Developing a Voluntary Institutional On-Line Information Network: The VIOLIN Pilot Project

Revised May, 2001

John H. Milam, Jr.
University of Virginia

Executive Summary

This report documents the VIOLIN pilot project conducted at the University of Virginia from Summer, 1999 through Spring, 2001. Funded under an AIR/NCES/NSF Research Grant, the purpose of this project was to develop and demonstrate a new type of voluntary, web-based data collection and dissemination tool. The project was coordinated with and incorporates features of the new IPEDS web data collections and Peer Analysis System. It is recognized that IPEDS cannot be continually redesigned and expanded and that a collection effort such as this could help better address emerging policy questions.

The Statement of Work for the grant included seven broad activities: (1) design a prototype of the voluntary core dataset (VCD/VIOLIN); (2) develop a web tool for its collection; (3) conduct a pilot test of the collection effort; (4) develop a prototype peer analysis tool using the newly collected data; (5) demonstrate the tool and resulting data set to key stakeholders and the IR community; (6) evaluate and document this project; and (7) prepare a request for proposal for fully implementing the design.

Twenty-four institutions participated in the pilot test, representing a variety of combinations of Carnegie classification and control. An online survey was used to obtain feedback about the project and to develop consensus about what data would be collected. Participants raised issues of concern about FERPA, public access, editing procedures, and technology. It was agreed that the Common Data Set represented the best type of data which could be shared in a voluntary collection such as VIOLIN. There was a 100% return rate for collecting the CDS from pilot study schools. Problems in collecting data from existing surveys such as the NACUCO Endowment Survey were addressed, particularly intellectual property rights and definitions. Other possible data collection efforts for VIOLIN, including a new proposal from NCES for consolidating redundant data collected by federal agencies, are presented.

Per NCES requirements, the VIOLIN prototype was designed with Microsoft software and web database technology, including Active Server Pages and SQL Server. Project websites were developed and maintained at UVa's Curry School of Education and at Willamette University. Dr. John Milam of UVa served as Principal Investigator and Project Director and Tod Massa, formerly at Willamette and now at the State Council of Higher Education for Virginia, served as Project Consultant. Numerous presentations have been made by the researchers at state, regional, and national conferences as part of the pilot study and to disseminate information about the project.

Overall, the pilot study found that VIOLIN is a viable and workable solution for collecting data that are either not collected elsewhere or for which the data are not currently available in electronic format. Any implementation of VIOLIN must balance the utility and interest in the data against increased reporting burden. It is critical, though, that VIOLIN remain voluntary and not be perceived as a "test bed" for potential IPEDS questions. Due to a cut in funding from NCES, it is proposed that VIOLIN be implemented privately and that the CDS be collected as the primary VIOLIN activity.

Introduction

Never before has so much change been taking place at once in terms of how higher education data are collected, analyzed, and disseminated. This is an exciting time for institutional researchers, policy analysts, legislators, accrediting agencies, professional associations, higher education administrators, and federal and state agency staff. There is much at stake in decisions to revise existing datasets and to build new web-based data collection and dissemination tools.

In examining these developments, there are at least three coordinated efforts. These include: (1) National Center for Education Statistics (NCES) work to redesign the IPEDS survey instruments, along with the implementation of web data collections and the dissemination of online data through the Peer Analysis System and IPEDS COOL (College Opportunities On-Line); (2) the NCES, National Science Foundation (NSF), and Association for Institutional Research (AIR) grant program for Improving Institutional Research in Postsecondary Educational Institutions; and (3) the National Postsecondary Education Cooperative (NPEC) Council and working groups designed to generate a national dialogue about a variety of data issues such as accessing survey resources, enhancing the quality and use of student outcomes data, best practices for data collectors and data providers, and technology and its ramifications for data systems.

It is in the spirit of these discussions and projects that a proposal emerged that a new kind of "Voluntary Core Dataset" be developed as a central data collection effort to address unanswered policy questions. This would meet the needs of a variety of users for institution-level analysis and would hopefully alleviate the need for institutions to respond to an ever-expanding number of surveys, replacing it with one web application geared toward peer comparisons.

Purpose

The purpose of this grant project is to develop and demonstrate a web-based collection and dissemination tool for institutional peer analysis of a new, "Voluntary Core Dataset" (VCD). This application will utilize IPEDS peer analysis tools and hopefully provide new data for decision-making at a variety of levels.

At the outset of the pilot, the project was dubbed by Roz Korb of NCES as "VIOLIN," with the acronym "Voluntary Institutional On-Line Information Network." The VIOLIN pilot project was housed at the University of Virginia, where Dr. John Milam, Research Associate Professor in the Curry School of Education, served as Principal Investigator and Project Director. Mr. Tod Massa, Director of Institutional Research and Planning Support at Willamette University at the time, served as Project Consultant. Mr. Massa is now Director of Institutional Research and Director of Technology Services at the State Council of Higher Education for Virginia.

Project Description

Objectives

The overall objective of this project is to develop and demonstrate a pilot version of the VIOLIN web-based data collection and dissemination tool. The Statement of Work for the grant included seven broad activities:

(1) design a prototype of the voluntary core dataset (VCD/VIOLIN);
(2) develop a web tool for its collection;

(3) conduct a pilot test of the collection effort;
(4) develop a prototype peer analysis tool using the newly collected data;
(5) demonstrate the tool and resulting data set to key stakeholders and the IR community;
(6) evaluate and document this project;
(7) prepare a request for proposal (RFP) for fully implementing the design.

Additional Requirements

In the process of ensuing discussions with NCES staff, who funded the pilot project, additional requirements for the VIOLIN pilot project were described:

(1). The resulting web applications must be built with NT Server, SQL Server 7 database, and Active Server Page (ASP) technologies;

(2). While the VIOLIN project web server might be housed at an outside location such as the University of Virginia, the web application must be able to interact with the NCES web server, in particular its utilization of the IPEDS data in the Peer Analysis System (PAS). Project staff were expected to work closely with NCES in understanding the PAS.

(3). It was also important to incorporate all of the lessons learned by NCES in the process of building its new IPEDS web collections, and include such features as edit checks, uploading of data files for submission, and record locks. Therefore, project staff were expected to work closely as part of the IPEDS Redesign Team and the IPEDS Training programs. Any VIOLIN collection will need to incorporate and be based upon currently available IPEDS data elements.

(4). The NPEC project on Accessing Survey Resources (ANSWERS) is considered essential to the identification of potential data elements for inclusion in VIOLIN. Project staff were expected to work closely with the NPEC ANSWERS project. Dr. Milam was subsequently retained as Consultant for this NPEC Working Group, which held its first meeting in July, 1999.

Designing a VIOLIN prototype

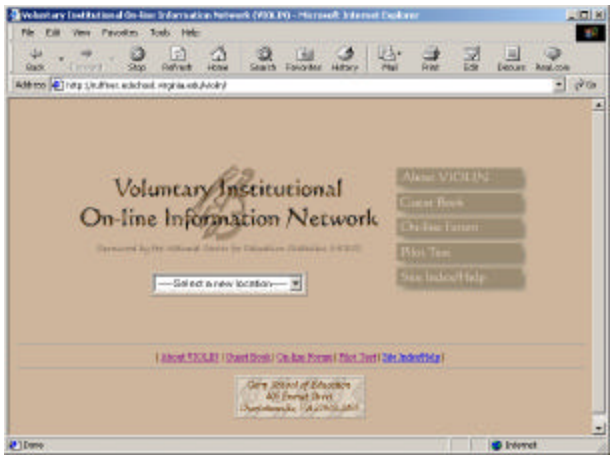
As part of the design of a prototype for VIOLIN, a set of 24 institutions was identified to serve in a pilot test group. These institutions were selected to represent a variety of combinations of Carnegie Classification and control. The following institutions participated in the pilot test.

