Arunachalam, V.
Documentation System and its Utilization in
Indian Agriculture.
Indian Journal of Plant Genetic Resources, 1
(1989): 89-97.

Documentation System and its Utilisation in Indian Agriculture

V. ARUNACHALAM

Indian Agricultural Research Institute, New Delhi

The need for developing an efficient Communication-Information-Documentation-System to suit Indian agriculture is keenly felt in the context of world germplasm of several crops that have been assembled, maintained and utilised in India. The documentation systems available in various countries are reviewed and it is suggested that the EXIR information retrieval programme could be adopted with modifications. An organisational network with establishment of Field Stations, Regional Centres and Genetic Resource Centres providing a link with National Bureau of Plant Genetic Resources (NBPGR), New Delhi is suggested to initiate documentation and allied activities like maintenance and distribution of germplasm in India. NBPGR can take the final responsibility of collating and executing the activities related to documentation. The spheres of research and information dissemination that would get catalysed as a consequence, are briefly described.

Documentation is the general term for the processes and activities for collecting, organising, storing, retrieving and disseminating recorded information (Weisman, 1972). During the recent past, several documentation systems have been developed to describe crop germplasm. In India, world germplasm of several crops has been assembled, maintained and utilised during the last two decades. A good deal of information on morphological, physiological, biochemical and biometrical characters including reaction to various pests and diseases has been collected on the germplasm of different crops. Yet, an organised system of documentation of this valuable information to meet the specific needs of a scientist remains to be developed.

With the establishment of a number of All India Coordinated Crop Improvement Projects for monitoring the breeding and extension activities at the national level, valuable data on crop germplasm and breeding researches are piling up over time. More and more genetic stocks are being added to strengthen the breeding base of crops and thus, increasing and diversifying the data further. Moreover, to monitor the activity of germplasm introduction and utilisation, the National Bureau of Plant Genetic Resources (NBPGR) was set up in 1976 in India. It is high time that we think of developing an efficient Communication-Information-Documentation-System (CIDS) to particularly suit the needs of Indian agriculture.
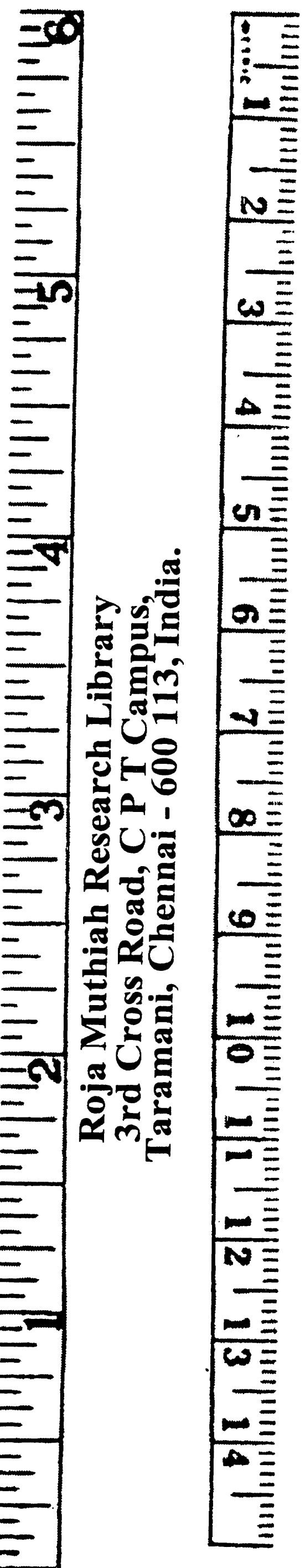
REVIEW OF AVAILABLE DOCUMENTATION SYSTEMS

Most of the documentation activities revolve around centres engaged in maintaining and utilising crop genetic resources. Such centres were named Genetic Resources Centres (GRC). Their major functions include exploration, quarantine, conservation, utilisation and documentation (Rogers et al., 1975). Each of these functions can be partitioned into their components which can integrate into a dynamic and continuous activity of a GRC (Fig. 1).

