types of applications used by campus data supporters should be considered, particularly with regard to encouraging the use of statistical packages that can be employed to overcome some of the weaknesses of our existing data environment. (48)

II. Planning Questions *(First Draft)*: Asking Difficult Questions on the Front End

Given the fact that the data findings contained in this report point to a number of deficiencies in our current campus data environment but also point to some promising future directions, future campus efforts should be vetted carefully based on what we have learned from the survey data. This section offers metrics and detailed follow-up questions based on the report observations and recommendations that can be used in the evaluation of future and ongoing initiatives (for example, see *Figure A-31* in the *Appendix* for currently planned projects reported by IDMG respondents). Obviously, these metrics are not relevant to every possible initiative, but promising initiatives are likely to span many of the below areas of inquiry and examining them in light of what we have learned from the survey is likely to be beneficial.

Metric 1: Will this effort lead to a significant increase in the efficiency of important data-related work on the campus (see input-yield matrix on page 37)?

Detailed follow-up questions: How will it increase efficiency (establish current status and how it is likely to improve efficiency and to what degree, quantify if possible)? What mission-critical areas are most likely to be positively impacted (specify as much a possible, e.g., undergraduate learning outcomes are likely to be improved because departmental staff will have direct and immediate access to high-level information on course-taking patterns)?

Metric 2: Will the effort increase the degree to which campus decisions are made based on information derived from high quality data?

Detailed follow-up questions: What areas of campus decision making are likely to be impacted? To what degree is data-derived information used in making these decisions? If it is not, what are the current barriers and how best can they are removed? How is this particular effort designed to overcome existing barriers?

Metric 3: Is this effort modeled upon past campus successes and designed to avoid past campus failures?

Detailed follow-up questions: What efforts have been attempted in the past in regard to this area of data-informed decision making or similar areas? Which have succeeded and which have failed, and why? How is this effort designed to build on past campus successes and avoid past campus failures?

Metric 4: Does this effort focus on the middle-portion of the decision-making flow chart (22) the area of our lowest campus ratings—where data is transformed from raw data to meaningful information)?

Detailed follow-up questions: What existing data does this effort draw upon? How does the effort seek to facilitate the transformation of the existing data into meaningful information? As seen in the diagram on page 22, which stages of the flow chart pertain to the current effort and how will progress be made in regard to each specific area?

Metric 5: Will this effort help departmental chairs and staff in their decision-making activities?

Detailed follow-up questions: What specific aspects (if any) of departmental chairs' decision-making portfolios does this particular effort seek to assist? How well are the needs of chairs and their departmental staff currently met in this regard and how does this current effort promise to improve the situation? To what degree have chairs and their staff been directly consulted in the design of this effort?

Metric 6: Will this effort increase our ability to plan, assess, and make good decisions in regard to human-related issues (retention, success, renewal, etc.)?

Detailed follow-up questions: What specific planning/assessment areas does this effort seek to improve (if any) and what human-related issues are directly related to it (if any)? Given the complexity of these areas, what types of methodologies are to be employed in this effort to assure its likely success? Will this effort require a fundamental restructuring of existing data (or collection of new data)? If yes, what technologies will be used to undertake this effort?

Metric 7: Will this initiative contribute to the consistency and accuracy of data on the Berkeley campus?

Detailed follow-up questions: Have the architects of the initiative surveyed the existing data and related data definitions on the campus in regard to this topical area? How does this effort attempt to sync existing and new definitions and communicate them to the broader Berkeley campus? What efforts will be put in place to assure the accuracy of data (particularly if their have been concerns regarding accuracy of data in the past)?

Metric 8: Does this effort include a build-out of user-friendly reporting tools (particularly web-based ones) that seamlessly integrate with preferred data analysis applications (e.g., Excel, BrioQuery, statistical packages)?

Detailed follow-up questions: What specific types of reporting tools are proposed? Have sponsors of the initiative discussed the design of these with potential users? How will the new reporting tools interface with popular data analysis applications?

Metric 9: What efforts are included in this initiative to communicate to the broader Berkeley campus the data that pertains to this effort and potential positive impacts of the effort to the larger community of Berkeley data producers and consumers?

Detailed follow-up questions: Does the effort include some form of communication plan? Has there been an attempt to identify potential customers (data producers/consumers) once the effort is complete and convey to them the purpose and potential benefits of it? How does the effort seek to encourage multi-directional communication (fostering productive, iterative feedback cycles between sponsors and potential beneficiaries).

Metric 10: How will this initiative increase access to data among appropriate campus populations?

Detailed follow-up questions: What specific populations (data producers/consumers) are likely to benefit from this effort? As access to data is increased, what mechanisms will be put in place to assure the proper and secure use of data? Will this effort help to overcome the sometimes parochial and insular nature of unit-specific initiatives on the Berkeley campus (i.e., how does the effort foster multi-unit or campuswide access to data)?

Metric 11: Does this initiative demonstrate the necessary sensitivity to the dispersed and diverse nature of decision-making activities and data needs on the campus?

Detailed follow-up questions: How will the effort seek to meet and respect the needs of diverse constituency groups on the Berkeley campus (e.g., different units, different campus clientele, etc.)? Does the effort seem like a one-size-fits all effort or is it flexible enough (perhaps modular) to adjust to a wide-range of contextual needs on the campus? How will effort walk the fine line between increased efficiency associated with greater centralization and our inherited tradition of unit autonomy and self-definition?

Metric 12: Is the effort sensitive to the needs and concerns of the range of functional groups (in relationship to data and decision making) on the campus?

Detailed follow-up questions: Have a wide range of functional groups (e.g., decision makers, CAO/MSO's, research and policy analysts, systems experts, administrative support staff, individuals who enter data and/or run a high volume of reports, etc.) been consulted in the design of the initiative? What aspects of the initiative reflect the need to respect the concerns of data producers and data consumers (i.e., ease of entering data, ease of extracting meaningful information, etc.)?

Metric 13: Have the architects of the effort carefully mapped-out relevant existing data use patterns on the campus and considered this in the design of their effort?

Detailed follow-up questions: What type of evidence have the sponsors collected to demonstrate existing use patterns of data and immediate and longer term data needs? Based on this analysis, how does this current effort offer the hope of increased efficiency (via increased integration and centralization of data) without the loss of greater flexibility associated with a smaller scale more modular approach? In other words, does the current effort appear to be scaled correctly (based on empirical evidence): not too small or too large?