The Student Data Backpack

Margaret Raymond

Margaret Raymond is director of the Center for Research on Education Outcomes (CREDO) at Stanford University

The world of education data is rapidly evolving. As accountability policies exert more pressure on schools to demonstrate student achievement, educators are becoming more focused on using available information about their students, resources and practices to understand current levels of performance and to glean possible paths to improvement. So-called evidence-based or data-driven organizations put rhetoric and conventional wisdom to the test, and thereby chart a truer course to effective teaching and subsequent student learning.

Or so the theory goes. As appealing as the rosy picture is in the abstract, the reality in schools and districts across the United States is more mixed. To be sure, there are local education agencies (LEAs) that have well developed information systems, or whose leaders understand the value of regular review of their schools' efforts and their effect on students' progress. But even these enlightened instances face the same hurdle that confronts so many others. The entire data-driven enterprise is a house of cards if the data that serve as its foundation are filled with errors, are incomplete, or do not capture the details that educators and policymakers most need.

To realize improvements in student achievement, accurate and complete data on students should be made available to educators, policy analysts and

other decision makers in a timely manner. Those decision-makers require confidence that the data have integrity, that they are complete and accurate. This chapter diagnoses the multiple points of failure in current practice that result in poor data quality — including people, data system architecture, and prevailing information management processes — to illustrate how the incentives to collect and manage data that are complete, accurate and timely are diluted. The diagnosis begs a new approach that will create one essential subset of education data, or information about students and their backgrounds.

The Student Data Backpack proposed here creates a different mechanism for collecting and maintaining certain kinds of student data that are currently collected in flawed ways. It begins with a data file that interacts with both parents and LEAs. The Student Data Backpack will be attractive to families because it provides an easy way for parents to enroll their child in school, but its real benefit is that it contributes to the completeness, accuracy and timeliness of critical data. This will give educators and policymakers a firmer foundation for their work, and parents will become fuller partners in their children's education. The Student Data Backpack also uses a social networking model to support and enhance parents' natural interest in their child's education.

Diagnosis of Current Data Quality

Poor data quality has important consequences for schools and students. Schools can lose funding if they undercount attendance or delivery of program services such as special education or Free and Reduced Price Lunch. Auditing data quality and correcting data errors are costly and LEAs usually avoid doing them in order to economize. But if details of student progress or teacher support programs are flawed, schools could allocate their resources imperfectly, potentially perpetuating ineffective practices or terminating successful ones.[1]

Given the value of sound data, one might reasonably question why LEAs do not undertake programs of improvement in their information technologies and practices. The manner by which schools and districts gather, store and update their data today is less the product of careful planning and design than of gradual evolution and marginal adjustments. Some effort has been devoted to crafting unifying solutions, but to date these efforts have been at the margin. A more general re-engineering of existing data collection, transfer and storage practices as is proposed here has not been considered.

The sections below summarize some pressing data quality challenges and offers some diagnoses.

Flawed Student Data Collection Practices

The initial point of data collection is the single most influential moment to ensure data quality, yet it typically receives the least attention.[2] When a parent wishes to register his or her child for school, a personal visit to the school or district central office is required. Identity and required immunizations are verified, and then parents fill out registration forms that are populated with the data fields the state, district or school requires. These data are then input by district personnel into computerized databases. Input errors are common, and parents are often dissuaded from providing full information about their child out of embarrassment or fear of having their child relegated to inferior opportunities. While distortions are probably inevitable, the typical set-up exacerbates rather than minimizes that risk.

After this initial encounter, parents are asked repeatedly to supply much of the same information in a variety of forms, such as emergency contact information, known food allergies, permission slips, and so on. However, rarely is any effort made to check the accuracy of the data or update the original data record, which can rapidly become outdated, especially the address and telephone information for mobile populations. The majority of mobile families do not forego telephone or cable connections when they move, but there is currently no mechanism for maintaining current information for mobile students.

There is little research on how widespread the problem is. In the course of developing data for a national study, one research group graded the student data sets provided to it by schools and found that while a few schools delivered flawless data, the average school had errors or missing values in over 20 percent of the fields.[3] State education departments have found it necessary to invest extensively in electronic data checkers to examine information provided by LEAs before allowing that data into state education agency (SEA) databases.