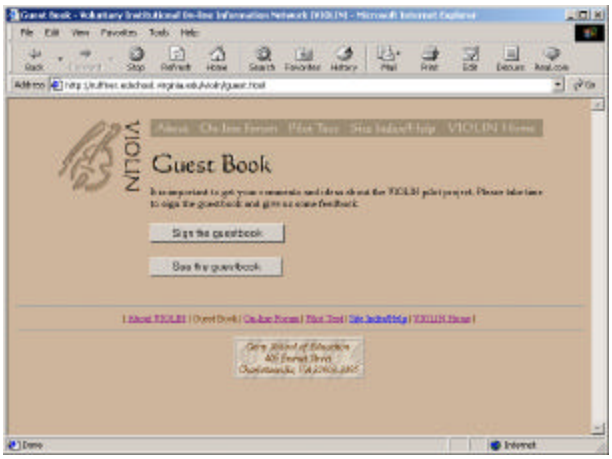
Table 1: List of VIOLIN Pilot Test Institutions

School	Type	Control	Contact
Allegheny College	LA1	Private	Marian Sherwood
Bowling Green State University	D1	Public	Bill Knight
Catholic University	D1	Private	Tracy Hunt-White
Cerritos College	2YR	Public	Steve Wong
Christopher Newport University	LA2	Public	Evan Davies
DePaul University	D2	Private	Gerry McLaughlin
Eastern Carolina University	C1	Public	Erika Ramey
George Mason University	D2	Public	Kathleen Cheney/Mike Wood

School	Type	Control	Contact
Georgetown University	R1	Private	Mike McGuire
Highline Community College	2YR	Public	Laura Saunders
Indiana University-Purdue University Indianapolis	D2	Public	Vic Borden
International College	LA2	Private	Fred Cenedella
Johnson County Community College	2YR	Public	Jeffrey A. Seybert
Manor Junior College	2YR	Private	Sally Mydlowec
Oakland University	C1	Public	Laura A. Schartman
Randolph-Macon Woman's College	LA1	Private	Dixie N. Sakolosky
Reed College	LA1	Private	Jon Rivenburg
Seattle Pacific University	C2	Private	Jerry Finch
University of Kansas	R1	Public	Deb Teeter
University of Portland	M1	Private	Karen Nelson
University of the South	LA1	Private	Paul Wiley
University of Virginia	R1	Public	George Stovall
Virginia Tech	R1	Public	Valerie Martin Conley
Willamette University	LA1	Private	Tod Massa

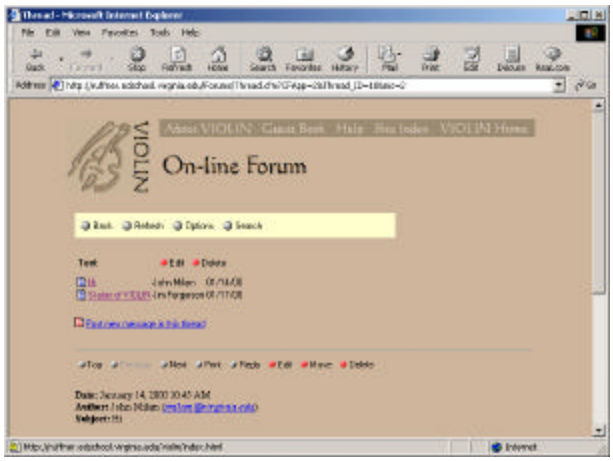
In order to brainstorm about the structure and nature of this kind of voluntary effort, an online survey was designed and administered for completion by the pilot school participants (100% return). A web server was developed for this purpose and housed at UVa's Curry School of Education. Software included NT Server, Internet Information Server, ColdFusion, Allaire Forums, and Access. The website was designed by Dr. Milam, with graphics support from Assistant Professor Dr. Lisa Heaton. The online interface and web database was developed using ColdFusion and Access. Screenshots of this website follow. The URL is available at: http://ruffner.edschool.virginia.edu/violin. A Guestbook feature was provided to solicit feedback about the project. This application allowed viewers to both submit and view comments.

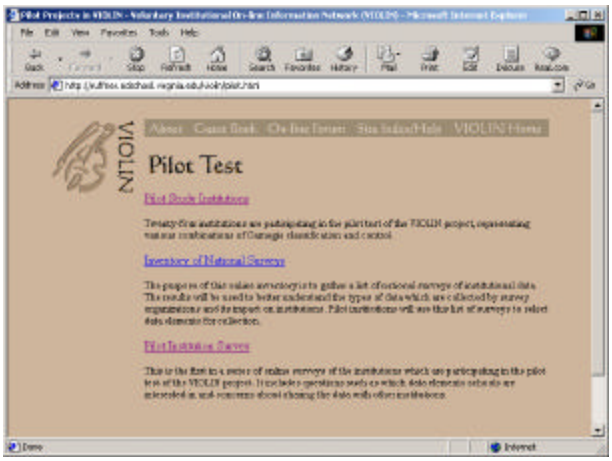




A threaded discussion group was also available as an application on the website. This was developed using Allaire Forums. For each of the pilot institutions, three pages were maintained. These in-

cluded a list of the schools, with contact information; a link to the Inventory of National Surveys application which was developed as part of the VIOLIN Project; and an online survey.





The online survey allowed participants to both submit responses and to view responses by survey item and by institution. The survey included the following questions:

(1) What data elements would you like to see collected in a new, national, voluntary exchange of data?

(2) Why are you interested in these particular variables?

(3) Are there surveys which already collect some or all of the data elements you are interested in? If so, what are they?

(4) How do you envision that these or other VIOLIN data will be collected and displayed via the World Wide Web?

(5) How would you design edit checks for these or other VIOLIN data?

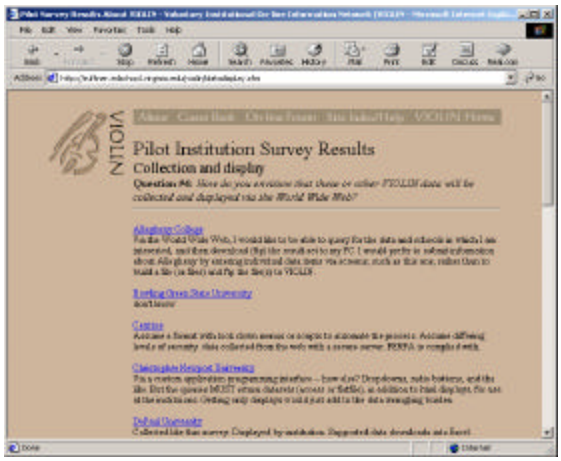
(6) Do you have concerns or issues about sharing VIOLIN data with other institutions that participate in the VIOLIN project?