The enormous data that would flow into a GRC would obviously require a powerful computer for efficient documentation. In fact, almost any type of relevant information on the germplasm, can be documented. It is then necessary to develop a list of descriptors to retrieve such information when needed. A few examples (Rogers, 1975) would help to have a clear understanding.

1. Reference : Accession numbers, collection information (collector's name, number, date), nomenclature information (scientific and common names), origin (country, state and locality), storage information (location, date of storage, etc.) and distribution information.
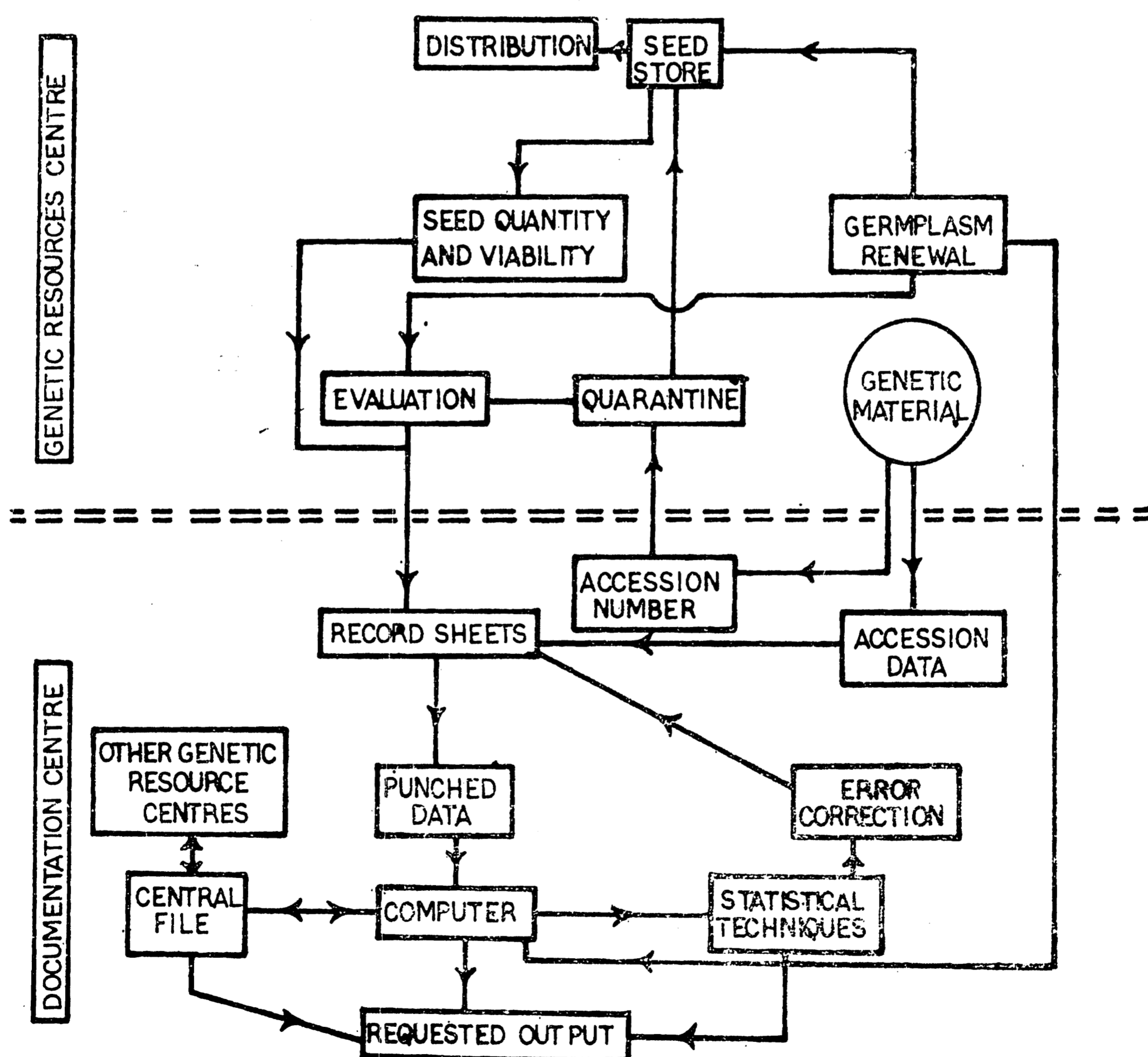

