

Carta: Studying College Pathways at Scale

20 December 2016

Tum Chaturapruek, sorathan@stanford.edu

Michael Bernstein, msb@cs.stanford.edu

Sara Cina, saracina@stanford.edu

Ramesh Johari, rjohari@stanford.edu

Andreas Paepcke, paepcke@cs.stanford.edu

Mitchell L. Stevens, mitchell.stevens@stanford.edu

Ben Dvorak, bdvorak@stanford.edu; Srinivas Malladi, malladi1@stanford.edu;

Daniel Ortega, dgalort@stanford.edu; Margaret Shen, marshen@stanford.edu;

Albert Tung, atung3@stanford.edu; Serena Wang, serenaw@stanford.edu

Pathways are the sequences of choices students make in college and into the workforce, and the organizational contexts in which those choices are embedded. Pathways are crucially important for postsecondary improvement, because variation in how students move through the very same schools is related to variation in evident learning and satisfaction with college (Chambliss & Takacs 2014) and socioeconomic prospects in young adulthood (Armstrong & Hamilton 2013). Additionally, students' gender and ethno-racial identities co-evolve with academic course selection (Charles & Bradley 2009; Charles et al. 2009).

Until recently, technological limitations have made pathway formation resistant to systematic observation at scale. Qualitative work has yielded rich insight but is constitutionally limited in scope and representativeness. Quantitative work is either cross-sectional or carried out in panel designs with few moments of data collection, obliging researchers to make broad inferences about process and causation. Detailed institutional data that might enable more nuanced analysis is carefully guarded by college officials in the interest of regulatory compliance and student privacy. Moving beyond observation is also essential for scientific inquiry into the evolution of college pathways can advance.

Carta, a web application for course exploration and planning, enables observation and experimental inquiry into student pathways at scale. Carta visualizes data from registrar records and student course evaluations at Stanford University to assist Stanford students in course selection, makes prior aggregated choices useful to current students as they consider and select courses over multiple terms, and makes visible the sequential choices that accumulate into academic careers. Access to Carta is free to the Stanford community and fully voluntary. Since its official launch in August 2016, Carta has served approximately 6,400 unique users.

Among them are 4,800 undergraduates representing 72% of the undergraduates at the university, with comparable adoption rates across all academic divisions.

Carta is a research project in ongoing development by an interdisciplinary team of Stanford faculty and student researchers. Carta enables fine-grained scientific observation of the practices of academic search, decision-making, course selection, and the progress of academic pathways to build tractable knowledge for postsecondary improvement at Stanford and elsewhere.

Ultimately Carta is not being developed for Stanford University alone. We are building it in order to create shareable algorithms and interfaces that can be effectively deployed across a wide variety of schools. We fully expect Carta to enable a distributed research program among colleges and universities nationwide.

In what follows we first describe Carta, its data sources, and user interface and report summary statistics on use to date. Second, we outline the wide range of research directions Carta enables, focusing specifically on consequences for both educational and computational social science. Finally, we specify important normative questions attending the development this research technology.

I. Building Carta: Design and Data Sources

From the beginning the Carta team has prioritized the student user experience. We devoted the first eighteen months of the project to developing an interface that would be intuitively attractive to undergraduates. As noted above, participation in and use of Carta is voluntary. We believe that wide adoption to date is primarily a reflection on the student-centric design and ethos of the platform. In this section we describe the interface, the data sources that feed it, and some usage statistics.

Interface

Figure 1 is a screenshot of the current Carta interface, which provides a number of modules of interest on courses including: basic course catalog information; information about weekly hourly workload (“Intensity”); when students take courses in their pathway (“Sequencing”); grade distributions; and data from student course evaluations.

Students interact with the platform in the following ways:

1. *Search*: students can search for classes by department name, course number, or course title.
2. *View*: students can click to view details of particular classes.

3. *Pin*: students can pin particular classes, which saves the them on the left-hand panel and initially populates class meeting times on a weekly planner.
4. *Unpin*: Students can reverse pinned selections.
5. *Toggle classes on schedule*: Students can choose, among their pinned courses, which ones appear on the weekly planner.
6. *Hover on a bar graph*: Students can hover on each bar graph to see precise percentage numbers for each bar.

