NEWTRUCK: A COMPREHENSIVE LONG-TERM PROJECT IN COMPUTER SCIENCE

Daniel C. St. Clair*
Department of Mathematics and Computer Science
Western Kentucky University
Bowling Green, Kentucky 42101

INTRODUCTION

Recent literature in both the computer science [1,2] and data processing [3,4] communities encourages the involvement of computing students not only in formal classroom experiences but also in "real world" learning experiences such as project participation and management, effective approaches to "real world" problem solving, interaction with people from different disciplines, interaction with people from similar disciplines, interaction with people in the "real world," budget management, formal report writing, and the ability to accurately document a system. ACM has recognized this need in both the 1977 report of C^3S [5] and in the more recent "Curriculum '78: Recommendations for the Undergraduate Program in Computer Science [6]." Recent articles in the SIGCSE Bulletin [1] as well as other computer education journals indicate that universities are responding to this need through one or more of the following:

1. Integration of "real world" projects into the existing curriculum.
2. Special projects courses added to the existing curriculum where students are responsible to the university only.
3. Special projects done in conjunction with the university and some outside organization. The student is responsible to both the university and to the outside organization.
4. Cooperative education in which the student works part-time for an outside organization.

In items 1 and 2, student reimbursement takes the form of course credit, while in items 3 and 4, student reimbursement is course credit and/or monetary payment. Projects may involve one student or a team of students. Projects typically span a single school term; however, multi-term projects are not uncommon [7].

*Present address:

Computer Science Department
University of Missouri-Rolla
Graduate Engineering Center
St. Louis, Missouri 63121

This paper describes a rather unique "real world" learning experience at Western Kentucky University. The project is the establishment, maintenance, and enhancement of a comprehensive database of national motor carrier information. The project is a result of a 1977 contract between W.K.U. and Tymshare, Inc.-Washington Branch. W.K.U. is a regional state-supported university with about 13,000 students. At this time, the project has involved W.K.U. faculty and students from the Department of Mathematics and Computer Science, the Department of Economics, and the Department of Business Administration. Tymshare, Inc. is a computer time-sharing corporation with an international reputation.

At the present time, the database, named Newtruck by the students who helped get the project started, has been created and a 195 page Newtruck directory has been published [8]. In addition, several reports have been prepared for various governmental agencies. The project is currently moving into the maintenance, enhancement, and consulting stages. While maintenance and enhancement are primarily the responsibility of W.K.U. computer science personnel, close contact must be maintained between computer scientists, econometricians, and business people both at W.K.U. and Tymshare, Inc.-Washington Branch. The consulting service offered to outside organizations requires close collaboration among all parties involved.

In its relatively short span of existence, the Newtruck project has helped W.K.U. provide pertinent "real world" experiences for many of its students in computer science. The remainder of this paper gives a detailed description of the project as well as a description of students' involvement in the project. Technical and administrative problems are discussed. Past and anticipated results of the project are described as they relate to the educational goals of the University and particularly to the computer science program. The paper concludes with suggestions as to how other universities could become involved in similar projects.

PROJECT DESCRIPTION

The basic idea behind Newtruck is simple. Each year, motor carriers are required to submit an annual report, similar in appearance to a book of 1040 income tax forms, to the United States

Interstate Commerce Commission. The I.C.C. then records a large subset of this information on magnetic tape. This information is recorded by carrier number (I.C.C. assigned), report schedule number, and report line number. On multiple column schedules, several items of data are associated with a single carrier, schedule, and line number combination. Although this is public information, its use has been limited due to the expense involved in data retrieval and analysis.

The objective was to archive the data by individual carrier from 1965, the first year it was recorded by I.C.C., through 1977. Additional data will be added to Newtruck as the tapes become available. The database management system selected had to be easy and inexpensive to use for the inexperienced programmer yet powerful enough to allow the experienced programmer to do several types of data analyses. Tymshare's relational database management system Express [9,10] was chosen for the job. Data are maintained for both Class I and Class II carriers of common freight. The class designation of a motor carrier is determined by I.C.C. and is based on the carrier's net income.