Fig. 1. Functions of Documentation and Genetic Resources Centre (after Rogers et al., 1975).

2. Environmental : altitude, climatic and ecological.

3. Organismic : morphological, genetical, physiological, biochemical, pest and disease resistance, rejuvenation (seed viability, storage time), tolerance to environmental stress (cold, lodging, drought, salinity, etc.).

4. Utility : breeding, genetics, agronomic performance, etc.

5. Ethnobotanical.

6. Consumer characteristics.

7. Industrial attributes.

8. Bibliographic.

Such diverse characters can easily be recorded in terms of scores. For many crops and characters, it was found adequate to employ a system of scoring from 1 (very low) to 9 (very high) indicating the magnitude and rate of gradation. The mean score 5 would enable one to easily identify high score (greater than 5) and low score (less than 5) categories (Matsuo, 1975). There is an obvious advantage of this system. Biometrical characters being continuous, may not maintain the same absolute values for a genotype over time and space. The scoring system would compensate for such inherent changes in the values of biometrical characters of a genotype and make information retrieval simple, easy and efficient. Where punch cards are the input medium for the computer, it would be possible to include information on a large number of characters in one card (each column providing information on one character).

A start was made in India during 1963-68 when a world collection of 5144 genetic stocks of sorghum was evaluated for 28 characters at 9 locations. Similarly 1532 genetic stocks of pearl millet were evaluated for 23 characters at 8 locations (Murty et al., 1967a ; Murty et al., 1967b). The data were catalogued and stored in punch cards but were not entered in any documentation system. Collections are available in a number of other crops too at different institutes/centres. How far a national CIDS can be initiated, will be discussed later.

Several systems have been developed on various computers to document information. For example, a system for documentation of Pisum collections was developed at John Innes Institute, UK. In the Federal Republic of Germany, a system named COLEM was written for a Siemens Computer. A fairly recent but efficient system was provided by the Taximetrics Laboratory, University of Colorado, USA. It is known as TAXIR, an abbreviation for TAX onomic Information Retrieval, a powerful and flexible system, capable of modifications to provide global communication (Rogers et al., 1975). It has already been modified and generalised to become EXIR (EX ecutive Information Retrieval, Ford-Lloyd, 1977). This system is oriented towards user and research needs. Two major groups of users are (i) primary users who can communicate with the computer, update or modify the data, storage files and also retrieve any item of information, and (ii) secondary users who can only use known commands to retrieve information.

Efficient information should conform to certain norms. It should be accurate and consistent. A screening process should be embedded into the system to

cull out irrelevant, redundant and outdated information. More than quantity, quality should be emphasised. The quantity should be commensurate with the user needs, both present and anticipated (Weisman, 1972). Flexibility of the documentation system should be ensured to make deletion and updating of information possible.

The retrieval process of such well-documented and stored information is executed by a user-oriented computer programme such as TAXIR, for example. TAXIR uses a systems approach for documentation and is flexible to take in dynamic changes that occur in its data profile. A thesaurus is provided to the user to enable the framing of queries for information retrieval and to communicate with the computer, in general. All descriptor states can be referred to using this thesaurus. A process of rapid search of the required information is set up using Boolean algebra approach as would be illustrated by an example later. TAXIR programme can also accept a pre-defined dictionary by the user. The query mechanism also makes corrections. For example, suppose the dictionary for country of collection is as follows : Algeria, Belgium, France, Franxe, Spain, Espana. The incorrect spelling of Franxe and duplication of Spain and Espana can be corrected by two statements :

1. CORRECTION COUNTRY OF Collection, SPAIN for all ITEMS with country of collection, ESPANA*

2. CORRECTION COUNTRY OF Collection, FRANCE for All ITEMS with country of collection, FRANXE (* is the code for end of a query or statement).

