



Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines

We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content

The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute

The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

AGPE: 1BPER/74/4C

Caps Crop Ecology & Genetic Resources Unit
Caps Plant Prod. & Prot Division

4 spaces

Headings of
paper in caps underlined

4 spaces

Start

TRAINING COURSES
IN GENETIC RESOURCES

Draft of base paper for 1974 Rome panel meeting - J.G. Hawkes

Training courses may be envisaged at five different levels, depending on:-

- a) amount of previous training
- b) level of training desired in the course
- c) ends in view, e.g. practical, theoretical or both.

The suggested levels and types of courses are as follows:-

1. Short Practical Courses (4 - 6 weeks)

These would lay great stress on standard practical techniques. No theoretical background would be necessary.

- a) Aims (i) Training in gene-bank management, e.g. documentation; seed storage, germination tests and rejuvenation; data management at a technical level.
ii) Training in field collecting techniques, e.g. practical sampling methods; herbarium, seed, etc. collecting; documentation of provenance data; photography, map reading and travel planning; contacts with farmers etc.
- b) Participants (i) Technicians with a good general education who are needed as trained assistants for the scientists.
ii) Trained scientists who need supplementary practical gene bank and field experience.
- c) Training Centres Seed banks and genetic resources centres, related to the crops for which training is to be provided and where trained personnel is available for giving practical instructions.

2. Medium-Length Courses (+ 3 months)

- a) Aims These would contain the practical training elements of course No 1, but would include some theoretical studies, e.g. taxonomy of crops collected; use of identification keys and descriptions; genetical basis of variation; plant breeding in relation to genetic resources;
- b) Participants The course would provide training for higher level technicians, those already holding diplomas in agriculture or allied subjects, to enable them to engage in collecting activities and gene bank management where fully-trained, scientific personnel are not available. It would be hoped that those showing promise on completing this type of course would move on to course No.3 later on.
- c) Training Centres As for course No. 1, making sure that trained teaching personnel in the practical and theoretical aspects of the course is available.

3. Masters (M.Sc.) Courses (1 - 2 years)

These would comprise lecture courses, practical laboratory work and a research project, followed by a dissertation.

- a) Aims To provide a complete training in the theory and practice of the exploration, conservation, evaluation and utilization of plant genetic resources (see prospectus of Birmingham M.Sc. course, now in its 5th successive year). The course could last for one year, as in Birmingham, or could extend for two years if desired. Some research experience is considered highly desirable.
- b) Participants This type of course is available for those already holding a first degree (or equivalent) in genetics, botany, agriculture etc.
- c) Training Centres These courses must be carried out in Universities, possibly linked to a short period of pre or post-course training at a gene bank.

4. Ph. D. Training Course (3 - 4 years)

These would be of the usual type, based on a taxonomic, genetical or physiological research problem related to some aspect of genetic resources work. They would either follow on from and be integrated with M. Sc. studies, as in No. 3 above, or would contain course work on genetic resources themes (U.S. system).

a) Aims To provide a higher level research-based training for students who would expect to be involved in due course in top-level posts at genetic resources centres.

b) Participants Students of excellent calibre who had taken a first degree (B.Sc., etc.) or a Master's degree under No 3 above.

c) Training Centres These courses must be carried out in Universities, as for No. 3 above.

5. High Level Refresher Courses (1 - 3 weeks)

These could be arranged from time to time either at Universities or at Crop Research Institutes where germplasm and related plant breeding studies are in progress. The recent Workshop held at the N.I. Vavilov Institute of Plant Industry is an excellent example of this kind of course organized jointly by the Institute and F.A.O.

a) Aims To act as a stimulus and a discussion forum to those already trained in genetic resources management and research.

b) Participants Those already holding M.Sc or Ph.D. degrees in this or related fields.

c) Training Centres It is envisaged that such courses would be held mainly in old-established centres where a reasonably large group of trained personnel would be available. Distinguished scientists from other countries might also be invited to participate (as with the Leningrad Workshop)

4.3 EXPLORATION: AFRICA

Africa is a vast continent, rich in genetic resources of several major crops and a number of minor ones. To assemble a reasonably adequate sampling of landrace and wild populations is a formidable task and will require well-financed and systematic explorations over a period of several years. A two-year plan can only be a beginning. I do not consider it feasible to attempt to saturate the continent with a crash programme because a) there is not that much expertise available and b) such genetic resources centres as there may be are not yet prepared for massive new acquisitions. Many of the nations do not have adequate personnel and resources to conduct exploration and collection locally, and expeditions from outside will be required. Some countries may not be politically prepared to cooperate in such enterprises at the present time. No matter how well organized and systematic the explorations may be, they will not be fruitful unless genetic resources centers are available to receive and maintain the materials collected.

The current situation with respect to genetic resources of several crops was presented for Africa in "Survey of Crop Genetic Resources in their Centres of Diversity: First Report" FAO/IBP D. H. Frankel (Ed.), February, 1973. Since these reports were prepared we have received information that the International Institute for Tropical Agriculture, Ibadan, Nigeria, has become active in plant exploration with personnel, vehicles, and resources allocated for the purpose. We also have encouraging news of financial

support for the establishment of a genetic resources centre in Ethiopia, and the International Crops Research Institute for the Semi Arid Tropics is moving rapidly toward an operational status. Any plan of operation for Africa must take these developments into consideration.

NORTH AFRICA: This region comes largely under the purview of the Mediterranean interests (see section 4.1). The most urgent concern at the moment is the replacement of landrace populations of wheat, barley and other crops by various forms of modern semi-dwarf wheats. It may already be too late, but it is thought that some of the older materials may still be found in the Atlas Mountains of Morocco and Algeria and/or in the more isolated oases of the Sahara. Other Mediterranean crops, such as chickpea, lentil, flax, rye, oats, and forage plants should be collected at the same time, but salvage of the remaining wheats is a most urgent operation.