The relationship between W.K.U. and Tymshare, Inc.-Washington Branch is two-fold. For Tymshare, it provides a valuable sales resource. Clients have nationwide access to the data on Tymshare's system directly through their own terminals or through W.K.U.'s consulting service. For W.K.U., Newtruck provides valuable "real world" experience for both students and faculty, it provides a unique and valuable research resource for both students and faculty, and it provides nationwide recognition for W.K.U. By agreement, Tymshare, Inc.-Washington Branch furnishes all computer time for data archival, maintenance, and enhancement in return for exclusive vendor rights for a set period of time. W.K.U. provides all personnel for the project. In return, W.K.U. is the sole owner of the database and receives a royalty for client use of the database. All royalties are returned to the project to pay personnel and equipment expenses and to purchase computer time for research use of the data.

Students' involvement in the project more closely resembles a combination of items 3 and 4 in the INTRODUCTION of this paper. (W.K.U. already has a course similar to item 2 [11].) At this point, students' association with the project has been on a work for hire basis with no university credit being awarded. The students' roles in the project include:

1. The design, documentation, and implementation of various subsystems such as data loading, data handling, and data verification. This includes the precise definition of problems which are often very obscure.
2. The interaction with people from various backgrounds at the University level, the corporate level, and the client level.
3. Project budget management. Students must be particularly aware of budget management since development time is furnished by Tymshare and client time is billable. Students are required to carefully manage all computer time.
4. Formal reports to University and Tymshare officials. The project plans to involve students at the client reporting level in the near future.
5. The preparation of an extensive database directory [8]. The directory itself was created, edited, and printed using system software.

Faculty make every effort to keep students involved at all levels of the project. They are not simply hired programmers. The result is that students gain practical experience in computer organization, systems, and languages as well as in information systems. Feedback on project performance from people outside the University is valuable to both the students and to the development of University programs. To date, four computer science students, two computer science faculty, one member from the economics faculty, and one business administration faculty member have been involved. As the project expands, plans call for the involvement of additional computer science students as well as students in economics, business administration, and accounting.

DESIGN AND IMPLEMENTATION PROBLEMS

At this point, Newtruck sounds like a standard database project. The scope of the project can better be understood by an examination of some of the problems that have been encountered to date.

One of the major problems was the sheer size of the database. For the eleven year period, 1965-1975, there was an average of 551 motor carriers per class per year. During this same period, the average number of schedules recorded per class per year was 24. Since each schedule contained an average of 38 data items, Newtruck contains over eleven million items of data for this eleven year period! Data was filed in arrays with each array representing a unique data item, such as assets, for a given class and a given year. Each array is subscripted by carrier number. The eleven year period required over 19,500 different arrays! This data alone filled an 808 cylinder, 3330 pack! Problems such as exceeding core storage and overflowing the volume table of contents of disks were common. Special data structures and data handling routines had to be designed to bypass these problems. "Simple" design problems included the development of a variable naming convention which could be computerized easily.

A set of major problems involved the overall inconsistencies in the data. Several of these inconsistencies are enumerated below.

1. The I.C.C. report forms changed from year to year and class to class. An item in a given schedule for a given year and a given class might not appear in the same schedule for subsequent years and classes.

2. The I.C.C. did not record the same schedules, lines, and columns for each year and class. In some years, as many as 32 schedules were recorded per class while in other years, only 12 schedules were recorded per class.
3. Carrier numbers were not unique from year to year and class to class. Since the first digit of each carrier's number indicates its class for a given year, any change in a carrier's class required the assignment of a new carrier number for the following year. Carrier class changes were caused by large changes in net assets or by a change in I.C.C. class definitions. The old number was frequently reassigned to another carrier. In addition, carriers passed into and out of existence during this time period.

Since the data are used primarily for time-series analyses, all of these problems had to be solved. Where possible, software was used to reconcile inconsistencies. For example, the carrier number problem was solved by assigning a unique number called a docket number to each carrier. Each carrier had to be traced throughout the eleven year period and each of his different carrier numbers recorded. Much of this reconciliation was performed using software. The remainder had to be done by hand.