Two fundamental problems are evident in this description. The first is that current data collection practice presumes that student data have a long half-life, but for significant numbers of students, the assumption is flawed when it comes to things like phone numbers and addresses. Second, once data are gathered by LEA personnel, there is no ownership of the duty to maintain currency or quality, since each opportunity for collection is treated independently.

Data Storage

The databases into which a student's information is entered create their own barriers to high quality data and its use. Each database is proprietary and has its own data dictionary (the list of variables that it contains and the formats for each variable). Once a state or school district adopts a vendor and its data dictionary, it is quite costly to swap vendors. Indeed, the vendors have created that barrier as a means to retain their customer base.

Two significant barriers result from the fact that variables and formats are not standardized. First is the difficulty of data exchange with other information systems, such as transportation management or food service applications. It is common practice that each system collects its own data on students, often at different points in time, so that inconsistent data on students exist across the various applications. The second is that unique data dictionaries make it difficult for schools and districts to use their data easily to file state mandated reports; often, customers must pay for an additional layer of software or programming to manipulate the contents of the database into the formats required by the state education departments. Thus, even if an LEA collects the "correct" data on students and their backgrounds, the way the variables are collected and the formats that the variables assume in different applications can make it challenging for LEAs to access and rely on the data they have on hand.

Recent developments point to a more promising future. The U.S. Department of Education has placed pressure on vendors of student information systems to make their database structures more uniform so that information can be exchanged across vendor platforms. Under the Education Data Electronic Network (EDEN), state education departments are required to use a uniform data dictionary when reporting on federal education program activities in their state, beginning in the 2006–07 school year. Early indications suggest that data coming from the SEAs are slowly converging on the EDEN requirements, which means that SEAs are shouldering the burden of translating the multiple coding formats from LEAs. It is clear that LEAs will inevitably be required to conform to the new variable formats.

The problem of having volumes of information isolated from each other—so-called silos—is common with information management systems generally, and has a well documented history in business and other fields.[4] Public education lags behind other sectors in the design and use of information

technology. So ingrained is the silo approach that in one state, California, a seven-year-long redesign of the SEA's data systems has created two free-standing databases with an extremely narrow set of overlapping fields and no plans to provide real-time linking of the databases.[5] (See RiShawn Biddle's pages in this volume for more on California's struggles.) In creating a new data system, one large urban district spent as much money on programs that would recode data so that the various silos (with their different data dictionaries) could use it as it did for the rest of the project.

These workarounds can be developed by states or districts to link their stand-alone systems, but they are expensive to develop and maintain. More importantly, they are marginal adaptations that fail to address the fundamental problem of interconnection — namely, how to create common standards for data and electronic data files to enable different software applications to share information easily.

A national effort by the Schools Interoperability Framework Association (SIFA) began in 1997 to establish common standards for data and data sharing. These standards enhance the ability of education software applications to exchange data across different departments within an LEA (e.g., instruction and curriculum planning, food service, transportation, or health), between schools (e.g., transfer of student records), or between LEAs and SEAs. After ten years of activity, SIFA has several common standards to show for their efforts; vendors can earn SIFA certification if they adapt their products to comply with the standards. The progress has been slow, but now that EDEN compliance is mandatory for SEAs, the pace can be expected to increase.

The common interface standards for student-level records make it possible to develop ideas such as the Student Data Backpack with confidence that student information system (SIS) vendors soon will be able to accept universally formatted data into their platforms electronically.[6] This capability would eliminate a lot of the conditions that lead to data input errors and redundancy in the current landscape, but still would not address some basic problems of maintaining accuracy and currency of the data.

Barriers Identified

The preceding discussion lays out how current practices contribute to the problems of low data quality and thus low confidence in analysis. Two main causes are at work: the first is that the incentives to collect and maintain full

and accurate data on students are flawed. The consequences of bad data quality are often felt long after the data has been collected, so the incentives to "get it all, get it right, and get it in the system" are pretty weak. Moreover, the people responsible for data gathering and input (most often school district clerks or school office managers) are largely uninvolved with any downstream use of the data, so they typically don't have a strong drive to ensure their work is accurate and complete. Once incomplete or inaccurate data are transferred into the system, they are costly to correct.

