

**From Data to Information:
Boosting Access and Usefulness of Both at U.Va.**

A collection of articles by Chip German, Deb Mills, Ed Tarlton, Scott Newman and Mike Jewell

Mr. German is the director of Planning and Policy Development for the Office of Information Technologies at the University of Virginia and former acting director for the Applications and Data Services (ADS) division of ITC; Ms. Mills and Messrs. Tarlton, Newman and Jewell are current members of the ADS division at the University of Virginia.

Overview

by Chip German and Deb Mills

At the University of Virginia, we are beginning to understand how wider access to University "business" information helps people who work here make better contributions to the institution's mission. But we are struggling to understand how that concept applies to a possible distinction between the "raw" qualities of "data" as contrasted to the "meaningfulness" of "information"—by which we mean data that has been placed in a context and thereby transformed to information.

For the purpose of this discussion, we'll use the term "data" to apply to the raw stuff of measurement of University business. In our computerized systems, raw data is organized by type at the most basic level. We give these data-types names, collectively known as "data elements," that allow specialists to understand how to record, refer to and retrieve them. But until context is clear—how those elements are defined and how they relate to each other—they do not represent "information," the next major step up in the progression of "meaning." Non-specialists can't do much with "raw data," as defined in this scheme, but they can do more with "information" because it is inherently more "meaningful."

Naturally enough, as the University has explored the concept of widening access to either data or information, some of the subtleties of this distinction have gotten lost or confused. People tend to focus best on this whole line of thinking when dealing with specific examples of data and information, rather than by discussing these notions more abstractly. In any case, the University's process began in earnest with the development of a policy aimed at making "data" more broadly available. [Consistent with the most common usage, we are referring to data as a collective but singular noun. Professionals in this field often use data in its more traditional form as a plural noun.]

Access to Data

The President's Cabinet approved the Administrative Data Access Policy (ADAP) on June 24, 1994. Since the policy was approved, the University has made significant progress towards its implementation. More than anything else, this policy is aimed at bringing a cultural change: the University is moving from an environment in which specific

administrative areas functioned as "owners" of University data to an environment where such data is "owned" by the University and viewed as a vital shared asset. The concept that the University owns the data is now widely accepted throughout the institution, largely due to discussions involving the University's "data stewards" (the central administrative officers whose units are responsible for producing and recording the raw data) during the development of the policy and its early stages of implementation.

At the same time, accessibility of data has dramatically improved in a technical sense due to the development of a special mechanism of computerized access that we named the "information" warehouse (see related article on the warehouse). As we have seen this mechanism evolve, some of us wonder if it would have been wiser to stick with the common name for this mechanism—"data" warehouse—because what is in it is often much closer to data elements than to information. Make no mistake—with the warehouse, data is much easier to get to and to understand. Larger numbers of people within the University community have access to more data than at any time in the past, and many are using it in exactly the creative fashion that we had hoped. However, access to data remains somewhat constrained by the more common traditional uses and perspectives of it.

For example, some of the data that our business-function users have most energetically advocated to be included in the warehouse is data that has legal restrictions on access, such as data about people (both students and employees). This is because people scattered across the institution have legitimate reasons to get to this kind of data, but old systems made it hard to do so. The warehouse makes it much easier for authorized people to get to—and make creative use of—legally restricted data even within the constraints of such restrictions. So, access to such data remains carefully controlled, and successful early projects have created a small corps of advocates for the warehouse approach to data access from a particular population that has previously been the most cautious about it. But, restricted access to these varieties of data has limited the range of people who could take advantage of the benefits of the warehouse mechanism.

Over time, access will improve as more data is made available through the warehouse without restrictions. However, as a result of current prioritization of the kinds of data to make available and the training necessary to understand it and to manipulate it effectively, access is not as broad as we might have hoped, nor is it as easy from a procedural sense to access the data as we expect it to be in coming years.

The continual increase in the amount of University data available through the warehouse will help address another problem as well. In a complex data environment such as ours, shadow systems often spontaneously appear over time as staff in departments try to make sense of data that is supplied to them in ways that are not immediately useful—ways in which the data is not easy to manipulate for "local" purposes. By providing data in flexible, easy-to-manipulate forms, staff don't have to develop such shadow systems to reshape the data into useful form—simple tools will do the job.

Still, in dealing with the thousands of data elements that exist in our environment, there is plenty of room for confusion. Sometimes a data element from one system carries the same name as a data element from another system, but few (if any) staff members could immediately tell you the difference between them. That difference is often quite important—similarly named data elements often involve data derived from very different sources for different purposes. This capacity for confusion means that we need another step before we reach our goals in the accessibility of data. We will realize those goals only through a redefinition of the data environment here. That step, in turn, will be a natural part of the re-conceiving of business processes that will have to be

done during the replacement of the University's core-business applications in the near future.