CARTA Catalog Search by department, course number or description

ENGLISH 90: Fiction Writing Pin this course

COURSES PINNED FOR THIS QUARTER

- CS 221: Artificial Intelligence: Principles and Techniques
- BIO 41: Genetics, Biochemistry, and Molecular Biology
- CHEM 35: Synthetic and Physical Organic Chemistry
- CS 103: Mathematical Foundations of Computing

QUARTER OVERVIEW

	Mon	Tue	Wed	Thu	Fri
10am	BIO 41		BIO 41		BIO 41
11am					
12pm	CHEM 35		CHEM 35		CHEM 35
1pm	CS 221		CS 221		
2pm					
3pm	CS 103		CS 103		CS 103
4pm					
5pm					
6pm			CHEM 35		
7pm					

AT A GLANCE

The elements of fiction writing: narration, description, and dialogue. Students write complete stories and participate in story workshops. May be repeated for credit. Prerequisite: PWR 1 (waived in summer quarter).

Go to Explore Courses

Terms: Autumn, Winter, Spring, Summer
Units: 5

Grading: Letter or Credit/No Credit
UG REQS: WAY-A-II, WAY-CE

Show Schedule

Instructors: Aut 16-17, Win 16-17, Spr 16-17

Classes with similar descriptions: ENGLISH 92, GS 197, ENGLISH 190T, EFLANG 698B, EARTHSYS 250

INTENSITY

Hours per week (240 students)

Hours per week	Percentage
1-5 hrs	~15%
5-10 hrs	~38%
10-15 hrs	~25%
> 15 hrs	~22%

Average quarterly units

Category	Average Units
With this class	~12 units
Stanford average	~15 units

Enrollment outcomes (3470 students)

Outcome	Percentage
Completed	~85%
Completed (was wait listed)	~10%
Dropped (was enrolled)	~3%
Dropped (was wait listed)	~1%
Withdrawn from class	~1%

SEQUENCING

Academic years of past students (240 students)

Academic Year	Percentage
Freshmen	~25%
Sophomore	~35%
Junior	~15%
Senior	~15%
Year 5	~10%
Masters	~5%
Other	~10%
PhD	~5%

ENGLISH 90 paired with your pinned courses

Course	When students took ENGLISH 90...	Average grade in ENGLISH 90...
Before BIO 41	21%	3.96
Concurrently with BIO 41	11%	3.93
After BIO 41	68%	3.96
Sample size	280	280

Commonly taken before this course: CS 106A, PSYCH 1, PSYCH 10, PWR 1, MATH 51, PSYC 135, ECON 1A, CS 106B

Commonly taken concurrently with this course: CS 106A, PSYCH 10, PSYC 135, CS 106B, PWR 2, PSYCH 1, ECON 1A, MATH 51

Commonly taken after this course: ENGLISH 190, ENGLISH 146, PWR 2, ENGLISH 92, PSYCH 10, ENGLISH 190F, ENGLISH 91, CS 106A

Your Quarterly Time Commitment History

Quarter	Time Commitment (hrs)
Autumn '15 - '16	~30
Winter '15 - '16	~30
Spring '15 - '16	~30
Autumn '16 - '17	~45

Pinned courses: CS 221, BIO 41, CHEM 35

Figure 1: The current Carta interface.

Data sources

Carta is supported by a range of official data sources from Stanford University. We also collect data directly through the platform.

At present the Carta platform makes use of three kinds of official institutional information:

1. *Enrollment data.* We utilize over fifteen years of official Stanford enrollment data. These include timestamped records of every course registered or dropped by each student, their grades in completed courses, as well as timestamped records of the declaration of a major. These data provide a complete longitudinal transcript of the course enrollment history of every student at Stanford. To protect student identity, our production databases work only with hashed student identifiers.
2. *Course evaluation data.* We utilize course evaluation data for the last three years of courses offered at Stanford. Evaluation data are based on student self-reports. Evaluations query students on course difficulty, instructor performance, achievement of learning goals, and reasons for taking the course. Evaluation data are anonymized by a third-party vendor for any and all university use and so cannot be linked with Carta's hashed identifiers.
3. *Course catalog data.* We utilize data on course descriptions, instructors, and scheduling derived from Stanford's course catalog. We use catalog data to populate basic information about courses. Additionally we use text mining techniques to identify similar classes and extract prerequisites.

Another data source is Carta itself. These data are of two types. The first describes Carta use and is collected automatically and continually. We maintain detailed logs of entry, exit, view, and click behavior, enabling us to describe how students navigate the environment with exceptionally fine granularity. The second type is information collected through direct user queries mediated by the interface. For example, we can build in-site surveys in both cross-sectional and longitudinal designs to learn more about course selection and pathway development over time. Such mechanisms can also be used to query users about course difficulty, engagement, and quality, or to allow users to contribute their own tags and reviews to courses, instructors, and programs.