Then these corrections are carried out internally but they are not made on the original input material. In other words, recompiling original material will regenerate the mistakes. This facility is a notable feature of TAXIR and can be used with advantage to prevent permanent loss of information and also in translation from one language to another (Hersh and Rogers, 1975).

Let us now illustrate information retrieval from the data in Table 1 stored in the computer. Suppose the following query is asked of the system.

TABLE 1   EXAMPLE OF DATA ON FOUR DESCRIPTORS STORED IN A
DOCUMENTATION SYSTEM

ID No.	FC	LV	GL	MF
100	Blue	Palmate	Vermont	April
102	Blue	Pinnate	Delaware	August
120	Red	Palmate	Maryland	May
103	Violet	Palmate	Maryland	June
114	Red	Palmate	Maryland	July
106	Yellow	Pinnate	Virginia	May

FC = Flower colour ; LV = Leaf venation ; GL = Geographic location ; MF = Month of flowering.

Query—List items with (FC, Red OR Pink OR LV, pinnate) AND MF, from April to June AND GL, Maryland*

The method of search consists in splitting the query into a number of simple components called operands. The values of the operands are entered against each stock (with an ID No.), 1 representing positive and O, negative response (Table 2). The query is then reframed using the operands and their logical connections. Boolean OR and AND operations will fix a final score 1 or 0 against each stock. Those with a score 1 are the stocks answering the query (Table 2). This method being a quantitative approach is perhaps the fastest of the search methods. It can hence be used to retrieve information answering a large number of queries accurately within a short-time (Estabrook and Brill, 1969).

TABLE 2. EXECUTION OF BOOLEAN SEARCH ANSWER IN TAXIR

Operand	Characteristic function					
	100	102	120	103	114	106
A. FC, Red *OR* Pink	0	0	1	0	1	0
B. LV, Pinnate	0	1	0	0	0	1
C. MF, from April to June	1	0	1	1	0	1
D. GL, Maryland	0	0	1	1	1	0
Query : A *OR* B *AND* C *AND* D						
Execution						
A *OR* B	0	1	1	0	1	1
A *OR* B *AND* C	0	0	1	0	0	1
A *OR* B *AND* C *AND* D	0	0	1	0	0	0
Answer : ID 120						

CIDS IN INDIAN AGRICULTURE

We now come back to our earlier question—are we at a stage to initiate a national CIDS and will the data on germplasm we talked about earlier be suitable and adequate for such a purpose ?

A moment's reflection on the use of documentation and retrieval makes it apparent that the genetic stocks whose data have been stored, should have been maintained over time to retain the genetic identity of the collections. Apart from ICRISAT, it is difficult to judge how many of places these germplasm entries have been maintained in India. Moreover, the data collected in the distant past on these entries cannot serve to initiate documentation. It is then clear that though we need a national CIDS, the necessary framework has to be set up before attempting such an activity.

Organisation

(i) As a first step, small groups of workers should meet and finalise the list of characters that need to be critically assessed and documented. Such groups

should be set up for each crop. Based on their report, it is then necessary to standardise the format for documenting the information on punch cards. The record system developed by Japanese group of IBP can serve as a model (Matsuo, 1972).

(ii) The record system should be flexible so as to incorporate the dynamic changes that occur in the germplasm over time. Global compatibility of the system should be given due weightage.

(iii) It is then necessary to identify an organisational network to feed accurate and consistent data to NBPGR, who will be finally responsible for CIDS. A possible network (Fig. 2) would require establishment of field stations (FS) at the fundamental level to feed data to Regional Centres (RC) which, in turn, will communicate them on to GRC. The field stations would function under the control of RC. A network of three GRC connected to RC and FS will, in general, be adequate to take care of the major crops for which germplasm is available. The network can obviously be expanded to suit various needs.