The other root cause is data "balkanization." Having multiple and isolated data systems in LEAs makes it difficult to ensure that all data are current, or that missing data are identified and addressed. There is a clear need for interoperability standards and vendors are incorporating them into their products. The progress in this area makes it feasible to conceive of new models of data collection, usage and integration such as the Student Data Backpack.

Clearly, technology impediments are not the only cause of information silos. The upside to interoperability extends beyond operating efficiency to the realm of clearer insight into the workings of schools. Political challenges arise whenever mention is made of consolidating information about schools, students and programs. The obvious opportunities that arise from integrating data silos, such as the ability to "connect the dots" about the performance of leaders or teachers, or the potential to expose favored programs or illuminate preferential resource allocations create significant anxiety whenever the subject is broached.

A new model of student data collection, one that advances beyond the marginal changes of the past, could be a vehicle for a variety of improvements that would lead to student data that are more accurate, complete, and timely, such as:

1. Making corrections, updates, and student moves available to school personnel in a timely manner.
2. Aligning the incentives for high quality data.
3. Creating greater capacity for parents to be full partners in their children's education.
4. Leveraging new technologies that can facilitate constructive sharing of information to improve student academic outcomes.

The Student Data Backpack

The challenge remains to develop a mechanism for gathering the data on students that results in better, more timely information for schools and districts. To be successful, the solution must provide adequate incentives for parents to regularly update their child's information. Parental sense of duty will only carry so far, so it is necessary to ensure that parents derive value themselves from their investment of time.

The notion explored here is the Student Data Backpack, an independent web-based data service that operates as a central clearinghouse for student data. Building on successful business models of internet information services, the Student Data Backpack envisions an electronic data file that exists independent of LEA information systems but supplies those systems with the data they need and provides feedback about the student to parents along with other valued resources.[7]

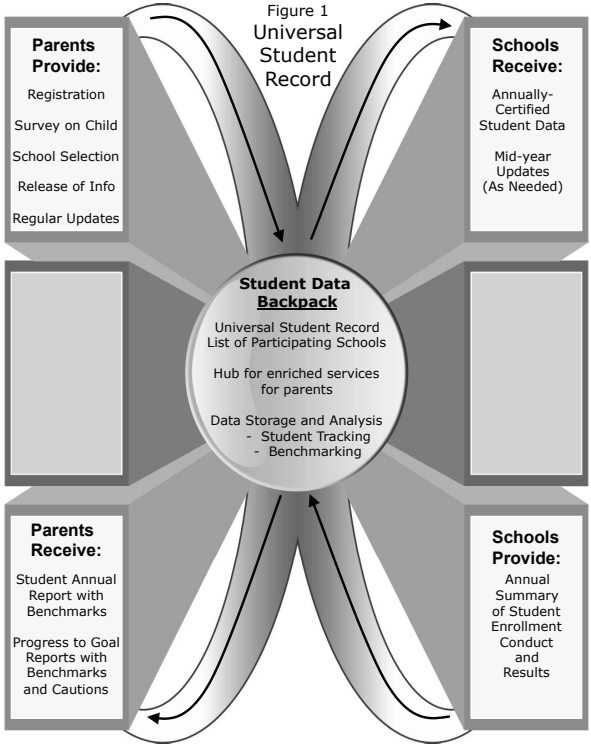
The Student Data Backpack contains a suite of products and utilities that parents can access over the internet. There are three essential components. First, it includes a universal student record (USR) containing students' personal information, enrollment histories, achievement results and academic experience. The second component is a data transfer function that interacts with LEA data systems to deliver and collect information on students. The third component provides parents with a variety of tools, resources and opportunities to interact with other users. The result is an online community for parents, centered in their role as "chief education customer," and extending to other facets of life for their children and themselves.