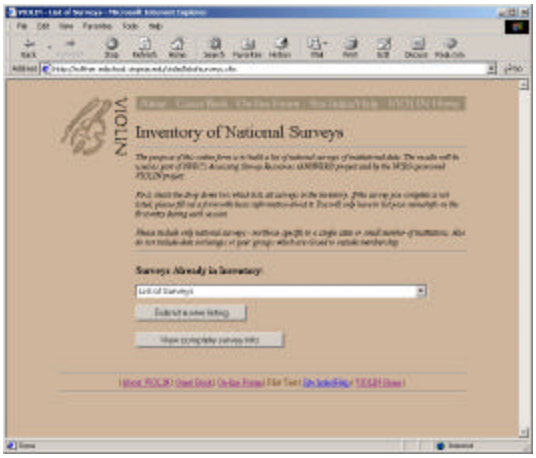
(7) Do you have concerns or issues about VIOLIN data becoming public domain after one or two years?

(8) Do you have any other ideas about how the VIOLIN pilot test should be conducted?

(9) Do you have any other comments about this project?

The following screenshot illustrates how the results were made available by survey item and by school. The second capture documents the Inventory of National Surveys, which pilot institutions were asked to complete for every survey which comes to their office.





Survey Results

The pilot institutions were interested in two broad categories of data: (1) data not found elsewhere, such as assessment survey results and financial aid information; and (2) data collected by others but not readily or easily available, such as the Common Data Set (CDS), US News, Delaware, and retention data in the IPEDS GRS. The most commonly mentioned data source of interest is the CDS. Schools are mostly interested in using the data for peer comparison. Some of the data cannot be obtained and are needed in order to run vendor models such as the US News rankings.

Some interest was also expressed in using VIOLIN to expand existing data exchanges. There is no intention to supplant existing data collection efforts which are limited to invited institutions, such as AAUDE, Urban-13, SUG, and HEDS. Rather, the additional data will help to supplement the exchanges and the web database software will help to make them more efficient and better integrated with IPEDS. As part of this project, dialogue was held with Vic Borden, who is coordinating the Urban-13, about the possibility of VIOLIN being used for the data collection and dissemination. While there is interest in the use of the software and the interface to IPEDS, some concern still exists about access restrictions to the data. Other discussions were held with Bill Knight about the Doctoral University Exchange and with John Vaughan, Deb Teeter, and George Stovall about the AAU Data Exchange.

Most pilot schools envision that VIOLIN would collect data with online screens comparable to the IPEDS web collection. There would be routine security for limiting access and approval rights. The web application would provide displays of the data in various formats, including the ability to download a dataset in Access, Excel, and AASCII format. It is expected that participants in VIOLIN would be able to upload the data as a flat AASCII text file or Excel-compatible spreadsheet via a browser upload button and/or FTP. In addition to interfacing with existing IPEDS data available through the Peer Analysis System and COOL, participants would like the application to pull in data from other sources, such as the Survey of Earned Doctorates, the NSF Graduate Student Survey, and the NSF Academic R&D Survey. Finally, pilot schools also wanted the peer analysis tool to be capable of producing graphs and tables of comparisons.

A variety of editing procedures are described by the pilot school participants as necessary for a project such as VIOLIN. These mirror many of the features built into the IPEDS web data collection. Recommended procedures include: triggers based on 10-20% variance from the past year for a data element; time-based stability; peer group comparability; allowable ranges; fit within longitudinal data; sign-

off by the data entry keyholder; the use of flags for outliers, with explanations about why the data are valid; and self-verification.

A number of concerns were raised about sharing data through VIOLIN. As expected, these came more from private institutions than public. All schools are concerned about obtaining a FERPA reading on this type of collection. Even with FERPA approval, some schools will not complete all items. This collection is voluntary and is not expected to be a test-ground for potential IPEDS questions. Therefore, pilot schools argue that the voluntary nature must allow them to complete only those items for which their institutions wish to share data. An agreement about confidentiality and disclosure is recommended, perhaps based on the AICUP statement signed by presidents. It is expected that the data will only be shared with other schools that participate.

If the project were publicly funded, the data would have to eventually be released to a larger audience in some form. According to NCES, it is possible that this larger audience may still be restricted to those who obtain an approved site license and agree not to publish data which could potentially identify individuals, such as cell sizes of 5 or fewer people. With private implementation of VIOLIN, this is no longer an issue. As part of participation, schools would recognize an agreement about the copyright issues in the data and how the data would be shared.

In planning for the implementation of VIOLIN, pilot school participants suggested a phased approach. It is recognized that the data retrieval tool is critical. Will this be limited to the Peer Analysis System for IPEDS? A detailed data element dictionary is necessary, with discussion of unique situations in reporting. There was confusion about the PAS, since it was not fully developed or stable during much of the pilot test time period.

Overall, the pilot schools were somewhat enthusiastic about the project. Some warned, though, that the balance between reducing burden by centralized collection versus additional web-based reporting is difficult to maintain. The success of this requires that reporting requirements eventually diminish, not get expanded because of this collection. Institutions see the utility of responding to a central repository, when the data are not available at all. However, when the data are available elsewhere, such as through a vendor, what is the point of duplicating the effort in submission and actually increasing burden?

For some, there is less incentive to complete a new voluntary effort. The success of IPEDS is tied to the stick of mandatory compliance, now with an accompanying $25,000 fine. Participants recommend that data not be collected if they are available elsewhere to institutions. Where data were collected elsewhere, but not available to the institutions for their use, there was interest, but only if the data elements are critical to the mission of the institution. Some confusion still exists over who will decide what data are collected and over the role and authority of a board of directors. Additional comments explored this point further. Pilot schools do not want to see VIOLIN as a test bed for potential IPEDS questions. One school explained that VIOLIN would "die on the vine" if it is perceived in this manner. VIOLIN must remain voluntary if it is going to succeed and the data elements must be fluid, dynamic, and very timely if they are going to be of interest and worth the extra reporting burden.

Additional Discussion

Over the eighteen months of the project, a significant amount of email traffic and phone calls were exchanged with participants from the 24 pilot school institutions. Dr. Milam and Mr. Massa also met with participants as part of state, regional, and national conference presentations such as VAMAP, PNAIRP, SAIR, and AIR. Consensus was developed among the schools that VIOLIN should begin by collecting the Common Dataset and each school agreed to provide a 1999 or 2000 copy of its CDS. En-

thusiastic statements were made that this would finally give something back to the institutions that provide the CDS data.

As it stands now, the CDS is developed by the publishers in conjunction with an advisory panel. The CDS consists of an agreed-upon list of questions, including many data elements that come directly from IPEDS. By providing a web-based repository for the CDS as a whole, participants believe that they would no longer need to complete these CDS or IPEDS items for the vendor surveys. It is this point which seemed to convince pilot schools that VIOLIN really does have the potential to reduce burden. Otherwise, it is simply a duplication of other already in-place collection efforts. According to CDS advisory committee member Peggye Cohen, this solution is already in place. "As a matter of fact, all of them [the publishers] are willing to take a school's CDS and their survey with only their specific questions answered (that's what we do)."