WEST AFRICA: In view of the current activity by IITA in assembling materials from Nigeria and nearby countries, the logical regions to collect in the immediate future should centre on Mali, Guinea, and Liberia. These countries have extensive uncollected resources of sorghum, pearl millet, rice, cowpea, yams, oil palm, Voandzeia, and coffee. Collection activity should be extended to Senegal, Sierra Leone, Ivory Coast and Ghana as soon as it becomes feasible to do so.

Local resources are minimal in the main target area. The French agricultural research organization, IRAT, has an office in Mali and has supported work on sorghum and rice in the past, but at present could

contribute little more than contact with appropriate Malian personnel. To obtain a reasonable sampling of sorghum, pearl millet, rice, and cowpeas would require at least two teams operating a minimum of two seasons in Mali alone. Technical assistance can probably be obtained through cooperation with IRAT stations in Senegal and Upper Volta, where experts for each of these crops are in residence. Preliminary conversations with the Ministry of Agriculture suggest the government of Mali would cooperate fully, but that a formal agreement might be required.

As far as I am aware, there is little agricultural research being conducted in Guinea. A special agreement between the government of Guinea and FAO might be required before an expedition could be conducted.

The governments of Liberia, Sierra Leone, Ivory Coast, and Ghana have shown every indication that they would cooperate fully with any legitimate efforts at assembly of genetic resources. Ghana has, in fact, a plant introduction officer, Mr. M. A. Adamsi, who has experience in collecting and would need only some additional resources and encouragement to sample the materials of Ghana. Transportation is usually the principal requirement. Senegal has both expertise and resources for collecting, but not for the maintenance of large collections.

PROPOSAL: Two landrovers or equivalent
Two specialists for collecting
Two local personnel to assist and facilitate.

The two teams could start in the north near the Sahara and work southward

across Mali and through Guinea to the coast if possible. Repeat the second year by different routes. With hard work, diligence and good luck a great deal could be accomplished in two years.

EAST AFRICA: Assuming that work shall get underway in Ethiopia shortly, the next most critical areas would be Tanzania, Zambia and Malawi south of the equator and southern Sudan and Somalia north of the equator. Because of the reversal in seasons across the equator it would be feasible for collecting teams to work both of these areas in the same year, passing through Kenya and Uganda en route. All of these countries have some agricultural research stations and there are a few FAO, US/AID and other experts that could be of assistance.

PROPOSAL: Two landrovers or equivalent
Two specialists
Local assistance to make two teams
Two years each north and south of equator

The major resources of concern are sorghum, pearl millet, cowpea, Voandzeia, forage grasses, finger millet, and wild rice. There are a number of minor crops to be collected as well.

TRANSPORTATION COSTS:

4 vehicles @ 5,000	20,000
operation first year 2,000 each	8,000
operation second year 3,000 each	12,000

SEED CONSERVATION

by Dr. Lennart Kåhre, Swedish Seed Testing and Certification Institute, S-171 73 Solna, Sweden.

The science of seed longevity

A seed embryo can be defined as a plant in a resting stage. This kind of latent life has since long stimulated and fascinated nature scientists all over the world.

At the end of the last century the botanists found that the seeds are respirating during storage. They started to study "the life span of seeds". Becquerel (1907) published one of the first reports in this connection, dealing with 500 species of 30 families conserved in Paris. The seed age varied between 25-135 years. He found the leguminous seed species to be long-lived e.g. Trifolium arvense L. had living seeds after 68 years (2 of 10 investigated). Another classic was Ewart (1908), who studied the viability of 600 seed species maintained in Melbourne. In addition to his own material he reviewed data obtained by other scientists, so totally he discussed about 1 400 species. He classified the seed species as to their longevity in the following three categories:

- a. microbiotic (seed longevity less than 3 years)
- b. meso- " " (" " 3-15 years)
- c. macro- " (" " 15- more than 100 years)

Most of the species were classed as microbiotic, but Becquerel's observation as to some seed species of Leguminosae being macrobiotic was confirmed. In addition Ewart mentioned some species of Malvaceae and Myrtaceae to be long-lived.

The storage conditions should be "optimal" for Ewart's classification to be valid, but 40 years later Orocker (1948) still questioned if we know for a single species what are optimal storage conditions. He raised that question after many years' progressive seed research in the United States of America. And in our time in the introduction to the book "Viability of Seeds" Prof. E H Roberts (1972) states: "... there are still contradictions, errors and differences of opinion, and the subject is still incomplete and fragmentary. But progress has been made: more fragments exist and some of them begin to tell a story". We are now able to prolong the life of seeds more than earlier was considered possible.

Factors influencing the storability of seeds

Some important factors regulate the seed longevity.

1. Anatomical structure
2. Maturaty stage
3. Mechanical damages
4. Moisture (relative humidity) - temperature and packages keep resp. down
5. Microflora, insects and mites

survived environment
antenna - closed pkg
The seed structure is the reason behind the longevity of some leguminous seeds mentioned above and reported in many other investigations. The "hard" seeds are well-known in *closed pkg of nature*
matured seed has a better physiological back-ground for storage over a longer period than unripe seeds. That was *keep resp. down New Zealand*

Brassica over four years storage. At the end 92 per cent of the full-ripe seeds were alive, but only 43 per cent of the unripe seeds. In red clover the corresponding figures were 58 and 6 per cent respectively.

- > Mechanical damages can destroy even the best seed (e.g. Esbo 1959, Moore 1972).

