

Sharing Information on Development

Working Paper 4: Technical support for information exchange

1 Tools and methods

1.1 The capture, collection and processing of development information is a major undertaking, especially in large organizations with many departments involved in activities in different sectors and geographic areas. Internal procedures must be designed to ensure that information is collected not only in a timely and cost-effective fashion, but also in such a way that the information is available to be shared with other organizations. Resources must be allowed for the development of appropriate forms, standards, and guidelines; for the training of staff in their use; and then for sufficient manpower to implement the procedures in the way in which they were designed. While some organizations have adopted computer technology to assist in this process, many organizations will continue to find paper-based systems an effective means of information collection, processing and dissemination.

1.2 *What tools or methods for information collection or processing have been effective in your organization? What kinds of tools or methods would you like to see developed in the future to assist in the processing of development information within your organization?*

2 Computer technology for exchange

2.1 The use of computer technology can make the process of collecting, disseminating and using development information faster, simpler and easier. Many agencies have already implemented computerized systems for managing information about their own development activities. The results from the survey of the Informal

Study Group indicate a wide range of different computer hardware and software in place. Agencies will not be able to standardize on specific computer hardware and software simply to facilitate the exchange of development information.

2.2 However the successful exchange of development information does not depend on the hardware and software used internally by an agency to process development information. Data can be effectively shared as long as it can be sent in a standard format from one agency to another. Standards are required for the definition and content of each data element; the *Proposed Common Exchange Format for Development Activity Information* represents this kind of standard. However this proposed format does not address questions of the physical format of computer-readable files. The standardization of physical format of machine-readable records is important to sharing information effectively in an automated environment.

2.3 One standard which might be suitable for the exchange of development information is the ISO 2709 standard for the exchange of bibliographic information¹. This standard allows the exchange of variable length textual data elements in variable length records. The ISO 2709 structure is well established, and is supported by retrieval software already used by some development agencies. However the record structure defined by this standard is quite complex: part of the record is a directory containing pointers to another section of the record where the values in all the fields are actually stored. For an agency using software that does not support this standard, writing programs to read and write records in this format can be a major undertaking.

2.4 A simpler approach to a standard format for machine-readable records could be developed for the exchange of development information. This standard might make use of a simple predetermined method of identifying the beginning of records and the beginning of fields. A record concerning a development activity would begin with a particular sequence of characters marking the beginning of the record. Following this would come the identifier of the first data element

¹ International Organization for Standardization. Documentation - Format for Bibliographic Information Interchange on Magnetic Tape. Geneva: ISO, 1981 (ISO 2709-1981).

(called a tag), followed in turn by the data actually contained in that field, perhaps extending over several lines in the physical file. Tag and data would repeat for each data element describing a particular activity. (Indeed this kind of record structure is sometimes called a *tag-data* format because the data elements follow one another within the record in the sequence *tag data tag data tag data*.)

2.5 The advantage of this second kind of format is that it accommodates variable length textual data and yet it is easy to produce from a wide variety of systems. However the use of this kind of format would require agreement on the exact form of the record identifier and the tags to be used to identify the data elements. A specific standards document would have to be written to spell out the precise format, including the tags or identifiers used for each of the standard data elements. The format should also include provision for the exchange of data elements which are not part of the standard, but which might be specific to the originating agency. In some cases, programs might still need to be written or used to transfer data from the local database to the exchange format or vice versa.

2.6 What standard or standards are suitable for a machine-readable format for the exchange of development information? Would your agency support the use of the ISO 2709 format? Would your agency support the development and use of a specific tag-data format for the exchange of development information?

2.7 Even if a machine-readable format for the exchange of records can be established, standards also need to be established for the characters which are used within each record. There are many conflicting standards for expressing characters in computer-readable format. One common standard, the ASCII character set, was developed in the United States for the storage, processing and exchange of information in English. No provision is made for characters needed in other languages, such as the e with acute accent in French, the a with angstrom in Swedish, or the o with tilde in Portuguese. Various extensions to the ASCII character set have been developed by different manufacturers. Some of these have become quite widespread, such as the extended ASCII character set used in IBM and IBM-compatible microcomputers. However none of these character sets are universally accepted and none handles all the

characters needed for all European languages. These standards also make no provision for languages that do not use the Latin alphabet, such as Japanese.