At several points in this discussion about potential data for VIOLIN, Dr. Milam met with representatives from the admissions guides publishers, including the College Board, Peterson's, and U.S. News. For a time, it seemed as if several of the vendors might be in favor of a central and neutral CDS repository. Later in this process, vendors stated that they would not be able to develop a shared vendor-sponsored repository. Rather, they are developing separate and duplicate CDS collections as part of their moves to web-enable their individual collection efforts. It is hoped that with broad support for the private implementation of VIOLIN for collecting the CDS, the publishers will work cooperatively with AIR and other associations to develop a central collection effort that will reduce burden, meet their data needs, and meet institutional data needs for peer comparisons as well.

Discussion with members of various IPEDS Redesign Teams suggests that there are different perceptions about whether institutions are willing to submit CDS data. All of the pilot schools were willing to share their CDS data. The researchers have not heard from any school at any of the presentations about VIOLIN that would be unwilling to share at least part of its CDS data, given expected assurances about FERPA, reciprocity, and confidentiality.

The clear consensus reached by this project is that a voluntary data collection effort such as VIOLIN would succeed only if it is able to collect timely and useful data that are based on established and clear definitions. Other data such as financial aid awards and assessment survey data such as the Student Engagement Survey would be too localized to sets of institutions to have broad appeal. The CDS is a rallying cry for institutions that wish to benefit from the massive increase in reporting burden, getting "something back" for all of their unpaid effort in reporting. Some pilot schools question why it is only the four publishers who are able to get access to the data which result from the Common Data Set. Note: the College Board does sell their database, which includes the CDS, and U.S. News releases some CDS data elements on its website. It is currently not possible to buy just the CDS from any of the four vendors.

Link to IPEDS Data

One outcome of the project is that any future implementation of VIOLIN will have much more open access to the raw, SQL Server IPEDS data which are at the heart of the Peer Analysis System. With the move within NCES to three IPEDS web collections (Fall, Winter, and Spring), this ensures that the online CDS will contain the most current IPEDS data available. For this reason, there will never again be a need to ask for IPEDS items on the CDS, further reducing the burden on institutions and better using the existing NCES datasets.

Relationship to the NPEC ANSWERS Project

At the outset, an important goal of the pilot project was that any effort in voluntary data collection must interface with the newly developed work of the NPEC ANSWERS project (which stands for "Accessing National Surveys with Electronic Research Sources"). It was articulated by NCES staff and others that the ANSWERS project would be used to identify areas or elements of data which are not currently collected and which could impact important public policy discussions. The ideal is that after potential data are identified by ANSWERS, recommendations would be made to the VIOLIN board of directors, which would then consider adding these needed data.

Throughout the VIOLIN pilot project, Dr. Milam has helped to tie a vision of VIOLIN to that of ANSWERS. Presentations about one project always mentioned the symbiotic nature of the other. There has been a great deal of enthusiasm for this role within ANSWERS, although the same type of questions were raised by pilot school participants and others. The most major concern was whether ANSWERS would result in VIOLIN simply becoming a "test bed" for potential, mandatory items in IPEDS, which would defeat the purpose of reducing burden. A second concern is that many surveys will have to be entered into ANSWERS online database and tools before it will be possible to determine whether new data elements of interest are currently not being collected.

In addition, pilot school institutions were asked to participate in one of the critical features of ANSWERS, the Inventory of National Surveys. This application is designed to collect in one place information about all of the many surveys which are completed at the institution level. Any and all surveys may be submitted. Response and design issues raised as part of the pilot test helped improve this application when it was incorporated into the final ANSWERS website.

New Developments

Since the status of the VIOLIN project was last reported at the SAIR Annual Meeting in October, 2000, a number of other developments related to VIOLIN have taken place. Most notable is discussion by NCES staff of whether VIOLIN might be expanded into an entirely different project and serve an even greater purpose. There is a great amount of duplication among federal agencies in their collection of higher education data at the institution level. This results in confusion in meeting policy needs and waste of precious federal and institutional resources. With the advent of the web, perhaps the concept of VIOLIN could be expanded to encompass all federal data collection at the institution level which is duplicated by more than one agency. This is something which has been discussed as an idea for some time, though a solution was not workable until the advent of web data collection. According to Dennis Carroll, this project has been a long time coming and will take a long time to develop fully.

Since the Office of Management and Budget (OMB) must be involved in approval of all official collection efforts, OMB could potentially recognize VIOLIN as the sole source for data which have in the past been duplicated across agencies or collected at different census dates. Therefore data elements such as Carnegie classification or region would be collected only once, with timely extracts provided to agencies so that they could focus on those questions that are of greatest interest to their unique missions. With this expansion, a new version of the VIOLIN concept has been discussed - ORCHESTRA, which would better suit the much greater role in reducing burden and improving the comparability and integrity of data.

Significantly, such an effort could not be undertaken unless the duplication of data variables across different federal agency datasets is well documented. The ANSWERS project has been discussed as the type of research effort needed to identify these variables, along with the appropriate use of definitions, census dates, etc. Such an "orchestrated" effort would go a long way toward promoting better data collection and dissemination practices.

Decoupling from NCES

Finally, NCES announced in December, 2000 that it would not continue to fund the VIOLIN project past the pilot stage and that the ORCHESTRA project would be conducted separately, beginning as a Working Group project within NPEC. The reactions of the institutional research and association communities were mixed. While there was some real momentum behind working with NCES to promote VIOIN, there is also great relief – for this meant that VIOLIN could be collected privately. Private, non-governmental collection of VIOLIN means that the data are not subject to FOIA and being made publicly released because federal monies were used for the project. Carol Fuller of the National Association of Independent Colleges and Universities and other representatives from private college and universities such as Mary Sapp are much more interested in supporting VIOLIN, now that it will not be funded by NCES. In the words of Dennis Carroll, "I am doing you a favor by not funding it."

Potential Collection Efforts

If VIOLIN is going to succeed, institutions must be able to quickly see "what's in it for them." For some schools, this is relatively easy, since they want access to the CDS. The following are examples of other potential VIOLIN data collections.

(1). Some schools need comparable financial aid or assessment data and are desperate to encourage other schools to participate in VIOLIN to share data for peer comparisons.

(2). For community colleges there are very little data about non-credit course activity, something critical to their mission. It is possible that AACC may be willing to work on a collaborate effort to collect these data from their institutions.

(3). There is great interest is faculty data by discipline, for which recent studies by one of the authors document that there are no reliable estimates of the total faculty population by discipline (Milam, 1999). Many graduate programs working with the Council of Graduate Schools are concerned about issues of faculty supply and demand.

(4). Many schools which complete data for U.S. News would like to get back all of the data elements which are collected in order to calculate the formulas themselves and better understand the rankings process.

(5). A recent study by NCHEMS for the Council of Higher Education Accrediting agencies CHEA documented the overlap and differences between data collections from member agencies. It is possible that VIOLIN could be used to collect consistent data in one place that could be shared across agencies and organizations, reducing burden and improving data integrity.

(6). The Urban –13 Data Exchange has discussed using VIOLIN as its software for data collection. This would ensure integration with the latest IPEDS data and allow member schools to focus more on the analysis of the data.

Vendor Collection Issues

Another important issue raised in interviews by the researchers is that of ownership. Should VIOLIN collect data from instruments that are owned or administered by non-federal agencies? An example of this was discussed in a recent AIR Currents report about VIOLIN and the IPEDS redesign. It

was recommended that VIOLIN could collect data from surveys such as Rand's Voluntary Support of Education, the NACUBO Endowment Survey, and the Delaware Study.