The moisture-temperature relationships are of outmost importance for the seed conservation. Also the type of seed is again of influence. Starby seeds contain more moisture than e.g. oily seeds. Harrington (e.g. 1973) has formulated two "rules-of-thumb" to explain the effect of moisture and temperature on seed longevity:

1. For each decrease of 1 per cent in seed moisture content the life of the seed is doubled (between 4 and 14 per cent moisture).
2. For each decrease of 5°C in seed temperature the life of the seed is doubled (at least between 0°C and 50°C).

It is obvious that the drier the seed the longer it will remain viable under warmer conditions.

Based on numerous experiments the Dutch grain specialist Kreyger (1972) made a classification of different seed species with respect to their loss of germination during storage and considering different moisture-temperature sets. His five categories are:

1. Very strong seeds

Rape, linseed, rough stalked meadow grass, barley

2. Rather strong seeds

Oats, beetseed, perennial ryegrass, field peas, green peas, creeping bentgrass

3. Less strong seeds

Brown beans, dwarf French beans, runner beans, wheat

4. Rather weak seeds

Red fescue, maize

5. Very weak seeds

Rye, onion

He has made diagrams on the relationships between the estimated maximal duration of safe storage and temperature for various moisture contents (expressed as relative humidities of the equilibrium atmosphere). Fig. 8-12 are taken from his recent publication: "Drying and Storing Grains, Seeds and Pulses in Temperate Climates". It should be noted from these diagrams that the absolute moisture content of the air varies with the temperature (lower temperature means less and higher temperature more absolute moisture).

Harrington (1973) means that the equilibrium moisture content may vary \pm 2 per cent due to variations not only in storage temperature but also in chemical composition of different seed lots and hysteresis effects.

humidities of the equilibrium atmosphere into seed moisture contents Kreyger has presented some tables (3-5). These data indicate that it is possible to keep seed viability for 20 years or more at a temperature of 0 to +5°C and a relative humidity of 40-60 per cent in case of very strong seeds i.e. 6-7 per cent moisture content in rape, 9-11 per cent in meadowgrass and 10-13 per cent in barley. In rather strong seed the relative humidity must be 40-50 per cent i.e. 9-11 per cent moisture content in oats and ryegrass, 8-9 per cent in beet and bentgrass and 10-11 per cent in peas. The less strong seeds cannot be kept above 40 per cent humidity and not above $\pm 0^{\circ}\text{C}$ to be alive after 20 years. That means 9 per cent moisture content in beans and about 10 per cent in wheat. The rather weak (e.g. red fescue and maize) and the very weak seeds (e.g. rye and onion) cannot survive for such a long period, except at a sub-zero temperature and at 30-40 per cent relative humidity.

Dr. L N Bass (1973), who is head of the National Seed Storage Laboratory in Fort Collins, Colorado, USA, has recently made a survey on seed storage in "Seed Science and Technology", the new publication from the International Seed Testing Association (ISTA). That covers 104 literature references. In connection with the more "difficult" seed species he summarizes that high viability of low-moisture content vegetable, flower and tree seeds was maintained during 20 years of sealed storage at -4°C . Onion seeds showed no loss of viability in up to 18 years. *free*

There are many choices in seed packages. Grabe & Isely (1969) have concluded results of an intensive long-term (6 years) seed packaging experiment, which originally was designed and initiated by Bass at Iowa State University. They found that cans were the only container that were completely moisture proof. The other packaging materials ranked in the following descending order: Polyester-polyethylene, cellophane, kraft paper and cloth.

seed storage

Harrington (1973) and others have found it absolutely essential to package only dry seed in moisture-proof cans. He recommends 6-12 per cent moisture content for starchy seeds and 4-9 per cent for oily seeds in this connection. The recommended type of packages include e.g. tin cans, glass jars with gasketed screw top lids and steel drums with clamp-down gasketed lids. Harrington & Douglas (1970) have described the procedures to be applied in more detail in chapter 22 in their text-book "Seed storage and packaging - applications for India". They point out the possibility to obtain and maintain a low seed moisture content by using silica gel and sealed containers instead of dehumidifying the whole storage. This procedure has also been used e.g. in Sweden where red clover seed samples are conserved since 1955 at about $+5^{\circ}\text{C}$ and by means of a desiccant a moisture content of about 7 per cent was reached very quickly. The germinability has been kept quite well in the past 20 years (Bingefors 1974).

the two 104/20 work

instead (humidify) to all in 2 p 13

From a practical point of view it is very important to clean and dry the seed well before it is stored. When drying the seed the seed temperature again is critical. The wet seed is damaged by high temperatures. There is clear evidence for that in e.g. Swedish experiments (Lindberg & Sörensson 1959 and Appelqvist 1965). When the seed moisture content is up against 35-40 per cent the seed temperature should not exceed 35°C.

Practical applications

The function of a gene-bank is two-fold. One has the long-term aspect (keep and not touch!). The other has the short-term intention (living exchange). So, in addition to the biological classification of species based on their seed storability there are more reasons to arrange more storage rooms in a gene-bank. In the book "Viability of Seeds" the seed storage laboratories in Fort Collins, USA (James 1972) and Hiratsuka, Japan (Itc 1972) are described more in detail.

It is important to establish storage facilities for a temperature range (something between + 5°C (standard) and - 20°C).

Fort Collins maintains about + 4°C with an average relative humidity of 32 per cent as standard. This temperature was selected as being suitable for most seeds. Three rooms are equipped for a temperature of - 12°C.

In Hiratsuka the storage for seed multiplication is kept under - 10°C and that for seed distribution at - 1°C. The seed moisture seems to be held at 4-6 per cent.

Völkensrode, Federal Republic of Germany has - 10°C and 6 per cent seed moisture, Bari, Italy - 1°C and 30 per cent relative humidity in the storage atmosphere and Izmir, Turkey ± 0°C and 6 per cent seed moisture.

Summary

In long-term storage of seed the most important factors are low temperature and low moisture content. Potts (1971) has characterized the life stairs of a seed. The aim of seed conservation is to keep the capabilities of the mature seed at a maximum for the longest possible time.