Parents register with the Student Data Backpack and receive a user account, similar to what occurs on other websites such as Amazon.com. Parents can use a single account to create profiles for each of their children. The Backpack gathers information from parents via a structured web survey designed to gather all the details needed to populate an EDEN-compliant universal student record. The USR is made up of variables about students such as date of birth, demographic characteristics, emergency contact information, English language proficiency, special education needs, and eligibility for subsidy programs such as Free and Reduced Price Lunch. The record also includes the unique student identifier each student receives from their state education department to support the linking of data over time for each student, a prerequisite for calculating learning gains from year to year.

Each time parents return to the Student Data Backpack site, they would be queried about changes in their child's profile. Should changes be made, a utility contained in the Student Data Backpack would initiate an update sequence.

The Student Data Backpack serves as a broker between parents and LEAs. Utilities associated with the Backpack would allow the parent to direct the student record to the school the student will attend. At the point of transfer, parents would have the ability to designate, beyond the uses required by the state or district, the degree of sharing of their child's information. For example, parents may be open to releasing the child's information to local social service agencies to see if he or she is eligible for youth-oriented programs. Or parents may be interested in releasing data to support ongoing research about national school improvement efforts. Thus for the first time, parents could exercise their discretion to release information about their children in a manner consistent with the original intent of the Federal Educations Records Privacy Act (FERPA).

Figure 1
Universal Student Record



They could choose which portions of their child's records could be shared at varying levels of disclosure.

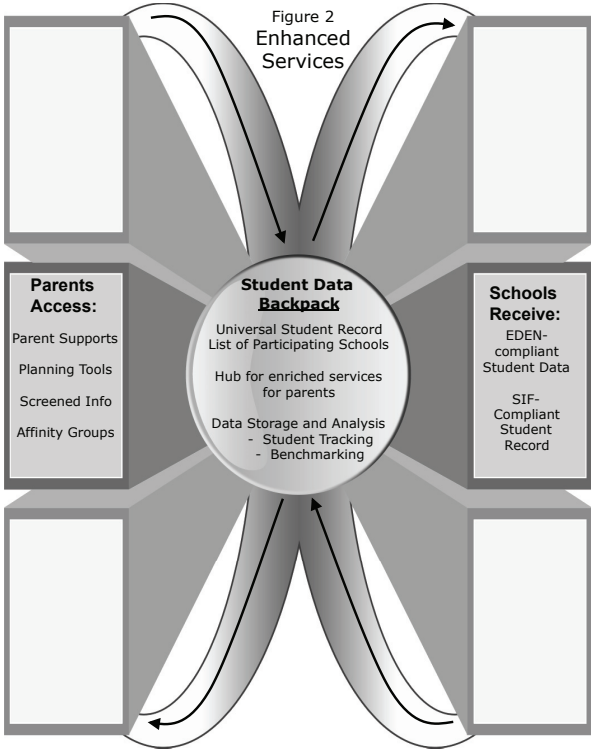
The handoff of a student's record would mark the official designation of a parent's choice of school for their child. Additional information required by a district or school to complete a student's registration would be exchanged at that time. The Student Data Backpack would then deliver the USR and the authorization for release of student information to the LEA data system. Any future updates would also be electronically transferred to the LEA on record for the student. Figure 1 shows a flowchart that describes the Student Data Backpack from the perspectives of the parent user and the school or district where the student enrolls.

Advances in interoperability standards made by the Schools Interoperability Forum Association make it feasible for the Student Data Backpack to be built with SIFA-compliant interfaces for data intake and transmission. The standards for exchange of data are being met with increasing prevalence as vendors make upgrades to their products, so it is reasonable to expect that the USR could interact with a growing number of information applications. Indeed, the number of LEAs and states that are incorporating SIFA requirements into their vendor agreements has more than doubled each of the past three years.[8]

Local education agencies would configure their student information system to accept incoming Backpack records, which would undergo the same quality checks that exist for other forms of data input. Once in their systems, the data become indistinguishable from other sources of data used by local education agencies. When the time comes for parents to receive report cards or other assessments of student progress, an output report would be produced in the student information system and transmitted to the Backpack platform, which would update the specific child's record and notify the parent.

Whenever LEAs deliver data about each student's activities and achievements back to the Student Data Backpack, it would then format the data on each student for easy viewing. Parents would have the ability to see electronic report cards showing the child's attendance, grades, and formative assessments results. As a bonus, the Student Data Backpack might include comparisons with similar students. It might also flag areas where a student might need additional effort and support.