In some cases, the surveying organization collects the data for its own publications or analyses, but does not make the data available in electronic format to participants for peer analyses. Each individual school assumes that it "owns" its own data, but are the collection method and data element dictionary protected by copyright or intellectual property law? The Higher Education Data Sharing consortium (HEDS) encountered threats of potential litigation over this issue recently, when it attempted to collect institutional responses to non-federally administered surveys.

There has been interest shown by pilot schools in collecting assessment data from instruments administered on campus. Must VIOLIN purchase the data from the vendor, or is it possible to obtain copies of institutional data directly from the schools? The answers to these questions are not clear and there are a number of copyright and intellectual property issues which will need to be resolved before this project can move forward.

Should only instruments that are not copyrighted or conducted by a commercial vendor be included in VIOLIN? This raises problems about whether such collections have been adequately vetted, with full documentation of census dates, definitions, and extract issues. One objective of the related ANSWERS project is to promote effective use of standard NCES, NSF, and CDS definitions. Any new data collection such as non-credit course activity would need to be carefully designed and this may fall well beyond the scope of the VIOLIN project.

If VIOLIN collects data that are associated with an instrument, such as the CDS, then its collection and delivery of the results must exceed the expectations for existing collection and delivery methods. If VIOLIN is to collect the CDS, it must be able to document to the four admissions guide publishers who work with the advisory group to develop the CDS that they will save money, get better data, have a better web collection process, have a better response rate, and help to reduce the reporting burden on institutions. All of these objectives are very possible with VIOLIN, which will incorporate state-of-the-art technology in web software, edit routines to ensure data integrity, and peer analysis tools. It is possible that such a move could generate significant good will from institutions toward the publishers, perhaps lessening some of the perceived animosity over rankings.

Additional discussions held by AIR-related groups with the vendor community have suggested that Peterson's, the College Board, and U.S. News need to understand much more about the proposed VIOLIN project before they can support using it to collect the CDS. For this reason, an initial discussion will be held with the publishers as part of the Higher Education Data Policy Committee (HEDPC) meeting at the 2001 AIR Forum in Long Beach. John Milam and Tod Massa will make a presentation about the VIOLIN pilot project at the Forum. There will be a special HEDPC open meeting to discuss data policy issues, including VIOLIN. This will also be a topic of the AIR Board meeting at the close of the Forum.

Developing a web tool for its collection

In order to meet the technical requirements of NCES for developing web data collection tools, it was necessary for both John Milam and Tod Massa to develop advanced skills in SQL Server 7 and in Active Server Pages. As part of this learning process, Mr. Massa worked in a volunteer capacity for AIR to develop the AIR Forum Scheduler software and other applications in ASP. He also moved them from Access to SQL Server. For this and other contributions, he was recognized with the "Outstanding Service Award" at the AIR Annual Forum in June 2000.

Dr. Milam went through a similar evolution, working as a volunteer to develop several web applications for AIR with ASP. He also moved the ANSWERS website and online tools from ColdFusion and Access to ASP and SQL Server. Non-VIOLIN funding was used to purchase copies of NT Server and SQL Server 7 licenses and software for other projects utilizing the same web server and desktop facilities.

This new effort in data collection could not be created in a vacuum. Rather, it had to be built upon the lessons and efforts of the IPEDS web data collections. For this reason, extensive discussions were held with NCES staff about the evolution and development of the IPSFA collection in August 1999 and the IPEDS Fall and Spring collections in 2000-01. During the months of this VIOLIN pilot project, NCES was learning about and addressing issues about uploading files, user support, keyholder locks, adequate SQL Server database structures, stakeholders, and edit ranges. All of these same lessons must be incorporated into any eventual implementation of VIOLIN. The problems and their resolution as part of the online IPEDS collections have much to offer the evolution of the VIOLIN/CDS collection.

To facilitate this shared learning process, Dr. Milam held regular discussions with NCES staff such as Susan Broyles and Dennis Carroll. Mr. Massa participated in the train-the-trainer program and helped train institutional researchers about the Fall web collection, IPEDS redesign, and PAS. Dr. Milam was to have served as a trainer, but moved from an institutional research position to faculty. Both have been involved in disseminating information about the new IPEDS web collection as part of conference presentations.

Dr. Milam served on the IPEDS Redesign Umbrella Team and on the NPEC Working Group charged with redesigning the Fall Staff and Faculty Salary surveys. He created an online feedback form for the AIR website to collect comments about proposed new tables of faculty/staff data and provided these comments in database format to NCES/ESSI staff. He also made presentations for NSF about its data collections and dissemination projects such as WebCASPAR and SESTAT on the web and served as faculty for the NSF Summer Database Institute in 1999 and 2000.

Mr. Massa participated in the NPEC Working Group on Reclassification of OMB Race/Ethnicity Codes, another area critical to the redesign of the IPEDS surveys and to the development of appropriate web applications.

Out of these many activities came the researchers' shared understandings of what would be expected in a VIOLIN-level collection. These requirements include the ability to upload a dataset rather than complete the online forms, as is still required with the Institutional Characteristics data. Key holders will need to be identified with various levels of access. Appropriate edits and ranges and room for documentation of anomalies will need to be built in. As will be discussed later in the section on peer analysis tools, the resulting dataset must be able to interface with the IPEDS Peer Analysis System, which would recognize the VIOLIN data as additional SQL Server Database tables of data which approved users may include in their analyses. Tutorials, training, and context-sensitive help must be designed into the applications to avoid overloading help desks.

In some ways, the development of the VIOLIN web data collection pilot project was faced with a "moving target." For the IPEDS web collections, as they were envisioned at the beginning of the VIOLIN project in June 1999, were very different than what was finally delivered and used for IPSFA in August 1999 and for the October 2000 web collection. Even the nature of the data elements to be included in these collections changed, mostly because of funding, but also because of the redesign process itself. One way to address this was accomplished through ANSWERS, for which Dr. Milam created special datasets for the Fall and Spring web collections. A third dataset was created to document for users

the "Deleted" data elements that were being removed from IPEDS. Some of these are now being brought back into the collection due to additional funding.

Based on the consensus reached with the pilot schools, it was decided that the prototype collection tool would use the Common Dataset as its data element dictionary and record layout. Dr. Milam included the record layout for the 1999 version of the CDS in the ANSWERS site, requiring extensive coding and creation of a data element dictionary. Mr. Massa also undertook this effort from another perspective. Recognizing that each year's CDS will change slightly with the addition and (hopefully) deletion of data elements, a SQL Server table was created by Mr. Massa to house the listing of data elements as they change over time. This table is then used to dynamically generate the collection forms by section for any given year. This alleviates the need to constantly update table structures and data entry screens with each new data collection year.

A Database-Driven Model for the Online CDS

At the heart of the VIOLIN design is a database-driven model created by Tod Massa, in which all online forms and screens are dynamically generated. This database model includes a special table of tables in the SQL Server database. Within each table, there are many different types of data elements. Depending on how these are listed and coded, each HTML screen is automatically generated to display a variety of data; including the previous year's data for a variable, an input box for submitting the current year's data, and any special editing routines which are required to ensure the data integrity of the variable.