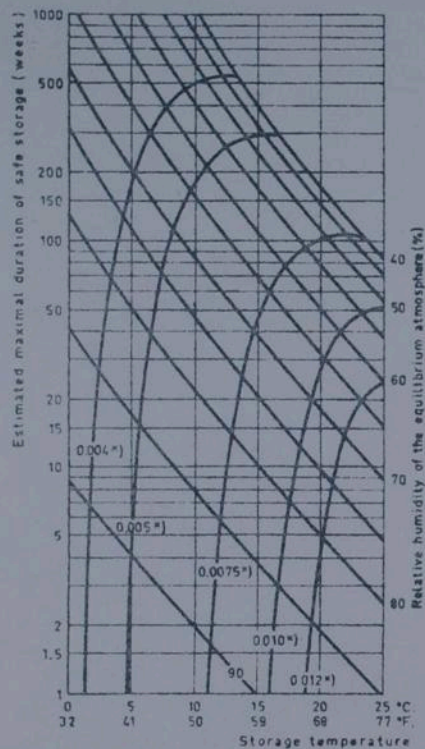
Literature references

- Appelqvist, L.-Å., 1965. Torkningsskador på oljeväxtfrö, ett kvalitetsfel som bör undvikas. - Svensk Frötidning, Årg. 34, Nr 9, 119-122.
- Bass, L.N., 1973. Controlled atmosphere and seed storage. - Seed Sci. & Technol., Vol. 1, No. 2, 463-492.
- Becquerel, P., 1907. Recherches sur la vie latente des graines. - Annales des Sciences naturelles. Sér. 9, Tome V, 193-311.
- Bingefors, S., 1974. Svenska lokalstammar av rödklöver 1 Sveriges Fröodlareförbunds stamkollektion. - Svensk Frötidning, Årg. 43, Nr 2, 24-26.
- Christensen, C.M., 1973. Boes of viability in storage microflora. - Seed Sci. & Technol., Vol. 1, No. 3, 547-562.

- Crocker, W., 1948. Growth of plants. Twenty years' research at Boyce Thompson Institute. - Reinhold Publ. Comp. New York, pp. 459.
- Esbo, H., 1959. Livskraften hos timotejfrö under långtidslagring. (With a summary in English: Vitality of timothy seed during long storage). - Växtodling 12, pp. 101.
- Ewart, A.J., 1908: On the longevity of seeds. - Proc. Roy. Soc. Victoria 21 (N.S.), 1-203.
- Grabe, D.F. & Isely, D., 1969. Seed Storage in moisture-resistant packages. - Seed World January 24, 2, 4-5.
- Harrington, J.F., 1973. Packaging seed for storage and shipment. - Seed Sci. & Technol., Vol. 1, No. 3, 701-709.
- Harrington, J.F. & Douglas, J.E., 1970. Seed storage and packaging-applications for India. - Nat. Seeds Corp. New Delhi, pp. 222.
- Howe, R.W., 1973. Loss of viability of seed in storage attributable to infestations of insects and mites. - Seed Science & Technol., Vol. 1, No. 3, 563-586.
- Hyde, E.O.C., 1954. The function of the hilum in some Papilionaceae in relation to the ripening of the seed and the permeability of the testa. - Ann. Botany (N.S.), Vol. 18, 241-256.
- Ito, H., 1972. Organisation of the National Seed Storage Laboratory for Genetic Resources in Japan. - In "Viability of seeds", 405-416.
- James, E., 1972. Organisation of the United States National Seed Storage Laboratory. - In Viability of Seeds, 397-404.
- Kreyger, J., 1972. Drying and storing grains, seeds and pulses in temperate climates. - Publ. 205, Inst. Stor. and Proc. Agric. Prod., pp. 333.
- Lindberg, J.E. & Sörensson, I., 1959. Upphettning- och torkningsförsök med spannmål företrädesvis vete. - Kungl. Skogs- o. Lantbr.akad. tidskr. suppl. 1, pp. 72.
- Moore, R.F., 1972. Effects of mechanical injuries on viability. - In "Viability of seeds", 94-113.
- Nobbe, F., 1876. Handbuch der Samenkunde. - Berlin, pp. 632.
- Potts, H.C., 1971. Seeds - development, structure and function. - Proc. 1971 Short course for seedsmen, Seed Technol. Lab., Miss. USA, Vol. 14, 37-51.
- Roberts, E.H., 1972. Viability of seeds. - Chapman and Hall Ltd., London, pp. 448.

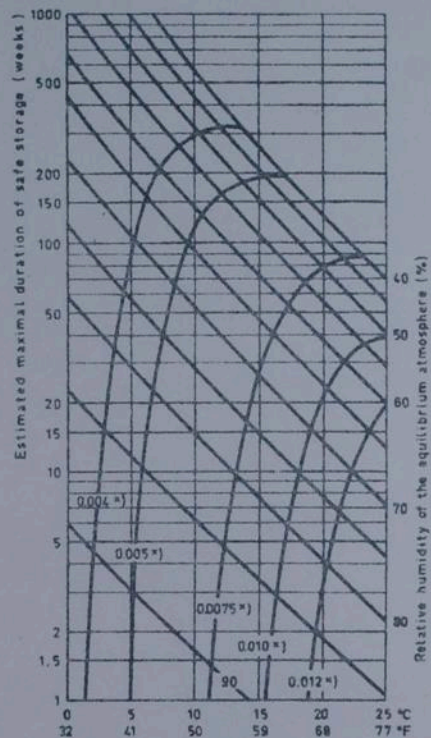
Organism of Gene Bank
A. Callan
o. Callan