At the end of the academic year, a completed transcript of the student's experience would be transferred electronically to the Backpack, which would



incorporate the final material into the permanent record for the student. Parents would have the duty to verify the record and notify the school if they intend to continue enrollment of their child in the school. If the student will return to the school, the school SIS can process the record as a continuation record; if the student leaves the school, the SIS will archive the record. The anonymous record would still be available to be included in analyses of personnel, programs, and services, but the school would not have the student in its active database.

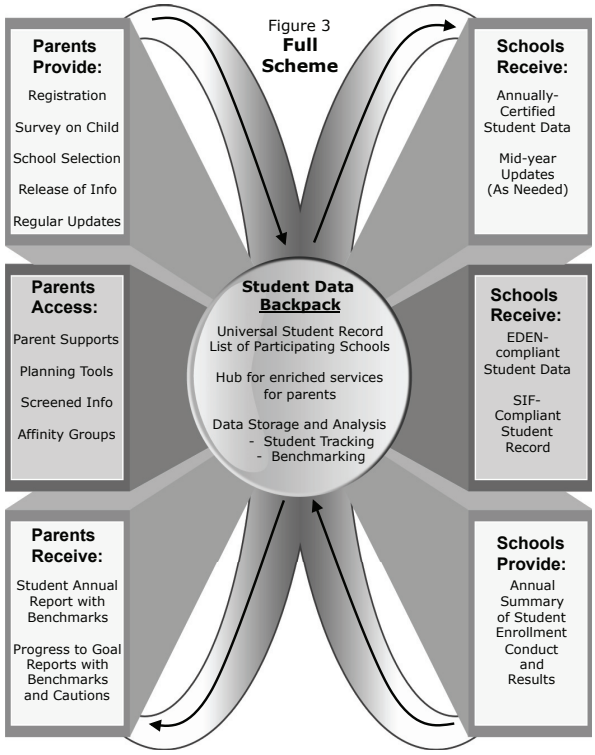
The Backpack could contain a variety of resources to help parents take a proactive role in supporting their child's development and education (see Figure 2). Examples of this kind of consumer-oriented content are seen today in the health and medical care sector; in the field of education, the potential impacts might even be greater since parents are also consumers, taxpayers, voters and advocates for their children.[9] One example could be a digital record of child immunizations and health. Alternatively, the Student Data Backpack might include access to parent discussion boards, informative videos about

parenting or child development, immunization and health record keeping, or planning tools to track their child's progress towards graduation and postsecondary endeavors.

Digital social networks are conventionally associated with younger users. MySpace and Facebook focus on younger American users and claim millions of user interactions per month.[10] But the phenomenon is rapidly becoming commonplace with other groups—grandparents have their communities, as do affinity groups such as Vespa drivers, vegetarians, and travelers. Interestingly, across all age groups, females are more inclined to engage in online social networks than males, a trend that bodes well for an education-oriented site.[11] These sites have learned that users not only benefit from the information or content that is available, but also derive personal value and satisfaction from affiliation and interaction. By serving the social needs of users, even to a limited degree, the information provider implicitly validates the participation of the user in whatever group they participate, and creates community-wide standards of conduct which have been shown in recent studies to positively affect user behavior. The effect of this should not be trivialized. For instance, people who participate in smoking cessation groups are encouraged to turn to the web community when cravings hit. This in turn supports and reinforces the original goals of the participant, typically leading to better outcomes than if the participant had been left to struggle alone.

As envisioned here, the same universal student record could serve as the common foundation in every Student Data Backpack, but various vendors offering a Student Data Backpack could develop their own blend of information and services to entice parents to use their version. The chance to tailor content to specific parental interests would motivate vendors to manage USR collection and storage and treat the resulting base of parent users as a receptive channel for the vendor's own mix of tools, information and services. This approach has been used successfully in numerous market niches, from financial services, to management of chronic health issues, to entertainment. The full schema is presented in Figure 3.

Because each Student Data Backpack vendor would have records on many other students in its databases, it would have the capacity to create many benchmarking profiles against which to compare a student's development. For example, a student might be compared with others in his or her grade and district, as well as with other students matching a personal profile, and so on.