With this model, it is possible to create new types of data collections virtually instantly, simply by listing the data element which is to be collected. Nothing is hard-coded in terms of HTML; rather, everything is created with ASP and the structure of the SQL Server database tables. In documenting edit routines, special SQL code is inserted alongside the variable name in a special column, to be run once the respondent has completed data entry. As in the online IPEDS collections, the user has the opportunity to override a flag which appears and state that while the data do not meet the edit they are still correct.

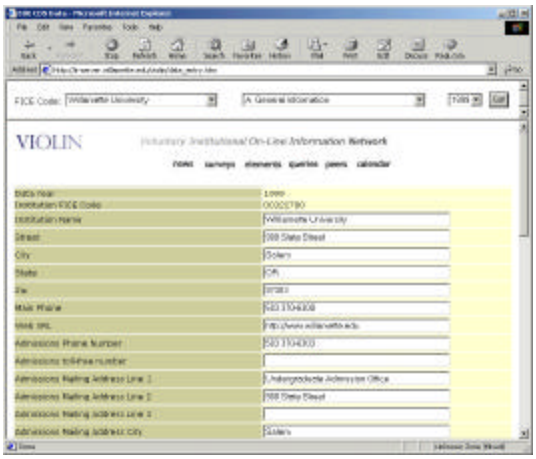
The Collection Tool

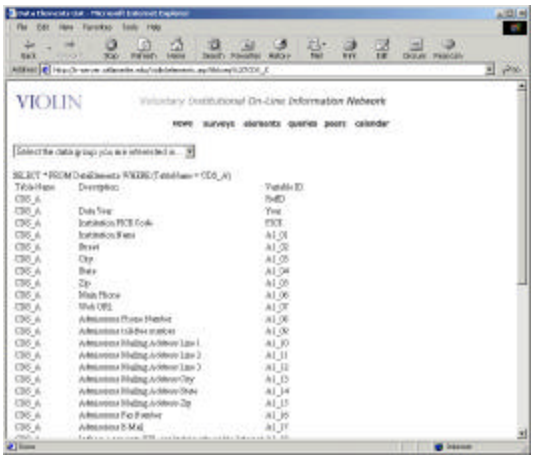
The final data collection tool developed for the pilot test or "proof-in-concept" of VIOLIN was developed by Mr. Massa using an NT Server he developed and maintained at Willamette University.

In the SQL Server database, there are a number of tables, most importantly the list of Data Elements. Fields include ID, table name, field name, description, field type, and collection (which allows for differentiation of years).

A total of 641 data elements are included in this prototype CDS. Database tables exist for each section of the CDS, from A through H. Tables also exist for FICE codes, institutions, and user groups. Other standard maintenance tables are also included. Extensive NT, Web Server, Secure Socket Layer (SSL), SQL Server, and application-based security are all employed to protect data entry and unauthorized access to the database.

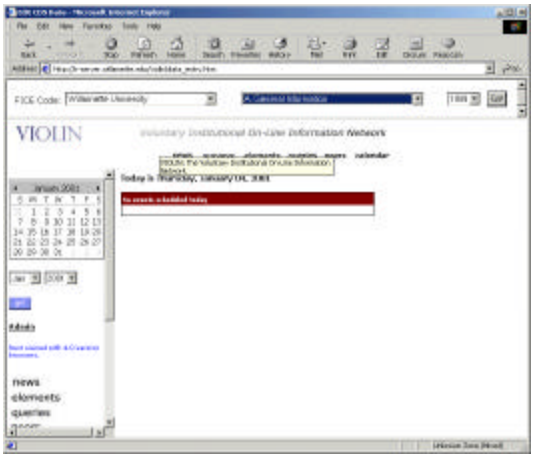
The basic data entry screen seen at the institutional level looks as follows. Users are able to select the institution, the section of the CDS they are interested in, and the year of data. An example of the display of data elements also appears.





Any final application would naturally include the standard institutional search and selection feature which is available in other IPEDS online tools such as COOL. For the purposes of this prototype, the focus is not on replicating these existing front-ends, but on developing a new approach that is unique to VIOLIN and interfaces seamlessly with the IPEDS Peer Analysis System.

A number of other features are included in the prototype web application, including a calendar, a news application, and documentation of data elements. These are displayed below as screen captures.



The prototype data collection tool does not currently include the editing, upload, or other verification tools. These would need to be developed in concert with the publishers, AIR, and sample institutions and must be tailored to the type of data collected. An extensive dialogue needs to occur if the CDS is adopted as the first VIOLIN collection to ensure that adequate edits are in place. The CDS advisory group and its listserv will be integrally involved in this process. It is expected that the vendors, working with AIR and the CDS advisory panel, will be able to document required edits and how missing data are handled so that the are willing to remove the CDS items from their individual surveys.

Conducting a pilot test of the collection

After the pilot institutions agreed to use the CDS for the prototype of VIOLIN, there was a significant and unplanned time lag while the web collection tool was created. In the future, this will be much less so, since the model of the prototype is based on dynamically generating the data entry screens

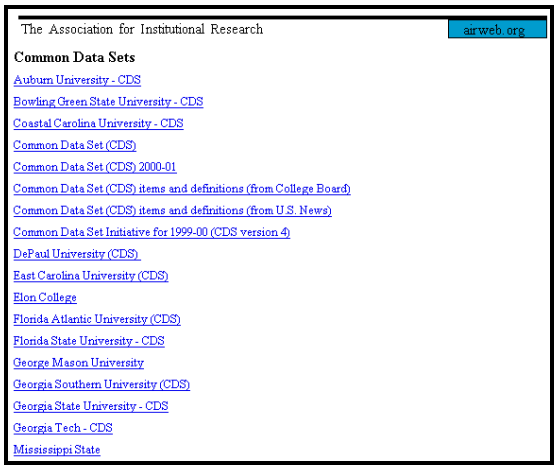
based on the table of database elements. For the pilot project timetable, this feature took much longer to develop than was originally expected. This is not do as much to the complexity of the task, which was still formidable, but to the workload of the researchers. John Milam directed the pilot school study and completed the programming, website development, and graphics support for the online survey and VIOLIN website. Tod Massa was the programmer of the data collection tool at the online CDS website.

Therefore, for this first round, the researchers agreed to collect the CDS in whatever format was available from the pilot schools, completing the data entry themselves. To date, data entry has been completed on a limited subset of schools. All schools submitted electronic versions of their CDS, though some schools took longer with repeated follow-ups to complete. Most of these came in Excel format. The holdouts were contacted and were the result of increased workload, not to any unwillingness to share the data in whatever format they were available. This does reinforce the point that some schools may not want to submit the data via data entry forms or uploaded file format. Some VIOLIN staffing will be necessary to accommodate this need for data entry from print.

While it would have been preferable to test the online collection of the CDS with data entry by the schools, it was not possible given the timeline and workload. However, a working prototype of all phases of this collection was built and demonstrated.

In a related effort, Dr. Milam maintains the AIR-sponsored website "Internet Resources for Institutional Research." As of Fall 2000, at least 39 institutions have provided links to their Common Dataset data. These CDS were voluntarily submitted without request. Other institutions provide the CDS as part of the institutional research office websites and a list of these is also maintained in the "Internet Resources" site.

Many schools reported at the AIR Technology workshop on "Building Websites," that was conducted by Dr. Milam for three years from 1997-1999, that they are interested in using the web to make their CDS more readily available. They would like to cut down on time spent responding to ad hoc requests for data by pointing people to the URL for their CDS. See the screenshot below.