(iv) The FS, RC and GRC should be adequately staffed. They can only then devote their undivided attention to germplasm maintenance, evaluation and documentation. This group of workers should be considered on par with those on breeding, genetics, agronomy and other disciplines. Enthusiasm among the staff of FS alone can provide accurate data to GRC and hence to CIDS.

RC = REGIONAL CENTRE,
FS = FIELD STATION , GRC = GENETIC RESOURCES CENTRE,
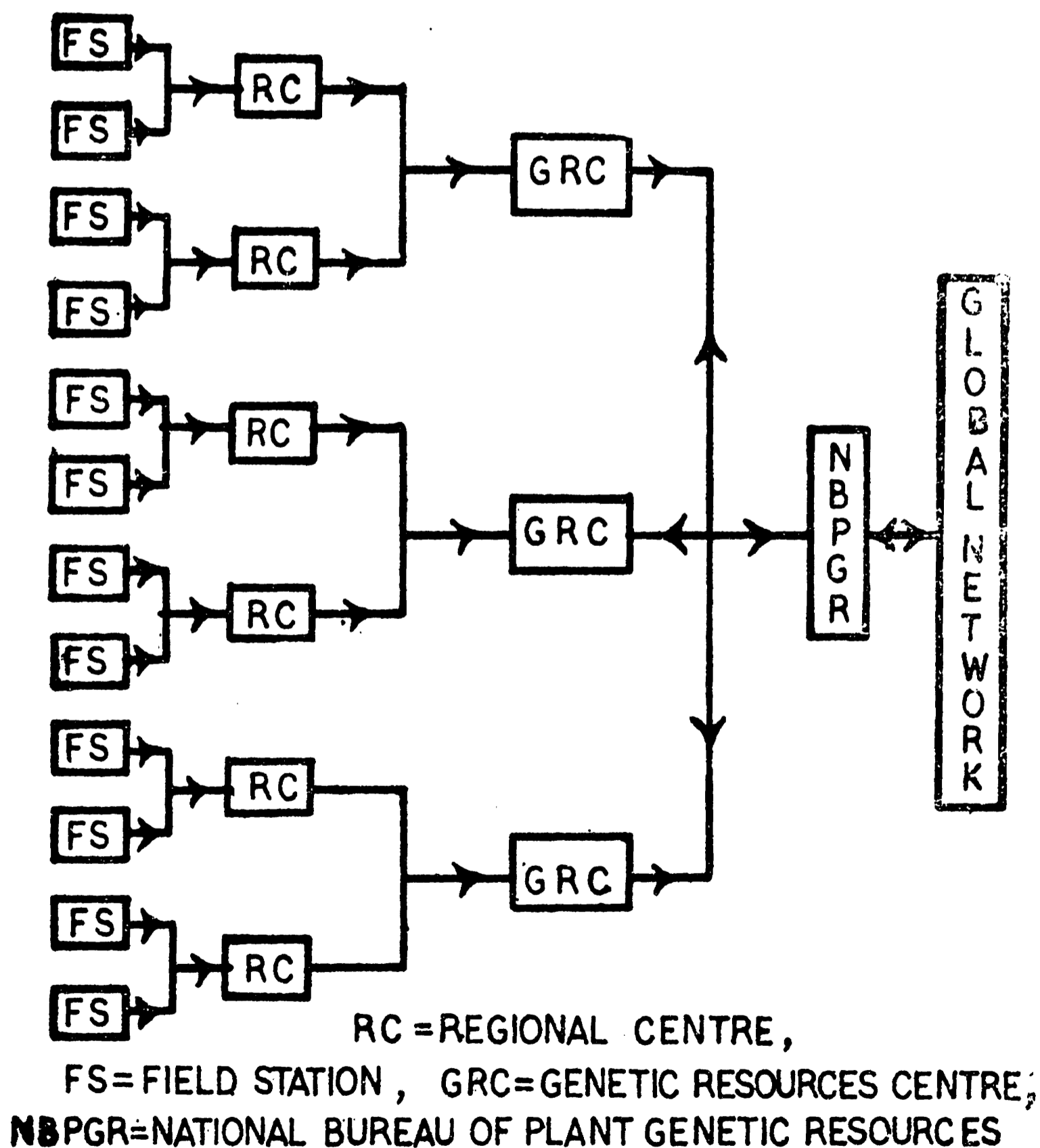NBPGR = NATIONAL BUREAU OF PLANT GENETIC RESOURCES