Verification of the data proved to be a major project in itself. To ensure correct transcription of the data from the original tapes to Newtruck, several small subsystems were created. The first subsystem went through the original data by year and class and made a tally of the number of responses for each schedule, line, and column. The second subsystem printed out complete carrier reports for a random sample of carriers. A third subsystem computed yearly and class totals for all carriers for such major items as assets, liabilities, and net income (loss). All of these subsystems were applied to the raw data and were written in Cobol. A similar set of subsystems was created in Express to check the data actually loaded into Newtruck. The latter was not done until a random check showed that data for several carriers was missing! The problem was a simple bug in an undocumented Cobol loading procedure! It took three days and a lot of hard work to locate and replace the missing data. No further remarks had to be made to encourage clear and accurate documentation.

Problems in communication between faculty and students, faculty and the people at Tymshare, and students and the people at Tymshare were common in the beginning stages of the project. Everyone quickly learned that in a project of this magnitude, communications must be clear and explicit in order to avoid unproductive work efforts.

Computing time was constrained to be between 6 p.m. and 6 a.m. on weekdays and all day on weekends and holidays. This restriction coupled with students' class schedules dictated that much of the work be done at odd hours, weekends, and during holidays. These conditions along with subsystem deadlines often meant long and sometimes trying hours. While it was anticipated that conflicts might become a problem under such conditions, they never surfaced. It is suspected that the students' enthusiasm and dedication to the project were the main contributors to this success. Various ad hoc discussions on egoless programming were often held over late night hamburgers and french fries.

Project workers had to be constantly aware of the fact that they were preparing a product to be marketed. While the actual marketing would be done by Tymshare, Inc.-Washington Branch, the marketability of the product would essentially determine the success or failure of the project. Students were very helpful in making suggestions, designing subsystems, and producing system documentation and descriptions to enhance marketability.

The administrative problems encountered were new and unique for a project of this kind. While similar projects are often performed on a grant basis, the Newtruck project is not. Continuance of the project is solely dependent on royalties produced by Newtruck. The Newtruck project is indebted to several W.K.U. administrators for their firm support.

OBSERVATIONS

Although the Newtruck project is a little over two years old, its positive impact has been firmly felt by the students, the faculty, the administration, and the programs of Western Kentucky University. As the project moves from the development phase to the maintenance, enhancement, and consulting phases, the observations cited in this section are expected to continue to apply as the project expands.

Perhaps one of the most visible impacts has been on the students. Those students directly involved in the project have developed a pride in being actively involved in a unique and nationally significant endeavor. This is evidenced by their dedication to the project and the impact of their discussions with their fellow computer science students. Students outside the project continue to express an interest in the project from the standpoint of becoming involved themselves and/or from the standpoint of the types of problems encountered. As the project expands, more students will be invited to participate.

In addition to pride in one's work, students developed in several other important areas. Some of these areas of development include the ability to:

1. Work with people in the outside world as well as with fellow students.
2. Provide accurate and complete system documentation. Since subsystems often required the use of previously written subsystems, students quickly became aware of the importance of clear and accurate documentation.
3. Ask the "right" questions in order to clarify project objectives.

4. Use perspective in decision making. They learned when to make their own decisions and when to ask questions.
5. Budget time and resources for a project.
6. Solve problems which don't have "textbook" answers.
7. Work effectively on large-scale computing systems.

Many of these objectives can be met with other forms of special projects. However, the magnitude of this project coupled with the success/failure possibilities produced a learning environment different from other types of projects.

As of January '79, three of the students in the project have graduated from the computer science program at W.K.U. Their reception in the computer science profession indicates the favorability of this type of experience for students.

Newtruck is also having a positive effect on the University's programs in computer science, economics, and business administration. The project has lent impetus to the development of an undergraduate course in database management systems. It has also been a contributing factor in the development of the Mathematics and Computer Science Department's master's degree proposal. The recent publication of the Newtruck directory [8] is beginning to elicit research and curriculum proposals from several University programs.