-12
+3



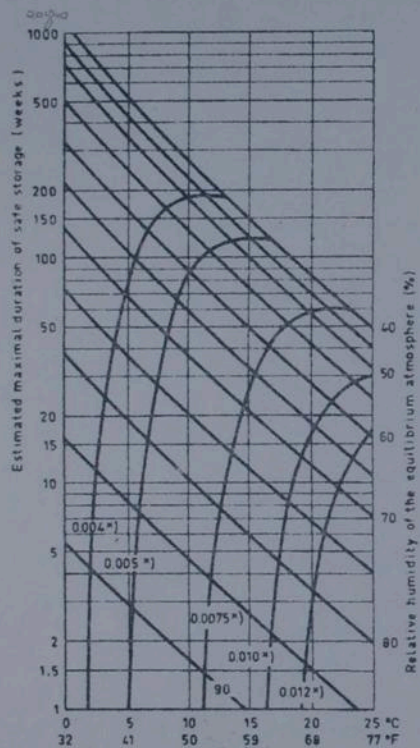
x) Curves indicating equal absolute moisture content of the air (kg/kg - lb/lb)

Fig 8 Relationship between estimated maximal duration of safe storage temperature for various moisture contents, expressed as relative humidities of the equilibrium atmosphere (very strong seeds, criterion germination, see Tables 9.3.4. and 5).



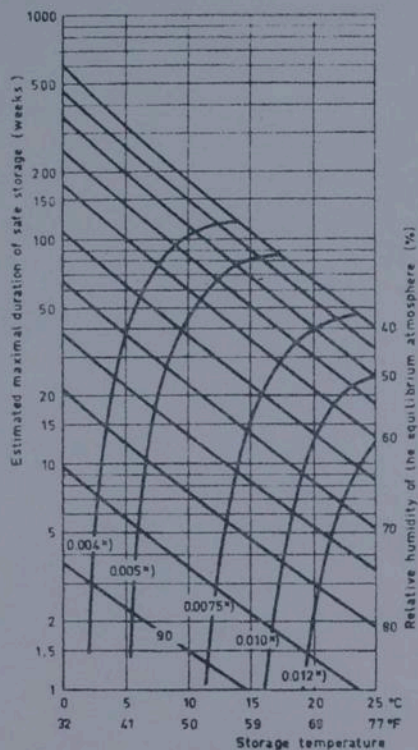
*) curves indicating equal absolute moisture content of the air (kg/kg - lb/lb)

Fig 9 Relationship between estimated maximal duration of safe storage and temperature for various moisture contents, expressed as relative humidities of the equilibrium atmosphere (rather strong seeds, criterion germination, see Tables 9.3.4 and 5)



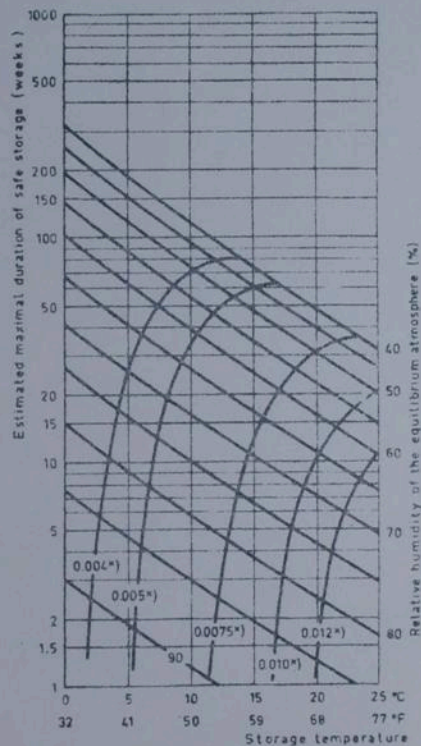
*] curves indicating equal absolute moisture content of the air (kg/kg - lb/lb)

Fig 10 Relationship between estimated maximal duration of safe storage and temperature for various moisture contents, expressed as relative humidities of the equilibrium atmosphere (less strong seeds, criterion germination, see Tables 9.3.4 and 5)



*] curves indicating equal absolute moisture content of the (kg/kg - lb/lb)

Fig 11 Relationship between estimated maximal duration of safe storage and temperature for various moisture contents, expressed as relative humidities of the equilibrium atmosphere (rather weak seeds, criterion germination, see Tables 9.3.4 and 5)



*] curves indicating equal absolute moisture content of the air (kg/kg - lb/lb)

Fig 12 Relationship between estimated maximal duration of safe storage and temperature for various moisture contents, expressed as relative humidities of the equilibrium atmosphere (very weak seeds criterion germination, see Tables 9.3.4 and 5)

TABLE 3. Moisture relations of 5 grains (lit. 13, 8, 5)

Relative humidity of the equilibrium atmosphere %	Equilibrium moisture content w (wet basis) %				
	wheat	barley	oats	rye	maize
92.5	25.5	27.5	25.2	28.5	22.5
90	23.0	22.7	22.0	24.4	20.7
87.5	21.0	20.7	20.0	22.1	19.2
85	19.6	19.4	18.6	20.3	18.0
82.5	18.4	18.3	17.5	19.0	17.2
80	17.6	17.2	16.7	17.9	16.5
77.5	17.1	16.6	16.0	17.0	16.0
75	16.6	16.1	15.4	16.2	15.4
72.5	16.1	15.6	14.8	15.4	15.0
70	15.6	15.0	14.3	14.9	14.6
67.5	15.1	14.5	13.8	14.4	14.1
65	14.7	14.1	13.3	14.0	13.8
62.5	14.2	13.6	13.0	13.6	13.4
60	13.7	13.2	12.5	13.3	13.1
55	12.8	12.3	11.7	12.4	12.4
50	12.0	11.4	11.0	11.7	11.9
45	11.3	10.6	10.2	10.9	11.2
40	10.7	9.8	9.5	10.1	10.7
35	9.9	9.0	8.7	9.3	10.0
30	9.1	8.3	8.0	8.6	9.3
			25 - 35°C		12 - 25°C
			77 - 95°F		53 - 77°F
			TEMPERATURE		