With benchmarked student performance, parents become more powerful advocates for their child in particular, and overall school quality in general.

Scenario of the Student Data Backpack

Lee Jones is a single parent of two school-aged children and has recently moved to a new community to pursue employment. With limited time and information, Lee is interested in enrolling his children in schools that will meet the challenges of one child's mild speech disability and the other's keen interest in science and mathematics. The school district in the new community has established an association with the Student Data Backpack, allowing Lee to register his children for school via the website.

Lee uses the web browser to locate the site. After a brief process to create an account, Lee discovers the site to be multifunctional. Lee completes universal student records for each of his children by filling out a web-based form that solicits names, dates of birth, recent school enrollments, and so on. The survey

also asks Lee to identify areas of particular interest for each child, as well as areas where additional school supports might help each child.

In Lee's new community, a variety of school options are available. Using the Student Data Backpack, Lee can access information provided by other organizations, such as GreatSchools.net. As a result, Lee finds a local charter school with an emphasis on science and mathematics that was recently listed among the best in the state. In addition, Lee uses the site's discussion groups and feedback forums to learn that the principal of another school is herself a parent of a special education student and is a strong advocate for her students.

Lee designates these schools for his children and requests that the universal student records be transferred to them. After a few minutes, the Student Data Backpack site acknowledges that the schools have accepted the records and confirmed registration.

Lee continues to explore the Student Data Backpack site and discovers that it provides a wealth of features. There is a College Readiness Tracker that contains a year-by-year checklist of steps that parents and students can take to be informed, prepared and ultimately successful in gaining college admission. Lee is surprised to see that even though college is several years away for his older child, there are many things that can and should be completed now. The Student Data Backpack site provides links to the Free Application for Federal Student Aid (FAFSA) website, which Lee can visit to get familiar with the process of applying for student financial aid for college. There is also a variety of student achievement tools that will help Lee understand how well his children are doing academically in relation to other groups of students — by comparing each child to others in the same school, same district, same state, or same demographic groups.

As a busy parent and newcomer, Lee is glad to find several discussion groups to join. The Single Parenting group has many affinity groups — some by age, some by geography. In a few short weeks, Lee is regularly participating in two and receiving support as a parent and helpful suggestions for handling his employment transition. Through the web community, participants share their local knowledge. Lee is ultimately able to make contact with several people in the same area and gets connected with employment leads as well as new friends.

Lee receives an email notice whenever the Student Data Backpack has been updated. He can then log in and see announcements, report cards, achievement test results, etc. All the data are added to the universal student records, and Lee is able to use the tools and resources of the site to monitor each child's progress. Because the child with science and math interests appears not to be keeping pace, Lee uses some of the Conversation Guidelines offered through the Student Data Backpack to help frame a constructive conversation with her teacher. Lee also taps into the Shared Knowledge section of the site to see what steps he might take to provide additional support and enrichment to his daughter.

Discussion

The Student Data Backpack offers a solution to a number of existing problems with student data collection and usage and does so in a way that may enhance the academic performance of America's students.

Both the technical feasibility of two-way interoperability and the benefits that users would derive are speculative at this point. To test the concept of the Student Data Backpack, it would be desirable to conduct a pilot. A number of schools would need to agree to test a small-scale version of the Student Data Backpack, importing the universal student record and delivering reports back to it on student progress. Content would be developed and offered to parents to test the kinds of resources and tools that they find useful. Real feedback from real users will reveal if the Student Data Backpack delivers valuable benefits.

The most important expected advantage of the Student Data Backpack is that it provides a value-laden means of engaging parents. It does this in two ways. First, they become the guardian of their child's USR (Who better than the parent to vouchsafe the record?). The incentives for accuracy and completeness of the contents of the universal student record are greatest—though probably not perfect—with the parents. While there may always be parents who are disinclined to serious involvement with their children's education, the most minimal requirements for using the Student Data Backpack are no more burdensome than what is required to register a child in school today, with the added advantage that it does not require a visit to a school during business hours.

