

Report of Task Force 3
Organization and Storage of Data

Oguz Caglarcan

The topic of discussion for Task Force 3 was the organization and storage of data. We had 8 items that were discussed. At the beginning, it became clear that, for each one of these 8 items, what we had to do was to make sure that we had a common understanding of what each one of the items meant.

1. The formats of data.

The way that was interpreted was that it related or referred to the internal representation of data within a system, as opposed to format in relation to inputting data into a system. The main concern in conjunction with format was in relation to the transportability of data from one system to another. The main comment was that as far as formats internal representation of data within a system was concerned, it was highly desirable that one be as much machine independent as possible. And those of you who are knowledgeable in a computer field recognize that there are certain types of formats that are more machine dependent than others. Also, the question of standardization of format, again, in relation to internal representation of data, was raised. There would be, I believe, it was recommended that, in terms of that question itself, that would require further study and analysis. But again, the emphasis was on transportability of data from one system to another.

Comment: The task force agreed that the most critical data format concerns were related to internal representation of data within the computer. The task force strongly recommends the use of "machine-independent" formats as much as possible to enhance transportability of data from one system to another. Therefore, any efforts toward achieving format standardization must address these concerns.

2. Media for sharing data.

Again, the interpretation was made, and basically the understanding was that if two user groups wanted to transfer data, would data come on cards, on tape; what sort of storage media would be used. Just about all possible forms of media were identified and the ultimate capability, the ultimate desirable was the on line communication. In other

words, two systems talking to one another through high speed data transmission, which would imply there would be some kind of communication network. However, it was recognized that an on line communication capability is an expensive proposition. It was also indicated that the security aspects of that kind of a transmission would be a concern. Also, it would require development and standardization, as well. Obviously, the positive side, in terms of why that would be the ultimate capacity desirable, was expressed in terms of the speed and minimal clerical manual operations. The second best selection was tape. The pros and cons, the advantages and disadvantages of what was user transportation, large storage capability, more standardization that exists as far as tape storage is concerned. The only disadvantage that was identified was that there would be some, in some cases, conversion may be required. In other words, if you do take data on tape from one system to another, it is not necessarily so that you would be able to automatically read that tape on your system. So there may be some conversion that may be required. As far as cards; basically the general opinion was that they're bulky, easy to destroy, or get damaged, and it is not necessarily so that every computer system has a card reader. Cost of shipping was also mentioned. Disc was also discussed and basically the opinion was that it was an expensive form and also there was a concern as far as standardization in relation to disc. In general discs are unique to a given system. Another one that was discussed was micro-fiche. One of the advantages; easy to mail, doesn't take up much space, human readable. However, it was indicated that, again, it's more error prone. The last one that was discussed was printed documents. All the other ones are obviously in machine readable format, while the printed documents would not be. The disadvantage was the fact that the printed documents are not machine readable, while the advantage was that they are human readable.

Comment: The task force analyzed the common types of media for sharing data and concluded that the magnetic tape is the most feasible medium while on-line transfer of data through high-speed communication lines is the most desirable method for sharing data.

3. Data base management.

Again, the interpretation that was made was that it means updating and maintaining a data base. It was agreed upon that management of data is a concern in that it affects certain critical factors, such as accuracy, integrity, completeness of data. And we also discussed, that information is definitely necessary in relation to different management methods that are being used when users are sharing data. In other words, a user who is sharing data with someone else would like to know as much about the management methods that

have been applied to that data base as possible. We also discussed the importance of identification of references. Reference material, sources of data, in conjunction with management. Again, relating to the critical factors of accuracy. There was a concern on that. We didn't go far on that, but I think the overall agreement is that is a topic, a point that needs to be further investigated.

Comment: The task force considered data base management to be an area of concern as it affects accuracy, completeness and integrity of data. The task force recognized that information and management methods used is necessary in transferring data. In addition, the task force recommends an investigation to determine the feasibility of developing guidelines and/or minimum standards for management of nutrient data.

4,5. We felt could be discussed together
Number Four being information sharing procedures and Number Five being communication with users. And basically, the interpretation of those two items centered around procedures or methods of information dissemination. In particular, in disseminating knowledge in relation to who has what and how to get to it. There were several means for accomplishing that, that were identified. Publications were mentioned, Newsletters, Conferences. Obviously, knowledge of user community very important, direct mail also; directories, dictionaries, catalogues in terms of disseminating the knowledge as to what exists.

Comment: Topics four and five were discussed together. Methods of information dissemination were emphasized for informing users on availability of nutrient data bases and systems. The following methods were identified:

- Publications
- Newsletters
- Conferences
- Dictionaries, directories, catalogs

6. Multiple use of data.
Basically, in addressing that topic, it was agreed that the emphasis should be on the implications of having more than one type of use or application, for a specified set of data. What are the implications of having more than one type of use on a given data set? It was felt that if we could enumerate those, it would be a large production. The ones that have been identified are responsibilities from the point of management of data. Inherent limitation, because of having multiple uses of data is the flexibility with which we could manipulate. This has to do with common

interests. It was emphasized that long range planning is very important, in terms of how that data would develop and evolve over a period of time.

Comment: This topic was discussed in terms of the implications of having more than one type of use or application for a specific set of data. The following areas of concern were identified for further investigation:

Assignment of responsibility for management
Policies and procedures for maintaining common interests
to facilitate continuing usefulness of data
Long range planning to promote growth of data

7,8. Turn around time and ready access.

We had quite a bit of discussion on these and the conclusion was that the question of turn around time and ready access were not really a concern in conjunction with the level of detail that we were addressing these topics and did not quite belong within this grouping. However, that does not mean they are not important factors. Therefore, the recommendation was that at this point, we do not address the question of turn around time and ready access.

Comment: The task force failed to recognize any concerns regarding these topics within the framework of the data organization and storage study area.

As far as specific discussions of the 8 items, that was basically it. The question of standardization, in general; we had a discussion on that. The general opinion is that there are definitely a number of areas where at least in terms of guidelines or minimum standards (guidelines was a key word) that there's a definite need for that in terms of the degree of information that should be made available when data is being shared in relation to formats, internal representation, as well as formal input and output.