TABLE 4 : Moisture relations of 10 sorts of seeds (lit. 8)

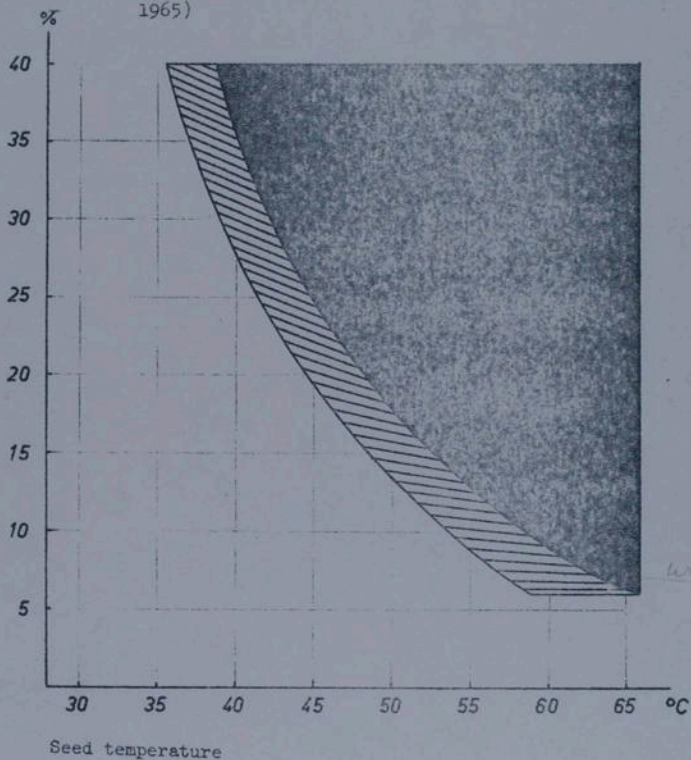
Relative humidity of the equilibrium atmosphere %	Equilibrium moisture content (w) (wet basis) %									
	Rape	Linseed	Caraway seed	Fodderbeetseed	Opium poppy seed	Onion seed	Perennial rye grass seed	Rough stalked meadow grass seed	Red fescue seed	Creeping-bent grass seed
92 ⁵	21.0	20.9	23.0	27.0	19.1	27.5	27.2	25.0	25.3	25.3
90	17.4	18.1	19.8	22.2	17.0	23.4	24.0	21.3	23.1	19.2
87 ⁵	15.3	16.0	18.0	19.8	15.0	21.1	21.6	19.5	21.3	17.5
85	13.8	14.5	16.5	18.4	13.7	19.2	19.9	18.1	19.8	16.2
82 ⁵	12.7	13.4	15.4	17.7	12.6	17.6	18.5	17.0	18.4	15.2
80	11.8	12.4	14.5	16.1	11.7	16.2	17.3	16.1	17.3	14.4
77 ⁵	11.0	11.7	13.7	15.1	11.0	15.0	16.2	15.3	16.2	13.7
75	10.2	11.1	13.0	14.2	10.3	14.0	15.3	14.6	15.3	13.0
72 ⁵	9.3	10.5	12.5	13.6	9.9	13.2	14.7	14.0	14.5	12.5
70	8.8	10.0	12.0	13.0	9.5	12.6	14.0	13.5	13.8	12.1
67 ⁵	8.4	9.5	11.5	12.1	9.1	12.0	13.4	13.1	13.2	11.7
65	8.0	9.1	11.1	11.6	8.7	11.5	12.8	12.8	12.6	11.3
62 ⁵	7.6	8.7	10.7	11.2	8.3	11.1	12.4	12.4	12.1	10.9
60	7.4	8.4	10.3	10.3	8.0	10.8	11.9	12.0	11.6	10.5
55	7.0	7.8	9.6	9.4	7.4	10.2	11.2	11.2	10.9	9.8
50	6.6	7.3	9.0	8.6	6.9	9.6	10.6	10.5	10.3	9.2
45	6.2	6.7	8.4	7.8	6.3	8.9	9.9	9.7	9.6	8.5
40	5.7	6.3	7.8	7.1	5.9	8.2	9.1	9.0	8.8	7.9
35	5.3	5.7	7.2	6.4	5.4	7.5	8.4	8.3	8.0	7.3
30	4.9	5.1	6.7	-	4.9	6.9	7.6	7.5	7.0	6.7
25	4.4	4.3	6.2	-	4.4	6.3	6.8	6.8	-	-
20	3.9	-	-	-	-	-	5.9	-	-	-

TABLE 5: Moisture relations of 5 sorts of leguminous seeds (lit. 8)

Relative humidity of the equilibrium atmosphere %	Equilibrium moisture content (w) (wet basis)				
	Brown beans	Dwarf French beans	Runner beans	Field beans	Green peas
92 ⁵	32.0	-	-	-	31.5
90	27.8	-	29.7	29.1	27.2
87 ⁵	25.2	25.5	27.0	26.2	24.8
85	23.1	23.2	25.0	24.4	23.0
82 ⁵	21.8	21.6	22.7	22.6	21.2
80	20.5	20.2	20.8	21.0	19.3
77 ⁵	19.1	18.8	19.2	19.4	17.8
75	18.0	17.6	17.8	18.0	16.8
72 ⁵	16.8	16.5	16.8	16.7	16.0
70	15.8	15.7	15.9	15.7	15.3
67 ⁵	15.0	15.0	15.1	14.9	14.7
65	14.3	14.5	14.3	14.2	14.1
62 ⁵	13.7	13.9	13.5	13.6	13.6
60	13.2	13.4	12.9	13.1	13.1
55	12.1	12.3	12.0	12.1	12.2
50	11.1	11.2	11.2	11.2	11.3
45	10.1	10.2	10.2	10.4	10.5
40	9.1	9.1	8.8	9.5	9.7
35	8.1	8.0	-	-	8.9
30	7.1	-	-	-	8.1
25	6.1	-	-	-	-
25 - 35°C - 77 - 95°F					