Fig. 2. A Possible Organisation of Communication-Information-Documentation-System.

(v) A third-generation Burroughs Computer is functioning at the Indian Agricultural Statistics Research Institute, New Delhi. This is a multiple-access time sharing system capable of taking remote terminals. However, direct communication of data of RC and GRC with this computer system is not immediately feasible, but can be a later possibility when satellite and earth stations are commissioned. Until then, punch cards can favourably be used as an input medium to the computer and data on all crops can be stored on magnetic tapes for storage and retrieval.

(vi) Since TAXIR and its modified varsion EXIR have been commended to have the potentiality for use in the global network (Hersh and Rogers, 1975; Ford-Lloyd, 1977), it is worthwhile to study this system and adopt it with modifications to suit the national programme.

Utilisation

The major utility of CIDS is to serve user needs of a wide variety of disciplines. However, CIDS can profitably be utilised to tackle a number of problems, some of which will be discussed below :

(i) It is an acknowledged fact that a good number of duplicates perpetuate in the germplasm accessions, necessarily escalating the cost of maintenance. For example, it has been conjectured that a barley collection of just 12,000 probably embraces a majority of the world's barley genes (Creech and Reitz, 1971). Thus selective additions to germplasm is more useful than more items added at random. A sound documentation system can be used to eliminate duplicates, whether they recur due to manual error or due to genetic causes. A method of matching genetic entries for a number of attributes used to characterise them using multivariate analysis or other techniques, should be developed to identify and cull out genetic duplicates.

(ii) The documentation process is the right answer for updating information on germplasm as and when genetic changes occur in the process of adaptation. A number of examples can be cited where a genetic stock acquired for a single favourable attribute was found to possess a number of them after a few years of its adaptation to a particular environment. PI 178383, a wheat collection from Turkey by Jack R. Harlan, introduced for its resistance to stinking smut in 1948, was later identified as a source of resistance to 4 races of stripe rust (*Puccinia striiformis*), 35 races of common bunt (*Tilletia caries* and *T. foetida*) and 10 races of dwarf bunt (*T. controversa*), all of which occur in Pacific Northwest. It had also unusual tolerance to flag smut (*Urocystis tritici*) and snow mold (*Fusarium* and *Typhula* sp., Creech and Reitz, 1971). Similarly, it is also not uncommon to find that a genetic stock loses its favourable attributes over time in the process of its adaptation to an alien environment. Evaluation of information on genetic stocks over time is easy with a documentation system and will enable one to decide the course of action regarding the maintenance of such stocks.

(iii) Another aspect of germplasm collection is the retention of genetic distinction. As and when genetic stocks are entered into the collection, it is necessary to

use methods of matching an array of attributes over time in order to decide whether the stock deserves maintenance as a separate entry. Such of those entries which show genetic similarity over time, can be pooled and maintained as germplasm complexes rather than as individual entries.

(iv) One of the potential uses of documentation is to discover the distribution of specific genes, like those of quality, disease and pest resistance and rare morphological attributes over space and time. Such information will be of direct applied value particularly in view of the fact that such specific genetic stocks are maintained and can be supplied to interested scientists.