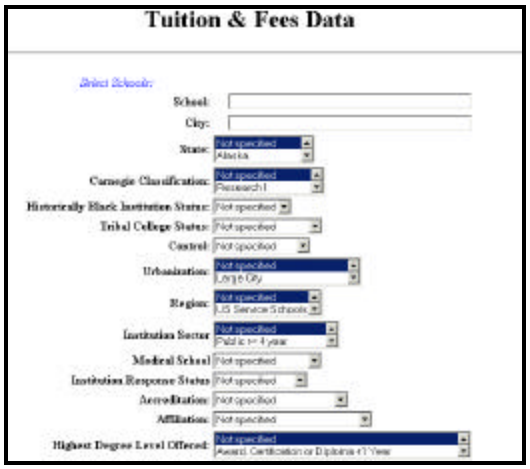


Developing a peer analysis prototype

At the outset of this project, Dr. Milam proposed to create a new type of peer analysis tool which would meet the needs of the new VIOLIN dataset. The revised, final project proposal dated April, 1999 includes the following plan:

After the web-based collection tool is visualized and prototyped, some sample types of peer comparisons will be developed using Microsoft web development tools such as Visual InterDev. Project staff will build web server, database application, and development server environments for this purpose. Discussions will be held with NCES about its own peer analysis tools, and comparisons will be made about the utility, training, and skill issues involved in using other data analytical tools for this purpose, such as commercial data mining software. NCES is working with the designer of Data Miner. There are many online analytical processing (OLAP) tools available. Out of this experimentation stage will come conclusions and recommendations about the best way to proceed in developing robust solutions to meet a variety of needs. The project should provide another source of feedback to NCES about these tools.

While it would incorporate features of the then available NCES search engines, the original Voluntary Core Dataset project proposal planned to expand these considerably. While still at George Mason University, Dr. Milam created a search engine and web application for IPEDS Tuition and Fees trend data from three years of the Institutional Characteristics file and shared this model with Dennis Carroll and Terry Russell. This tool is currently available at the URL: http://data3.gmu.edu/cfdocs/ipeds. Mr. Massa created a similar search engine and web application for IPEDS and other data, baccalaureate origins data taken from the annual Survey of Earned Doctorates and used by the Higher Education Data Sharing (HEDS) consortium. A screenshot of the tuition and fees search engine appears below.



A number of discussions were held during the first half of the pilot project by Dr. Milam and Mr. Massa about the value, utility, and problems of different institutional search engines and peer tools. This review included IPEDS COOL, the IPEDS peer analysis system, the IPEDS Early Release tool, NSF WebCASPAR, the Higher Education Data CD and other products from Minter and Associates, HEDS, and the AIR/NCES/NSF grant project conducted by Vic Borden and Tim Thomas to study "forms and formats for delivering information derived from IPEDS data sets." Each tool was tested and evaluated for suitability to VIOLIN's purposes.

Soon after the grant for a pilot VIOLIN project was awarded, a beta version of the Peer Analysis System (PAS) was created. Both Dr. Milam and Mr. Massa were beta testers for the online tool and provided feedback for its improvement. Once the design was tested and released to the public, both researchers continued to be heavily involved in using, demonstrating, presenting, and teaching about the

PAS. This was necessary as part of the VIOLIN project because the PAS (like the IPEDS web data collections) has remained until recently a "moving target," with new features, datasets, and capabilities evolving over time. Now, after a tumultuous but successful Fall IPEDS collection, NCES is much better prepared for future collections.

In planning for the project, NCES did not envision a separate peer analysis tool for VIOLIN. Rather, the web data collection in VIOLIN would be migrated like a new IPEDS dataset table to the PAS, where participating institutions would access the new data as if it were within IPEDS. The effort to develop a new and better peer analysis tool was subsequently dropped and renewed efforts were made to better understand the PAS.

In order for the VIOLIN project to interface with the IPEDS PAS, it was necessary for the VIOLIN developers to work closely with the PAS developer and the NCES web development staff. As part of the ANSWERS project, Dr. Milam worked to understand and implement NCES standards for online content. These involved standards for graphics, navigation schemes, meta tags, links, email, non-NCES material, SQL database design, new content approval, active server page coding, and other topics. Dr. Milam was able to successfully place the ANSWERS website on the NCES server and because of this it is expected that an NCES-compliant version VIOLIN could also be created.

Copies of the SQL Server database, ASP code, data element dictionary, documentation, and flow charts were obtained and used to study and analyze the PAS model. Dr. Milam met with the PAS developer, Mohamad Sakr, and later emailed questions to Mr. Sakr and to Sam Barbett of NCES. These materials were shared with Mr. Massa, who then spent considerable effort analyzing the PAS model and code to see how it could incorporate VIOLIN.

In addition to adding new tables of VIOLIN data to the PAS, other options have been considered as part of the VIOLIN project. These include duplicating the entire PAS on a non-NCES server, which would then be used for secure access of VIOLIN participants.

As part of dialogue about VIOLIN and the PAS, several vendors including Minter and Associates contacted the researchers to better understand where NCES intended to go with the PAS and whether other methods for disseminating IPEDS data would still be in place. NCES announced that it would phase out, with the completion of traditional IPEDS paper collections, the release of FTP files on its existing IPEDS website. All future data collected via the web would only be released via the PAS, where they would be available alongside critical IPEDS historical datasets.

Since the turnaround time for release of IPEDS data with web collection has changed from several years to less than two months, the PAS is envisioned as the best way to promote the timely release and use of data. Dr. Milam worked with John Minter and others to express concerns to NCES about extracting data from the PAS to NCES. NCES made the decision to provide complete dataset tables in comma-delimited, CSV text format as part of the PAS. Complete data dictionaries for each table are also provided. If additional manipulation of IPEDS data is necessary for vendors or others who do not for whatever reason wish to use the PAS, it is possible that VIOLIN could provide special data tabulations. This is possible since VIOLIN would have access to the redundant SQL Server IPEDS data.

An additional component for VIOLIN is envisioned which goes beyond what is available in the Peer Analysis Tool. Users will be able to cut an individualized version of the IPEDS and CDS datasets using HTML forms to select variables, years of data, and institutions. They will be able to choose or upload a file with a list of USERIDs for peer schools. They will be able to pick among the various tables of IPEDS and CDS data; then within these pick specific variables. The resulting dataset will automatically merge the different tables and present the data in a comma-delimited format for downloading and addi-

tional analysis. This concept was presented by Dr. Milam to NCES at a recent meeting of the SHEEO/NCES Network. It was mentioned to the subcontractor within Pinkerton Associates who is in charge of the Winter 2002 collection, who informed Dennis Carroll that this was an additional feature envisioned for the Peer Analysis Tool. It is certainly recognized within the context of VIOLIN that there is a critical need to get data to users without the front-end of the Peer Analysis System.