2.8 There are various other international, national and de facto standards for characters in languages other than English, including standards used for specific kinds of data, such as text, bibliographic or library cataloguing records². While these standards provide a hardware- and software-independent means of coding a wider variety of languages, they are not widely supported by computer manufacturers or software developers. Use of one of these character coding schemes for machine-readable development information would require computer programs on most systems to translate characters from the format used internally by an agency into the standard format used for exchange purposes, and vice versa.

2.9 *What character sets would you like to see in information your organization receives? What character sets would you like to use to send your development information to other organizations? Is it worth keeping accented or special characters in the data or should simple English characters be substituted for accented or special characters required in other languages? Would your agency be able to support the translation from one character set to another when development information was being imported or exported?*

3 Physical format

3.1 The physical medium used in the exchange is also an important consideration. Computer files can be exchanged online, and via a wide range of media, including compact disks, magnetic tapes, and floppy disks. Compact (CD-ROM) disks provide low per-disk production and distribution costs for large amounts of storage in relatively large runs

² See, for example, International Organization for Standardization. *Information processing - Coded character sets for the transmission of text*. Geneva: ISO, 1983 (ISO 6937); International Organization for Standardization. *Extension of the Latin alphabet coded character set for bibliographic information interchange*. 2nd ed. Geneva: ISO, 1983 (ISO 5426-1983); American National Standard for Information Sciences. *Extended Latin alphabet coded character set for bibliographic use*. New York: ANSI, 1985 (ANSI Z39.47-1985); and Network Development and MARC Standards Office. *US MARC specifications for record structure, character sets, tapes*. Washington: Cataloguing Distribution Service, Library of Congress, 1990.

(more than 250 copies); however they require special hardware and software to be used. Magnetic tapes come in different densities and in different formats (cartridges and reels). Floppy disks used with microcomputers come in different sizes as well as different densities. Smaller organizations, such as those with only microcomputer equipment, may only be able to produce or receive data in one of the lower density floppy disk formats. While the physical medium used to distribute files may seem a trivial problem, when updates of project information are to be sent frequently to a number of organizations, sending them all in one or two formats reduces the time and cost involved in the preparation of the exchange records.

3.2 Is there a need to establish standards for the physical media used in exchanging development information or is the question of formats best left to individual agreement between the two agencies involved? Can your organization process information in all formats, including CD-ROM? Can your organization produce information in all formats, especially the common low density formats, i.e. 1600 bpi magnetic tape, 720Kb 3.5-inch floppy and 360Kb 5.25-inch floppy disks?

3.3 Information can also be shared through allowing outside users access to information on a computer used by the originating agency. In this case, the data is not sent to the external users, but rather outside organizations are given access to the data or a subset of data on the host machine. The data should still be in a standard format to facilitate its use by external users; this approach only seeks to avoid the cost to the originating agency of preparing disks or tapes and the cost to the recipient agency of storing a copy of an external database. However remote access involves other additional costs. There may be hardware and software costs to both agencies to provide the means for an external agency to connect to the host computer, and there are certainly on-going telecommunications costs based on the amount of use made of the service. Remote access also gives rise to added training costs, as users must learn to use the software on the originating agency's computer. The originating agency must also plan carefully to ensure the security of the host computer and the data stored on it.

3.4 While computer and telecommunications technology is wide spread, it is not necessarily readily available in all organizations, especially institutions in developing countries. However information

can still be exchanged, regardless of computer technology, if it can be produced in print or microform format. Yet reproduction costs (in the case of microform) and printing and shipping costs (for information in paper form) can make these forms of distributing information relatively expensive, particularly if the information must be updated on a frequent basis.

3.5 Is remote online access to your agency databases a more effective approach than the exchange of data in machine-readable form? Under what conditions is remote access preferable? Is your agency prepared to provide in hardcopy or microform the same information provided in machine-readable format? How often can this information be updated?

4 User-friendliness

4.1 Within a computerized information system, the use of coded information for certain data elements is important. Codes are quick to enter, and simple to validate when data is being collected. They are often language-independent, facilitating the exchange of information between organizations using different languages. However they may also be difficult for the end user of the information, since their meaning is not always obvious.

4.2 One method of overcoming this disadvantage is to have the computer translate a code into a more complete and clear description. This translation could take place at various points within the information processing and distribution chain: when the information is being collected; when information is sent to other agencies; when information is received by the other agencies; or when the information is being accessed by the final user. Regardless of when the translation takes place, the end result is that the user can search and display a full, descriptive form corresponding to the coded value.

4.3 *What provisions already exist within your organization for the translation of coded values? What provisions would you like to see in place in your organization or in other organizations?*