In the next phase of research we will link the above data sources with demographic data as well as information on post-collegiate jobs and careers. Demographic data will be essential for us to assess pathway variation by gender, socioeconomic status, and ethno-racial identity. Other statuses, such as participation in varsity-level athletics, are surely important for a full picture of pathway ecology (e.g. Shulman & Bowen 2002). Data describing occupational and career trajectories will enable us to observe any relationships between pathway variation and labor

market opportunities. We will obtain work and career data from multiple sources including LinkedIn and the Stanford Alumni Association.

Usage

Carta launched widely to the Stanford population on August 1, 2016. Since that time we have accumulated over 6,400 unique users including 4,800 unique undergraduate users. This constitutes 72% of the Stanford undergraduate population. On average, Carta has received 1,700 weekly active users; the peak thus far is 3,000 active users, coinciding with the week before the start of classes for the Fall 2016 term. Table 1 describes the user activities of Carta to date since launch. Table 2 breaks down Carta users by class year. Table 3 breaks down Carta users by school.

Activity	Mean per user	Standard error
search	46.1	0.9
view	39.4	0.7
pin	9.9	0.2
unpin	3.7	0.1

Table 1: User activities on Carta

Class year	Count	What fraction of Carta users are X?	What fraction of X are Carta users?
Freshmen	1657	34%	93%
Sophomore	1184	24%	69%
Junior	1043	21%	64%
Senior	973	20%	60%
Undergrads	4857	100%	72%

Table 2: User breakdown by class year

Major	Count	What fraction of Carta users are X?	What fraction of X are Carta users?
Undeclared	2947	59%	78%
Engineering	891	18%	69%
H&S Interdisciplinary	489	10%	64%
Natural Sciences	211	4%	69%
Humanities and the Arts	197	4%	55%
Social Sciences	195	4%	57%
Earth Sciences	61	1%	58%

Table 3: User breakdown by school

II. Research with Carta

Our primary goal is to scaffold research to inform and improve student progress through college. To achieve this goal, Carta must entail a close synergy between social-scientific inquiry and computational design. If it continues to enjoy voluntary and steady use and can be deployed in multiple institutional settings, Carta might substantially redefine social-scientific understanding of the college “treatment” on student learning, identities, and careers. At the same time, iterative inquiry produces insight vital for improving the design of the platform itself. The entire enterprise depends on applying computational advances to benefit social science, and using social-scientific inquiry as a mechanism to advance the computational design and the quality of the user experience.

A. Educational Social Science

As a tool for educational social science, Carta enables progress on at least three topics: (1) academic decision making and pathway evolution; (2) pathways and identities; (3) the relationship between college pathways and the evolution of young adulthood. We provide brief summary rationales for each domain below. We fully expect that areas of inquiry will multiply as the enterprise matures and becomes distributed across multiple researchers and institutions.

1. *Academic decision making and pathway evolution.* Pathways are fateful for students but have long been resistant to observation at scale. In a manner analogous in some ways to ethnographic research, Carta enables researchers to observe how students pursue academic pathways *in situ*. Beyond observation, Carta provides a mechanism for interacting with students even while they are in the process of considering and selecting courses, both cross-sectionally and over time. In light of the prior qualitative findings on student pathways (Armstrong & Hamilton 2013; Chambliss & Takacs 2014) we are interested specifically in: (a) sources of information (in addition to Carta) on which students rely when considering courses; (b) the relative influence of significant others -- friends, teammates, roommates, parents, and professional advisors -- at the moment of particular decision events and over the college career; (c) variation among students (e.g., by gender, socioeconomic status, class year, or extracurricular commitments) in the process or outcome academic decision making.
2. *Pathways and identity formation.* Identity formation is the process through which human beings come to define themselves as certain kinds of people, for example as “men,” “women”; “Black,” “Latino”; “smart,” “gifted” -- or “the sort of person who does well” in a particular school or field of study (Steele 1997; Walton & Cohen 2007). Prior research demonstrates that identity formation in college evolves reciprocally with academic course selection. Substantial variations by gender and race in field persistence suggest that academic, identity, and equity issues are intertwined (Bowen & Bok 1998; Charles & Bradley 2009; Goldin 2015).