Demonstrating the tool and prototype

Numerous presentations have been made about the VIOLIN project by the project staff at the state, regional, and national levels. Some of these have been combined with presentations about the NPEC ANSWERS project. These presentations include the following:

- ACE-sponsored SARA financial aid users group (2000)
- AIR Foundations Institute (1999, 2000)
- AIR Technology Institute (1999)
- AIR/NSF Summer Database Institute (2000)
- Association for Institutional Research annual Forums (1999, 2000, 2001)
- Council of Higher Education Management Associations (CHEMA, 2000) annual meeting of association presidents and staff
- National Postsecondary Education Cooperative council meetings and steering committee (NPEC, 1999, 2000)
- Ohio Association for Institutional Research and Planning (2000)
- One Dupont Circle IPEDS group (1999)
- Pacific Northwest Association for Institutional Research and Planning (PNAIRP, 1999)
- Southern Association for Institutional Research (SAIR, 1999, 2000)
- Virginia Association for Institutional Research (1999, 2000)

For most of these presentations, individualized PowerPoint slideshows were prepared. For some of them, a live demonstration was conducted using a modem and/or notebook computer with a web server and duplicate copy of the VIOLIN website.

Presentations and VIOLIN Board of Directors

NCES and AIR staff have made additional presentations in which the VIOLIN project was discussed. Over the course of the pilot project staff, it was clear that the project had a life of its own that extended far beyond the activities of Dr. Milam and Mr. Massa – a sign of its potential success. AIR staff worked to create a VIOLIN Board of Directors and presentations were made to potential members and their respective organizations about the value of VIOLIN and how it would serve different constituencies. Potential board members were identified from the following organizations, including the four publishers:

American Association of College Registrars and Admissions Officers
American Association of Community Colleges
American Association of State Colleges and Universities
American Council on Education
Association for Institution Research
Career College Association
College and University Personnel Association
Council of Graduate Schools

National Association of College and University Business Officers
National Association of Independent Colleges and Universities
National Association of Student Financial Aid Administrators
National Center for Education Statistics
National Science Foundation
NPEC Accessing Survey Resources Chair
Peterson's
SHEEO Network Executive Director
Society of College and University Planners
The College Board
U.S. Education Department/Office of Postsecondary Education
U.S. Education Department/Planning and Evaluation Service
U.S. News
Wintergreen/Orchard House

One of the first tasks envisioned for the Board was the development of a set of general operating policies for VIOLIN. For example, when will data be released? Only schools which supply a specific dataset table will have be allowed access to the same data.

As discussed earlier, the NPEC Accessing Survey Resources project and VIOLIN have mutually supportive purposes. A product of the ANSWERS work is the identification of critical data elements for inclusion in VIOLIN. It was decided that the Board would determine, based on the priorities assigned by this NPEC workgroup and other information, what data elements will be included in VIOLIN.

Finally, it was decided that the Board will grant access to VIOLIN data to data collectors and researchers based on the recommendations of the AIR/VIOLIN contractor. Specifically, the AIR/VIOLIN contractor will accumulate requests for access to the VIOLIN data, including the purposes and reporting methods for such data use.

Evaluating and documenting the project

Short written and oral reports with project updates were provided upon request by NCES and AIR. The most recent of these was incorporated into the VIOLIN website as part of the "About" page. Special mini-reports were prepared for NCES staff to deliver to the NPEC Steering Committee and other groups. An interim report was also provided to AIR as part of the grant requirements for project funding.

Evaluations of conference presentations were reviewed when available, including written evaluations from AIR and SAIR. More informal feedback was solicited from audience participants as part of presentations at most of the meetings listed above. Generally, interest and enthusiasm for the project has been relatively lukewarm except for the proposal to use VIOLIN as a repository for CDS data. Most other uses are seen as too small (such as a single data exchange) or as an additional reporting burden. However, the database-driven model behind VIOLIN easily allows for the creation of special collections. The designer has merely to document the variables within each table and how these are to be dynamically generated within ASP and HTML for data entry.

An online Guest Book and threaded discussion group were hosted on the project listserv, but these tools were much underutilized and did not provide useful feedback. A methodological log was maintained by Dr. Milam for a period of time, but the extensive email correspondence and meetings with pilot institutions precluded using this as a helpful documentation tool.

Information about the project has been disseminated as part of conference presentations by Dr. Milam, Mr. Massa, and NCES and AIR staff. Detailed information has also been communicated in writing as part of the *AIR Alerts* print and online series and the Electronic-AIR listserv, both published by AIR. VIOLIN was discussed and evaluated by several different NPEC Working Groups involved with the IPEDS Redesign and is mentioned in the NCES redesign documentation.

Preparing a request for proposal (RFP)

Dr. Milam was notified in December, 2000 by Dennis Carroll of NCES and Terry Russell of AIR that future funding for implementation of VIOLIN had been cut and that it would be unnecessary to prepare a request for proposal. Since several of the project activities such as the online survey of pilot schools have been completed for some time, a list of requirements and deliverables is still outlined below. This listing is somewhat ambiguous, given the tentative nature of the design. If the "ORCHESTRA" design discussed by NCES for eliminating redundant federal data collections is implemented, the requirements will be far different and more complex than if VIOLIN is used privately to build an online CDS repository. The RFP which is outlined below is for the implementation of the CDS repository design.

Requirements for Full Implementation of VIOLIN

(1). The website developer must use NT Server, Internet Information Server, SQL Server, Active Server Page, and SSL technologies to provide a secure web server with adequate bandwidth to house the project.

(2). The developer must meet all standard NCES-type requirements for website design, including but not limited to navigation, SQL Server database structure, ADA compliance, and meta tags.

(3). The developer must maintain a complete, redundant version of the IPEDS database, including all currently available tables.

(4). The developer must work closely with AIR, the publishers, and the VIOLIN Board to incorporate evolving changes in expectations for web data collection, as documented in the online IPEDS and NSF web collections and new data element implementations such as new OMB-approved race/ethnicity codes.

(5). The developer must work closely with AIR and the VIOLIN board to incorporate training, tutorial, and context-sensitive documentation and presentations related to VIOLIN.

(6). The developer must design a dynamic, database-driven model for data collection which will meet the needs for all types of VIOLIN datasets, including the Common Dataset.

(7). The developer will work closely with AIR staff and committees and the CDS advisory panel and publishers to develop appropriate editing routines and online checks for data integrity.

(8). The developer will create web applications for secure collection of VIOLIN datasets through online screens and various standard methods of File Transfer Protocol.

(9). The developer will conduct timely web-based collections of VIOLIN data as determined by the Board of Directors, implementing appropriate technology and edit routines.

(10). The developer will screen and provide recommendations to the Board of Directors of requests for licensing agreements of the VIOLIN data.

(11). The developer will provide extracts of the data to the Board of Directors, participants, licensees, and other designated entities in approved formats in a timely and accurate manner.

(12). The developer will coordinate all actions with the VIOLIN Board of Directors and will take on additional responsibilities and activities as assigned by the board, such as working with potential data exchanges and/or groups of institutions to meet their data needs and resolving copyright or intellectual property issues.

(13). The developer will maintain current, advanced knowledge of data policy issues at the federal, state, and institutional level. This will include full participation in appropriate AIR, NCES, NPEC, NSF, SHEEO, CHEMA, NCHEMS, and other organizational activities as appropriate to promote the effective use of VIOLIN data. In particular, it is important for the developer to work with the AIR Higher Education Data Policy Committee, which will be chaired by Dr. Milam beginning in June, 2001, and with the CDS Advisory Panel.

(14). The developer will maintain current, advanced knowledge about all aspects of the IPEDS system. It is expected that the project director will visit and network regularly with the publishers to develop ties between these vendors and the project work.