To move the University to an integrated system of administrative applications requires further evolution of the "data culture," among other important factors. The essential business-process redesign that we both desire and need for our integrated-system approach clearly will also lead to complete rethinking of University data, from how the elements are defined, to where they are stored, to how they are accessed. The Administrative Data Access Policy and related activities have laid groundwork explicitly and effectively for the attitudes that the University needs to shape this kind of rethinking. Data stewards are gaining confidence in the more open institutional-data environment, even as it grows increasingly open with each addition of data to the warehouse. The stewards are themselves already endorsing—and even inventing—revolutionary changes that would have seemed inconceivable to them several years ago.

These revolutionary changes will include thinking about how we organize data at a very fundamental level. The notion of a so-called enterprise data model (see related story about "information architecture") and how that model plays out in many ways affecting access to individual data elements will be a central part of the planning for an integrated system of business applications at the University.

In addition to the fundamental obstacles of the current data environment, there are technical challenges to easy access as well. The sources of data, such as strip files and VSAM files, make it difficult to selectively make portions of the data generally available and other portions not available. Again, the warehouse mechanism provides at least a partial answer. With the warehouse, it is easy (compared to the pre-warehouse situation) to establish views of data. Financial accounting system (FAS) data in the warehouse has two views, and everyone receives them. Sets of data from the Office of Sponsored Programs (OSP) have legally restricted social security numbers and, therefore, there are two views—one view with the number and one view without.

Sets of student data have all generally available data in one view and the legally restricted information in another view; however, access to this data is currently limited by school type. For legally restricted data, all undergraduate schools have access to all other undergraduate schools' data. In the past each graduate and professional school had access to only that school's data, but with student data on the warehouse, all graduate and professional schools have access to each other's data.

Human resources data sets have significant amounts of data that are legally restricted, but almost all unrestricted data (with the exception of salary-related fields) is available for general consumption. Legally restricted data is available to persons with authorizations according to their vice-president/major budget unit (MBU) or division/administration codes.

The warehouse mechanism has forced us to confront issues of particular sensitivity regarding specific sets of data, and in all cases, data has become much more open. A good example of the cultural change in which we are engaged came with the passage by the 1996 Virginia General Assembly of a requirement for agencies to provide public indexes of all of their databases, beginning with databases created or substantially changed after July 1, 1997. Opposition to this requirement quickly became evident at the University, but not on the same basis as it would have been a few years ago. Then, the data "owners" would have fought bitterly to preserve their right to determine who had access even to summary information about their data and to determine how it would be used.

In 1997, however, the University's data stewards met with Forrest Landon (see a Landon story), executive director of the Virginia Coalition for Open Government, to discuss the University's reasons for concern—that the legislation was written in such a way as to impose a heavy and unnecessary administrative burden on institutions such as ours. At the meeting, the data stewards unanimously expressed their support for the notion of openness and freely accessible public data. The Virginia Coalition was impressed by the University's enlightened view of public access to institutional data. It formed a firm partnership with the University on this issue, helped ensure that the Virginia Press Association and the State Library of Virginia joined that partnership and successfully campaigned for guidelines that reduce the burdens of compliance.

We've seen this pattern of changing attitudes toward access to data in other ways as well. School officials who were concerned that individual departments in the school would be able to view data about other departments—data that previously had been considered confidential and not shared among departments—have told us that department chairs had discussed the matter and concluded a more open environment made good sense.

This is not to say that there are not pockets of concern and even resistance to some of the fundamental notions of open access to institutional data. The most common concerns center on internal control and security. To address those concerns, we have asked the University's audit department to help us review our increasingly open environment. Early discussions suggest that the internal control and security currently in place in the policy and its implementations are adequate. However, audit will continue to review the environment, data element by data element, to ensure that—when necessary—the University makes data available with appropriate restrictions or with required training so that it is used properly.

Appendix A depicts of the evolution of access to institutional data as we see it at the University. We have made good progress toward the goal established by the University's leaders—a much more open environment in which employees have the opportunity to make more valuable contributions because they can work from a broader base of data. We are entering a critical phase of business-process analysis on the way to replacement of administrative applications. The success of a deeply rooted redesign of how the University executes its "business" work will directly affect the ultimate success of our efforts to open up the data environment. At the same time, the University's initial success with the Administrative Data Access Policy has helped pave the way for that redesign.

Easier Data Input

Simultaneous with our work to open access to institutional data, we are also working to make the mechanisms for input of data into University systems more efficient and sensible, streamlining processes wherever possible.

For example, throughout the University, department staff use simple paper forms that they designed to gather expense reports from faculty and staff who have traveled on University business to justify reimbursements. Those departmental forms are aimed at simplifying the process of gathering expense details from travelers who would find the University's normal forms obtuse. But this means that data goes through many stages of recording—first by hand by the travelers on the department-designed expense reports, second by the department staff who then transfer the data (usually with typewriters) to the University's paper forms and finally (after many stages of review and approval) by a data-entry unit that takes expense details listed on the University form and "keys" them into the computerized financial systems. The entire process's inefficiencies add up to delays in reimbursements getting back to the travelers.