Carta enables researchers to empirically describe patterns of academic choice-making by demographic variables such as gender and race, and to experimentally control the amount and kind of information students have to consider how their academic choices might be associated with “people like me.” In light of prior findings by others we will be specifically interested in variation in how different kinds of students respond to similar information. For example, do women respond to information about aggregate course grades differently than men? Do engineering majors utilize different information available to them on Carta than humanities majors? Might iterative course selection be related to the development of an academic self -- and if so how might we observe and/or “nudge” that process?

3. *Pathways from college into adulthood.* The journey from childhood to full independence has steadily lengthened in American society. A life stage demographers now call early adulthood stretches well into the third decade -- creating opportunity for more flexible personal development and career exploration but also new risks of financial and relational instability (Settersten & Ray 2010). Important questions for researchers and college officials attend this new reality, to wit: what is the relationship between academic pathways and post-college trajectories, and how might colleges best prepare students for their lives after college?

Bringing these large questions down to tractable size for the short term, Carta researchers will be interested specifically in the following: (a) whether academic pathways variably predispose graduates to pursue particular lines of employment; (b) the extent to which academic pathways mediate, or obviate, the role of demographic characteristics such as gender and socioeconomic status in occupational trajectories; whether there is variation by pathway in students’ reported satisfaction with college as “preparation for life” in the years after graduation.

All of the above is valuable for college administrators and policy-makers. Incremental choices sum to academic careers. Students recognize this in the abstract but may not be well equipped to see potential implications of academic choices in the short term. Neither students nor human advisors alone have had easy access to the kind of information now available to them instantaneously through Carta. As the platform and the academic culture of Stanford (and other institutions) coevolve, we eagerly anticipate contributing to change in policy and practice that will encourage decision making informed by the best possible available information.

B. Computational Social Science

A second research domain stems from Carta itself as an object of inquiry and development. Carta presents unique challenges in three dimensions: (1) the choice in which information is displayed to students; (2) the influence and observation of students’ objectives in processing this information; (3) the design of algorithms that aid students as they make decisions.

1. *Information display.* What information do we choose to display to students, and how? The range of data available is vast, and by necessity we make choices in how we display information. A key challenge here is to help students understand possible biases in the various kinds of information they receive through the interface. For example, Carta presents information about the performance of past cohorts of students who have taken particular courses. However there may be many forms of selection bias present in these cohorts. Relatively stronger students may take a particular course in Spring instead of Fall terms, for example. In the absence of this knowledge a student may mistakenly infer that a Spring offering is “easier.” An important line of inquiry is understanding (through interface design, observational study of user behavior, and algorithmic development) how we can mitigate this kind of bias in interpretation.
2. *Preference elicitation.* Students have a range of goals that might influence course selection. For example, some students will be most interested in minimizing risk of low grades, while others may be interested in maximizing new opportunities such as exposure to fresh subjects and modes of inquiry. Varied goals will lead to different desired behavior when interacting with Carta, and the same student may exhibit different goal-seeking at different points in the college career.

Administrators share the concern of balancing risk and opportunity. In our initial conversations in developing Carta, administrators and staff (including Stanford’s professional academic advisors) also expressed a wide range of views on the types of student outcomes they consider important in shaping pathways. For example, one goal for a liberal arts curriculum is to ensure that students obtain a wide variety of experiences during college. This is consistent with increasing the range of opportunities to which each students are exposed. At Stanford, one particular manifestation of this goal is ensuring engineering and science students’ meaningful participation in humanities courses. Additionally there are students for whom managing risk is a critical concern. Academic advisors are well aware that students enter college with highly variable academic preparation and expectations. At some points in their college careers, students may be too academically ambitious for their own good.

Carta can provide mechanisms for considering how to balance risk and opportunity to students on these matters. One of the most important elements of Carta is the implicit elicitation of students’ preferences through their interaction with the platform. A goal of our work is to transform our observation of student behavior into interpretation of preferences. In some cases, this suggests novel mechanisms of data collection such as user-generated tags or responses describing interests and goals. In other cases it suggests combining action with ultimate outcome (e.g., course enrollments) to infer how students perceive of risk and opportunity. This process requires a combination of data analysis from Carta itself with qualitative investigation of how students use the platform and integrate it with other sources of information.

3. *Algorithms for student decision support.* Short-term academic decisions can have long-term consequences. This means that “good” outcomes cannot simply be measured through grades or evaluations at the end of a single course. Students have goals beyond a particular academic term and beyond college itself. Similarly, schools have goals for the education they offer far beyond students’ assessment of any particular course.