Seed
moisture

Relationship between moisture content and critical
temperature in Brassica oilspecies. The dark and
shadow areas indicate reduced germination (Appelqvist
1965)



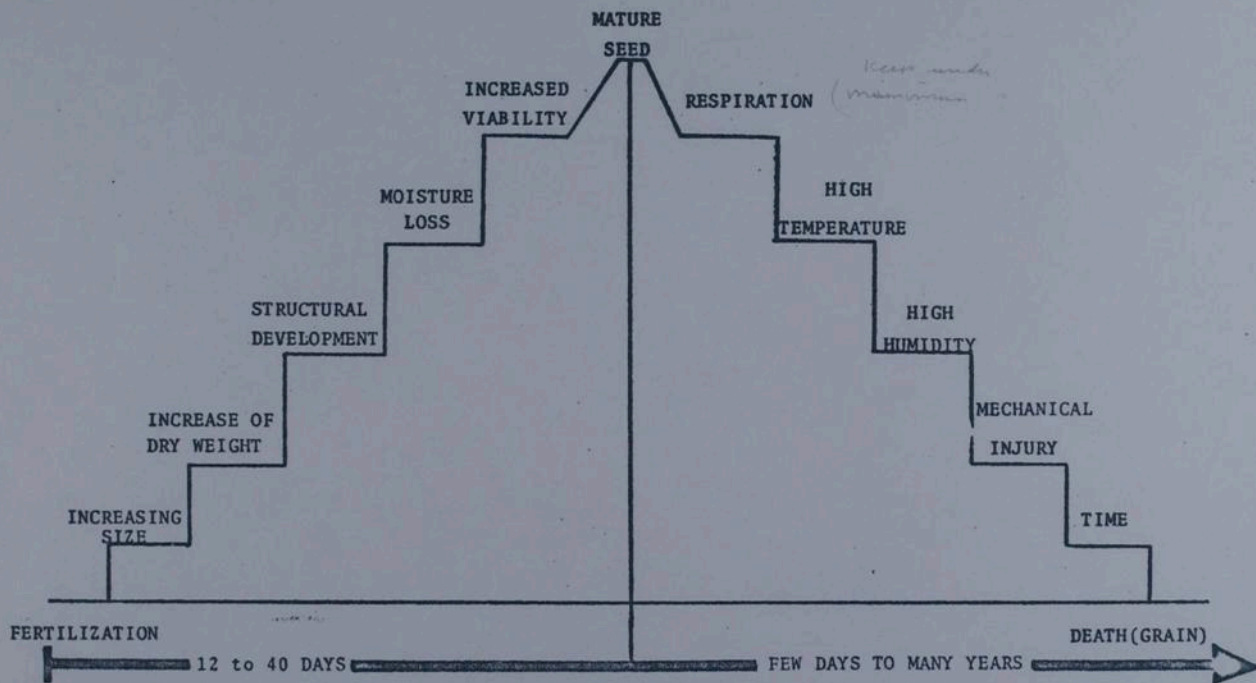


PLATE 5

THE RISE AND FALL OF A SEED'S CAPABILITIES

Recommended Conditions for Seed Storage
at Genetic Resources Centres

(Working paper for the FAO Panel on Plant
Exploration and Introduction, 1974).

E.H. Roberts

Department of Agriculture and Horticulture,
University of Reading, Earley Gate, Reading
RG6 2AT. U.K.

1. Maintenance of genetic integrity. A large amount of genetic mutation is associated with loss of seed viability in storage. It has been estimated that a fall in viability to 50% produces an amount of mutation equivalent to treating fresh seeds with 10,000 r of X-rays; and about half as much mutation is produced even if viability only falls to 80% (Roberts, 1973). It is suggested, therefore, that long-term seed storage systems should aim at maintaining viability at high levels, say 90% or above.
2. Advantages of infrequent multiplication. When seed stocks begin to lose viability, the seeds need to be grown in order to replenish the stocks. Particularly for those samples which contain mixtures of genotypes it is usually argued that it is necessary for this multiplication to take place in ecological situations identical or very similar to the areas where the samples were originally collected, in order that the original genetic complement of the sample is/subjected to different selection pressures and therefore altered. The procedure is costly and involves a great deal of time and organisation by skilled staff. The seed stocks themselves are subject to the dangers of complete loss, mechanical mixing and possibly cross-pollination and genetic selection. It follows that the frequency of multiplication should be reduced to the minimum. Thus any extra costs of providing a really good storage system should be set against the costs, dangers and inconvenience of multiplication.
3. The size of seed samples to be stored. Most seeds show roughly the same weight/volume relationship (Fig. 1).

Fig. 1. Seed weight/volume relationships

<u>Species</u>	<u>16/ft³</u>	<u>Kg/1000 cc</u>	<u>vol. (cc) occupied by 1 kg.</u>
Lucerne	48.0	0.768	1,300
Wheat	48.0	0.768	1,300
Maize	44.8	0.718	1,390
Walnuts	40.0	0.641	1,560
Rice	36.0	0.577	1,733

The number of seeds per unit volume varies greatly between species and no doubt the volume of the sample to be stored should be determined by seed number rather than seed weight. Nevertheless, it would seem reasonable as a basis for the present discussion to assume that each sample for long-term storage should be of approximately 1 Kg. (this would seem to be particularly desirable in the case of infrequent multiplication is proposed), and that the average size of sample would be about 1 Kg. Thus, from Table 1, it would

follow that an ample average size sample would be about 2,000 cc.