CONCLUSIONS

The Newtruck project is similar to programs at other universities which are designed to introduce students to computing in the "real world." While it possesses the advantages of these programs, it differs from many of them in that:

1. It is a unique contribution of national significance which involves diverse departments of the University working together with an outside organization.
2. The project is open-ended. Students' participation extends beyond semester boundaries. Their dedication helps to ensure continuation of the project.
3. The project is easily expandable to other forms of transportation.
4. There are a myriad of opportunities for student involvement. Database design, maintenance, reporting, and consulting illustrate a few of the opportunities in computer science.
5. Work is evaluated almost entirely by people outside the university environment. Often, the evaluators are people from outside the area of computer science.

The opportunities for universities to engage in projects of this nature appear endless. While short term university-company projects are not a new idea [7,12], the concept of a long term project is new. It requires not only the support of university faculty and administration, but the willingness of an outside organization to become involved in a long term project. Such long term organizational commitments do not seem unreasonable in view of the current computer manpower shortages [13]. Furthermore, the choice of a project must be done carefully to ensure that the project offers a variety of opportunities to students and faculty. While the involvement of disciplines outside of computer science requires additional organizational work, it lends latitude to students' experiences and better helps establish the project within the university.

The establishment of a project can occur in several ways. The Newtruck project started with a proposal submitted to the Department of Transportation. Although the proposal was not funded by D.O.T., Western Kentucky University and Tymshare, Inc.-Washington Branch decided to pursue the project independently of government support. The Federal Register, a weekly publication of the U.S. Government Printing Office, and Federal Notes, a biweekly publication of the Federal Development Association of Saratoga, California, provide information on government grants and granting agencies. This information may be used to submit solicited as well as unsolicited proposals.

Computer support services should be chosen carefully. Nationally known computer services organizations such as Tymshare, Inc. can provide such things as software support, product accessibility, and a sales staff not available from a university computer center. Such organizations often provide toll-free telephone services which make them readily accessible. Even if a project is not funded, it may result in a contractual arrangement similar to that used in Newtruck.

Regardless of the funding approach used, the concept of a continuing contractual arrangement between a university and a private organization is not commonplace. Such arrangements have to be carefully tailored to each project, to each university, and to each outside organization. Although this is time consuming, the results can produce a growing, self-supporting resource of national significance for the university.

REFERENCES

1. SIGCSE Bulletin, Vol. 10, No. 1 (February, 1978).
2. Goddard, A.R., "Relating Computer Science Programs to the Needs of Industry Through Internships and Cooperative Programs," SIGCSE Bulletin, Vol. 6, No. 1 (February, 1974), pp. 165-168.
3. Heiker, V., "The Education of the Systems Analyst," Journal of Systems Management, Vol. 25, No. 9 (September, 1974), pp. 18-24.
4. Heiker, V. and Galli, D., "Is Today's Systems Education Adequate?," Journal of Systems Management, Vol. 27, No. 1 (January, 1976), pp. 24-27.
5. ACM Committee on Curriculum in Computer Science, "Curriculum Recommendations for the Undergraduate Program in Computer Science," SIGCSE Bulletin, Vol. 9, No. 2 (June, 1977), pp. 1-16.

6. ACM Committee on Curriculum in Computer Science, "Curriculum '78: Recommendations for the Undergraduate Program in Computer Science," Communications of the ACM, to appear.

7. Perry, J.M. and Sondak, N.E., "The Project Experience in Undergraduate Computer Science Education," SIGCSE Bulletin, Vol. 10, No. 2 (June, 1978), pp. 21-28.

8. St. Clair, D.C. and Wassom, J.C., NEWTRUCK Directory of Motor Carrier Data, Bowling Green: Western Kentucky University, (June, 1978).

9. EXPRESS Introductory Manual, Weston, Mass.: Management Decision Systems, 1976.

10. EXPRESS Reference Manual, Weston, Mass.: Management Decision Systems, 1974.

11. Crenshaw, J.H., "Team Projects in the Undergraduate Curriculum," SIGCSE Bulletin, Vol. 10, No. 1 (February, 1978), pp. 203-205.

12. Powers, M.J., "The Directed Project: A Poor Man's Internship," SIGCSE Bulletin, Vol. 10, No. 1 (February, 1978), p. 97.

13. Hamblen, J.W., "Computer Manpower in the United States-Supply and Demand: An Update," SIGCSE Bulletin, Vol. 9, No. 4 (December, 1977), pp. 19-37.