Our algorithmic challenge is to build a platform that can enable students to manage both short- and long-term goals when making decisions. This includes finding strategies to leverage the pathways of past students to help current ones understand what is possible, or to break myths of what is assumed to be difficult or impossible for students to achieve.

The above are challenges for any computational social system, but have a special valence in our instance because we are building Carta simultaneously to be appealing to users and a transparent scientific tool. In this way Carta provides an ideal laboratory for developing computational social-scientific environments.

III. Normative Questions and Goals

Studying student pathways at the level of fidelity enabled by Carta comes with substantial ethical and organizational challenges. We are building the platform even while the larger higher education community is coming to recognize the transformative potential of data ubiquity and interoperability.¹ Inherited legal regulations and normative practices for safeguarding student and institutional privacy were developed in an era of paper records, and when expectations for measuring student and institutional performance were much different than at present. We view this challenge not as a deterrent but an opportunity: to begin new conversations about responsible use of student data; to develop sound procedures for ensuring responsible use; and to accrete appropriate expectations about reciprocal understanding between students, schools, and academic researchers. We here raise three questions to inform thoughtful deliberation as we move forward.

What is the proper or ideal relationship between mentors/advisors, students, and digital tools built to support them? Based on the development of information technology in other domains, we think it is wise to presume that digital tools have great capacity for improving academic mentoring and advising, course selection, and career planning at Stanford and elsewhere. Yet we recognize that human relationships are the essence of mentoring as well as personal and intellectual development. Stanford has appropriately invested a great deal in ensuring that all its students have access to experienced mentors. By virtue of this investment, alongside its global leadership in technological innovation, Stanford is ideally poised to develop hybrid tools for

¹ Stanford has taken a leadership role in specifying the most salient implications of these changes. See the resources compiled at the Stanford/Ithaka project “Responsible Use of Student Data in Higher Education” (ru.stanford.edu).

navigating the college journey. Once we have incubated them here, tools like Carta and the ethical scaffolding surrounding them can be shared with other schools.

What kinds of student pathways are acceptable and ideal? Who gets to decide? Schools and their students have complicated production functions. Academic engagement and achievement are unquestionably important goals, but there are others: excellence in athletic and other non-academic activities; timely completion of degrees; likelihood of lifelong employment; personal growth. Carta and other technologies can illuminate the course and iteration of college pathways, but no technology by itself can specify appropriate, correct, or ideal pathway purposes or goals. We believe Carta and other attendant technologies surface fundamental questions about the purpose and process of college.

How should schools and researchers distinguish between data describing students and official academic records? Colleges and universities have a fiduciary obligation to safeguard, and verify into perpetuity, the official academic records of their own students. As academic organizations they also are obliged to improve their own practice in light of accumulating information and knowledge. Given the ever-growing amounts and kinds of information schools are collecting about students, these two obligations may come into conflict unless clear distinctions are made between official academic records and those data that can be utilized in the service of organizational improvement. Newer technologies for aggregating, storing, and deploying data raise important questions about data integrity, but they also offer heretofore unimaginable opportunities to make universities better, more equitable, and more self-reflective institutions.

Works Cited:

- Armstrong, Elizabeth A., and Laura T. Hamilton. 2013. *Paying for the Party: How College Maintains Inequality*. Cambridge: Harvard University Press.
- Bowen, William G., and Derek Bok. 1998. *The Shape of the River: The Long-Term Consequences of Considering Race in College and University Admissions*. Princeton: Princeton University Press.
- Chambliss, Daniel F., and Christopher G. Takacs. 2014. *How College Works*. Cambridge: Harvard University Press.
- Charles, Maria, and Karen Bradley. 2009. "Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries." *American Journal of Sociology* 114: 924-76.
- Charles, Camille Z., Fischer, Mary J., Mooney, Margarita A., and Douglas S. Massey. 2009. *Taming the River: Negotiating the Academic, Financial, and Social Currents in Selective Colleges and Universities*. Princeton: Princeton University Press.
- Goldin, Claudia. 2015. "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College." Technical Report, Harvard University.
- Settersten, Richard A., and Barbara E. Ray. 2010. *Not Quite Adults*. New York: Random House.
- Shulman, James L., and William G. Bowen. 2002. *The Game of Life: College Sports and Educational Values*. Princeton: Princeton University Press.
- Steele, Claude M. 1997. "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist* 52:613-629.
- Walton, Gregory M., and Geoffrey L. Cohen. 2007. "A Question of Belonging: Race, Social Fit, and Achievement." *Journal of Personality and Social Psychology* 92:82-96.