(v) It is known that several desirable genes are located in local germplasm. For instance, sorghum shootfly resistance is found in local genetic stocks from Andhra Pradesh (Maldandi), Gujarat (BP-53), Maharashtra (Karad Local) and some others. Some entries from Karnataka (Allu Jola and Dagadi) have very compact earheads. A search for such gene sanctuaries, both national and inter. national, will be feasible with a good documentation system.

(vi) The maintenance of germplasm over time, especially of cross-pollinated crops, such that they retain their original variability is a major subject of research by itself. However, when records of several attributes of genetic entries are stored over years, it is possible to evaluate them over time. Such evaluation will help to forewarn the maintenance unit about the deteriorating variability status so that appropriate remedial measures can be taken in time to prevent genetic erosion.

(vii) An important and pertinent topic of CIDS relates to the optimum number of genetic stocks that need to be maintained in every crop. Earlier reports (Creech and Reitz, 1971) indicated that real variability with regard to identifiable gene effects did not probably need maintenance of more than 12,000 items in barley and similar number in other crops. It is, of course, not possible to obtain the optimum number for each crop by theoretical means. Yet a norm should be set, say of the order of 20,000 beyond which there should be a note of alarm for adding on to the germplasm. If we consider the number of genetic entries that is drawn upon for breeding and other purposes in any crop, one has to face the hard fact that a really low proportion (less than 1%, in general) has been utilised. The utility index can be improved by making available readily retrievable and well-documented information on germplasm backed up by a well-organised set-up of supply of pure seeds. A great deal remains to be researched upon and organised in these areas.

(viii) Of salient theoretical interest is a number of population genetic studies that can be planned from CIDS. Evolutionary pattern and genetics of crop plants, mapping of linkages of specific gene loci and discovering new genetic relationships at intra- and inter-specific level are a few areas for exploration.

(ix) Finally, a sound CIDS can help to stimulate many other areas of research. Results on well-planned breeding and agronomic experiments can be documented along with detailed records on meteorological parameters. Evaluation over time can help to trace crop-weather systems and make it possible to formulate optimum

cropping systems to suit specific weather conditions. Weather-pathogen relationships (in preference to host-pathogen) can be used to identify new ecological races of pathogen. Based on an analysis of performance over time, it is possible to retrieve information on resistant genotypes for specific areas and ecological conditions.

REFERENCES

Creech, J. L. and L. P. Reitz. 1971. Plant germplasm now and for tomorrow. Adv. Agron. 23 : 1-49.

Estabrook, G. F, and R. C. Brill. 1969. The theory of the TAXIR accessioner. Math. Biosci. 5 : 327-340.

Ford-Lloyd, B. V. 1977. Data storage and retrieval systems in genetic resources information exchange. Proc. 139th Annual Meeting of the British Association for the Advancement of Science.

Hersh, G. N. and D. J. Rogers. 1975. Documentation and information requirements for genetic resources application. In O. H. Frankel and J. G. Hawkes, eds., Crop Genetic Resources for Today and Tomorrow, p 407-448. Cambridge University Press, Cambridge.

Matsuo, T. 1975. Gene Conservation—Exploration, Collection, Preservation and Utilisation of Genetic Resources. JIBP Synthesis, Vol. 5, University of Tokyo Press, Tokyo.

Murthy, B. R., V. Arunachalam and M. B. L. Saxena. 1967a. Classification and catalogue of a world collection of sorghum. Indian J. Genet. 27A, : 1-312.

Murthy, B. R., M. K. Upadhyay and P. L. Manchanda. 1967b. Classification and cataloguing of a world collection of genetic stocks of Pennisetum. Indian J. Genet. 27A : 313-394.

Rogers, D. J., B. Snoad and L. Seidewitz. 1975. Documentation for genetic resources centres. In O. H. Frankel and J. G. Hawkes, eds., Crop Genetic Resources for Today and Tomorrow. p 399-406. Cambridge University Press, Cambridge.

Weisman, H. M. 1972. Information Systems, Services and Centres. Becker and Hayes Inc., New York.