For parents motivated to maximize their child's development and academic experience, the Backpack can serve as an enriched portal for the delivery of tools, chat boards and information that help them understand their child's progress, engage with other concerned parents, and locate support and services if needed. The simple fact that content and services are available through the Student Data Backpack might bestow on the offerings a level of legitimacy that could increase parents' comfort in seeking out resources they may not otherwise investigate.

The Student Data Backpack could help parents explore the current array of education options for their children. Because the Backpack is independent, it can serve as a neutral platform for information. It would be relatively easy to partner with existing resources like GreatSchools.net that offer parents information about the schools in their area (created from state education department data on student performance by grade and school). Parents would be free to interpret the information for themselves and make their choices about which schools they wanted for their children.

It's true that the appeal of the Student Data Backpack won't be universal. But parents will have to exert some effort to enroll their children in school, whether it involves a visit to a website or a visit to a school's central office. Even if the only steps a parent completes are those of registration, the parent would still gain a key portion of the Student Data Backpack functionality and benefit from its ease of use. What's more, parents would gain from the efficiency of a single entry of information, instead of the multiplicity of forms now required. The Backpack would leverage SIFA compatibilities so that multiple documents could be populated from a single set of information.

Clearly, the benefits for parents from the Student Data Backpack would rise with increased use. As parents investigate the Backpack's information about local schools, the typical information asymmetries that hinder parental choice would be reduced. Simply seeing a list of schools might promote deeper inquiry into the options available for their children.

The largest direct benefit to parents comes from the opportunities for social networking online—parents can join communities that are about *them*. The degree of affinity and identity that result from participation in virtual communities can be beneficial even if the ties are weak.[12]

The Backpack also expects to offer schools and districts faster access to complete information about the students they serve. With SIFA-compliant

interfaces, the Backpack becomes an efficient and prompt way for students entering a district or school for the first time to put their essential data in the hands of educators. Moves, transfers or corrections are updated electronically with less involvement from school personnel. Further, the independence of the Backpack from all vendors of student information systems would prompt all vendors to hasten their compliance with SIFA and EDEN standards.

Once a critical mass is achieved, schools could rely on the Student Data Backpack to provide a superior alternative to their registration practices. As clean, current and accurate data become available, one could expect schools to rapidly migrate to that model of acquisition especially if it affords the opportunity to reduce expenses by streamlining personnel.

Perhaps the greatest benefit of the Student Data Backpack lies in the improved access to data to inform decisions about education improvements. Cleaner and more complete information can alert educators to the impact of their efforts; teachers and school leaders can target instruction more effectively; and researchers can make greater contributions to help schools, teachers and students become more successful.

Endnotes

1 See Chrys Dougherty's chapter, "Getting FERPA Right: Ecouraging Data Use While Protecting Student Privacy" in this volume.

2 Ligon, Glynn. "The Data Quality Imperative" ESP Solutions Group, 2007.

3 CREDO Data Quality Report Cards, Spring 2006.

4 Gulati, Ranjay. "Silo Busting: How to Execute on the Promise of Customer Focus." *Harvard Business Review,* May 2007.

5 CALPADS and CALTIDES are the student and teacher databases respectively in California. They are both under development–a prime chance to create a formal nexus between the two collections–and have only a convoluted single path to link the two together.

6 Collins, Laurie and Larry Fruth, et al. "The Right Data to the Right People at the Right Time." *Data Quality Campaign Quarterly Issue Brief,* June 2007.

7 One successful example is the Information Services Group (http://www.informationsg.com), which amasses information from a variety of commercial collection points to synthesize industry and sector reports for businesses, investors and regulatory agencies.

8 *Annual Report.* Schools Interoperability Framework Association, 2007. http://www.sifinfo.org.

9 See WebMD's most recent financial reports to see how successful the web-based education and patient support: http://investor.shareholder.com/wbmd/releasedetail.cfm?ReleaseID=326466.

10 As of July 2008, MySpace claimed 73 Million users and Facebook reported 37 Million users. This user base is predominantly in the 18-24 year old group; they are the user group that would use the Student Data Backpack in the next ten years.

11 See http://investor.shareholder.com/wbmd/releasedetail.cfm?ReleaseID=326466 for information on usage differences by gender.

12 Granovetter, Mark. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91, November 1985.