N73-19210

VIEWS ON COMPUTER PROGRAM DOCUMENTATION AND AUTOMATION

Dr. Herbert R. Grosch
National Bureau of Standards

The topic of this talk is not really computer program documentation and automation. Because the Bureau of Standards is not very advanced in these fields, this paper will confine itself to comments about various aspects of the problem of documentation and description. The first aspect I will discuss is what has been called the problem of semantics.

Scientists have access to three levels of data. The first is the personal information that is often written on the backs of IBM cards or 3 by 5 cards: the kind of information that is used to make a particular project work. The second level, more formalized and organized, but usually not more mechanized, includes printed catalogs, formally kept laboratory note-books, and computer printouts with comments for a unique program. Finally, there are the mechanized data banks and the mechanically formalized documentation of the computer.

The topic of this symposium is how to transform the second level of information into the third level of information with a minimum of thought, time, and attention. This will be much more difficult than you may realize. Information has no independent life of its own. What it means to each of us depends on our individual backgrounds, training, and experience. The same set of figures that means a useful window for blastoff to an engineer on the ground means an extremely narrow passage for an astronaut who is going to have to squeeze through that same window.

Similarly, the information produced by documentation of a program, if presented to several different groups, may mean entirely different things to these groups. The solution to the problem of documentation, in my opinion, is not strict formatting or extensive use of standards. Nor is heavy investment in professional regularization of documentation and of the surrounding data the solution. Instead it is the recognition of the fact that people look at problems differently.

If this recognition were taken to an extreme, it could be argued that everybody should document his own work and do everything by hand. But that would be an impractical solution. After all, the purpose of the English language, of installation standards, and of programming manuals is to provide a common ground for the transmission and use of information. The goal of this symposium is to look for ways to provide this common ground quickly and inexpensively. This is an admirable goal. But unless work toward this goal is accompanied by a recognition of the differences between the ways men look at information, the goal will not be realized.

An example of the problem created by these differences can be seen in the field of software exchange. If a programmer tries to keep track of his system so that he can hand

it on to a successor working in the same position for the same organization on the same computer and solving essentially the same problem with just a different set of input numbers, the chances are that there will be very few problems. One reason is that the information to be exchanged is done almost always between men of similar backgrounds and training, at least as far as this particular problem is concerned.

However, when a company decides to send a program to COSMIC so that it can be distributed to people with similar backgrounds and training to be used in solving a similar problem on similar equipment, the task of transmitting the same information is much more difficult, as most of you know. The paper presented by the COSMIC representative illustrates this perfectly.

Finally, if a person decides to go into the business of selling programs, i.e., transmitting programs to strangers working on different problems on different machines, the problem of documentation becomes almost insurmountable. If you talk to people who are in the business of selling and maintaining programs, you will find that perhaps 80 percent of their costs are essentially maintenance and documentation costs.

The prospects for using machine language alone in seeking a solution to the problems of documentation are quite discouraging. The problem with machine-independent language is that it fosters the illusion that it is possible to document without rethinking; i.e., without thinking about what needs to be changed to use a program in a new system or new machine.

The difficulty of documenting cannot be eliminated because documentation cannot be done automatically; thought must be involved. However, the automatic approach does have the important advantage that changes can be made easily while documenting. Automation streamlines the process of picking out and putting together the elements needed to document a particular program.

Another aspect of the problem of documentation is its cost. Nobody knows within 30 or 40 percent just what computing and data processing cost the Federal Government. One probable estimate is that computing and data processing, communications, and library science information technology with its information retrieval, documentation, and microfilm cost the Federal Government well over $10 billion a year. It costs the United States as a whole perhaps $50 billion a year.

One reason for this enormous cost is that all the computers, information retrieval devices, and other specialized devices operate at no more than 5 to 30 percent efficiency at best. This is certainly not very efficient.

Despite the fact that it is patently impossible to run anything at 100 percent efficiency, there are some ways of reducing costs by a factor of 2 or, keeping costs the same, increasing throughput by a factor of 4. This symposium is an attempt by the technical community to do this singlehandedly. It is doomed to failure because management and administration have not kept up with technological advances. To put it somewhat facetiously, the technologist makes things as good as they are, and the manager makes sure they do not get any better.

The kind of organizational and administrative change and the detailed management needed to see that technological advances are used is lacking not only in Government but also in business to a great extent. Moreover, it is almost impossible for technological

improvements to be used properly. The principle by which people are promoted from technical positions to high-level managerial positions effectively excludes anyone with a real desire to achieve efficiency.

The only solution to this problem seems to be to attack the problem this way. There do exist large-scale, on-line systems that work because they have to. NASA is an example of one, the airline reservations systems are another, and the stock market is yet another. These systems are publicly visible and are subject to economic accountability. If the astronauts do not land alive, if an airline reservations system does not work properly, or if someone loses $50 million worth of stocks, then somebody will be fired. These systems should serve as models for the solution to the problem of how to keep our costs within reasonable bounds.