At the moment, there is no in-hand solution. State regulations require that reimbursements for travel expenses have real—not digital—signatures of approving officers, and they must be accompanied by certain original receipts. These state procedural requirements represent technical impediments that stand in the way of straightforward digital solutions for the travel-reimbursement system. But, the attention to the problem has brought progress on related fronts. It is part of a University effort aimed at figuring out how to do such things better, one component of which has been the Electronic Forms Project (see related story). The project has completed work on several simpler forms and has succeeded in creating a wider expectation in the University community that a range of transactions, including the process of entering data into central University systems, will be done soon by digital means directly by faculty and staff in their offices and by students from computer labs or their residences.

Another project is aiming at a more immediate benefit (even if ultimately paper-based) to persons who use forms to put data into the University systems. The U.Va. Forms Directory project will make about 120 administrative forms available on the web in Portable Document Format (PDF) with fields that can be filled in on-line. A searchable directory will provide access to these PDF forms, as well as forms that reside on other web sites.

These projects may also be working in a subtle way to better inform the community about the data environment. The more faculty and staff understand the correspondence between what they are entering—now directly in some cases—into the system and what they can retrieve from it, the less they will need additional training and explanation to turn data to which they have access into practically useful information.

Data to Information

With a range of first-phase successes on the data front, the University now finds itself in a dilemma. We have found ways to make institutional data much more broadly available to the University community, but we have collectively discovered something the data stewards knew intuitively all along—the way data exists at the University means that people who have access to it must have significant explanation and/or training in order to make sense of it. In short, before data becomes useful information, someone has to contextualize it, either by training (which represents one hurdle to open access) or by explanation in situ (which in essence represents another type of hurdle). In the latter case, the explanation may be so cumbersome and obtuse that no one bothers to read it or, worse, bothers to access the data at all.

The solution, which—when you think about it—we've been using for many years, is to allow persons whose skills are in clarity of explanation (writing or teaching specialists) to work with data specialists in creating presentations that couple the data with clear and useful explanation. But because the explanation adds great bulk to the product of such a presentation, presenters usually find that they either have to sacrifice detailed data or have to reduce the range of types of data in order to get the important points across in the available time or space. A first-rate example of successful application of this approach is the University's annual "President's Report." The report communicates many points of University data, highly contextualized (it has become information, by our definition), to an audience of persons who don't specialize in working with those types of data.

However, the limitations of that solution are just as evident when you think about applying it for another purpose to a specific audience. The President's Report may be of interest to the University's faculty and staff, but it doesn't (as a rule) empower them to

make better contributions to the institution by giving them data in sufficient detail for them to creatively manipulate it for new purposes. There is clearly a middleground, somewhere between the President's Report and the warehouse mechanism, that will provide enough contextualization to promote both greater access and greater utility of the data, as well as detailed data.

Again, an example of such a middleground is helpful—the University's Data Digest, produced by the Office of Institutional Assessment and Studies. Here you can view relatively detailed data with enough explanation of context to ensure appropriate understanding. The Data Digest is a model that seems worthy of emulation by departments dealing with slightly more specialized information.

Another concept that seems to hold significant promise is a particular type of "data mart." Data marts reduce the range of data to a more manageable span. For example, a data mart could provide a subset of data that reduces some confusion about data elements simply by making available only the one that users most commonly need. The data mart itself does not necessarily make the data easier to manipulate—the same query tools that work on the data warehouse may be required for the data mart—but at least the context and appropriateness of the data may be clearer.

Expanding the concept of the Data Digest, combined with formatting that allowed such "views" to be easily downloadable as popular spreadsheet files, and increased use of data marts may be the next logical steps in the evolution of the University data/information environment. Providing information to our employees is empowering to them; providing raw data without adequate context may prove, more often than not, simply confusing.

APPENDIX A: Evolution of Data Access Environment			
University of Virginia			
	Former	Current	Future
Access	Mainframe systems had difficult security and access procedures. Employee access very limited.	Information Warehouse access procedures and privilege-granting are easier, but both the training required to help people know how to access general administrative data and limited range of views that contain only such data represent continuing barriers.	Integrated-systems approach to administrative applications will reduce training barriers and remove some structural impediments to providing a wide range of freely accessible views, at least for general administrative data.
Usage	Often limited to snapshots of data -- usually not downloadable for local manipulation; had to be re-keyed for that purpose.	People can download data and manipulate it for local purposes. Much more data is available for comparative purposes (across previous organizational boundaries) than was true in former conditions.	Re-conception of business processes will lead to data structures that lend themselves to meaningful views that many employees can understand and use appropriately.
Culture	Low acceptance of the notion of free access. High expectation of misunderstanding of data.	High acceptance of the notion of free access. Medium expectation of misunderstanding of data, tied to the complexity of the current data structures and definitions.	High acceptance of the notion of free access. Low expectation of misunderstanding of data.
Potential for misuse	Low; controlled by limits on access (prior restraint).	Potential for misuse is higher, but community agrees it will be handled by policy enforcement.	Same as current.

virginia.edu, Volume II, Number 1, Spring 1998

*Published by the Office of Information Technologies (OIT) and the
Department of Information Technology and Communication (ITC)
at the University of Virginia
Copyright 1998 The Rector and Visitors of the University of Virginia*