4. Storage conditions for very long-term storage. In the case of species whose seeds show orthodox behaviour (i.e., most species), the available evidence indicates that seeds can be stored for very long periods if they are stored at a sufficiently low temperature and moisture content. Moisture contents can be safely dropped to 2% in most species and certainly to at least 5% (i.e., in equilibrium with about 10-15% air relative humidity in starchy seeds or about 25% in oily seeds). Thus there would certainly be no danger in storing orthodox seeds at moisture contents at 5% or in equilibrium with air at 25% R.H. (This is in agreement with the views of Harrington, 1972). So far as temperature is concerned, there does not seem to be any dangerous lower limit, even at values approaching absolute zero; and all the available evidence suggests that there would be a continual increase in viability period with decrease in temperature to very low values (Roberts, 1972). From these considerations it is concluded that orthodox seeds for genetic conservation should be stored at about 5% moisture content (or in equilibrium with air at about 25% R.H.) and at the lowest temperature which is economically feasible.

5. Economic factors determining the temperature of storage. If it is recognised that, from the biological point of view, the lower the storage temperature the better, then the major factor determining the temperature to be used in practice is the capital cost of the store and associated refrigeration equipment necessary to maintain a given temperature. There is little difference in the cost of storage plant designed to maintain any temperature within the range $-10^{\circ}\text{C}.$ to $-20^{\circ}\text{C}.$; but there would be an increased cost of about 15% if the plant were designed to maintain a temperature of $-25^{\circ}\text{C}.$ for example. There would be little difference in the running costs of operating plant at say $-10^{\circ}\text{C}.$ or $-20^{\circ}\text{C}.$, providing the store was not being entered very frequently. Consequently it is suggested that the main storage facility at Genetic Resources Centres should be run at $-20^{\circ}\text{C}.$ and that orthodox seeds should be stored at about $5^{\circ}\text{C}.$ in moisture-proof containers at this temperature.

6. The cost of long-term storage facilities. The cost of storage facilities designed to operate at $-20^{\circ}\text{C}.$ is not very great but varies with the storage capacity required. The relationship between volume and cost is given approximately by the equation $\text{cost} = K \times \text{vol}^{2/3}$ (for cost in £ sterling and volume in cubic metres the value of K in U.K. is currently about 250 for "off-the-peg" units). Thus there is a worthwhile economy of scale.

In order to estimate roughly the cost of storing various numbers of samples, current prices of various sizes of "off-the-peg" low-temperature storage units have been taken. Fifteen percent has been added to the basic price in order to allow for provision of stand-by refrigeration equipment which, it is felt, should be standard for long-term seed storage facilities. If the units on which the estimates are based are surrounded by a brick structure it is estimated that their life should be greater than 100 years. The refrigeration equipment would need replacing every 12-25 years at about 80% of the current capital cost.

number of samples which can be held in a store of given volume.
 (a) The samples of seeds will be stored in cylindrical containers of 2000 cc average volume, and each cylinder will occupy a square-sectional prismatic space of 2,600 cc. (b) The containers will be placed on moveable shelving of the type found in library stacks so that in a rectangular store, the only working space that is necessary would be two passages 90 cm wide running at right-angles. Table 2 has been compiled incorporating these assumptions.

Table 2 (attached)

7. A comparison of current seed storage practice with that proposed here.

It is impossible to make comparisons between different practices of long-term seed conservation which are based on experience. There may be considerable errors in extrapolating what happens at very low temperatures and moisture contents from what happens at higher values and there may be a tendency to over-estimate the period of viability (Roberts, 1972, 1973), but it is the only method which is available. Nevertheless such estimates will give some indication of the order of magnitude of the multiplication intervals to be expected and the relative comparison between different storage systems is probably reasonably valid.

Table 3 shows the conditions employed in three seed-storage banks and compares them with the conditions proposed here in terms of the expected periods for viability of three common crop species to drop to 95%.

Table 3. Expected multiplication intervals under various systems of storage.

<u>Organisation</u>	<u>Temperature</u>	<u>Moisture content</u>	<u>Estimate multiplication interval, years (time taken for viability to drop to 95%) *</u>		
			<u>Triticum aestivum</u>	<u>Hordeum distichon</u>	<u>Vicia faba</u>
National Seed Storage Laboratory, Fort Collins, U. S. A.	4°C. **	32% R.H. (i.e., about 9% moisture content for non-oily seeds; about 4-6% for oily seeds)	8	95	17
Plant Genetics Resources Centre, Izmir, Turkey	0°C.	6% moisture cont. for cereals; 7-8% for oily seeds	27	540	90
National Seed Storage Laboratory, Hiratsuka, Japan.	-10°C. ***	4-6% moisture content	93	2,700	270
Conditions proposed here	-20°C.	5% moisture content.	328	27,000	1,300

* Calculated from Seed Viability Monographs (Roberts and Roberts, 1972).

** Three rooms are also available at -12°C.

*** Seed for distribution is held at -10°C.

Even allowing for large errors, from Table 3 it will be seen that for a number of common species, at least, problems connected with multiplication of seed stocks are likely to be minimal at a storage temperature of -20°C .

8. Recalcitrant species. No generalisation can be made yet about storage conditions appropriate for those seeds where the moisture content cannot be dropped below some relatively high critical value. It is suggested, therefore, that conditions for the storage of such recalcitrant seeds should not be specified at present and that immediate practical efforts should be concentrated on the conservation of orthodox species by seed storage.

9. Conclusions. Plant Genetic Resources Centres should include major facilities for the storage of orthodox species at 5% moisture content and -20°C . Such facilities should be designed to store samples which have a mean size of about 1kg. A routine programme of tests will be necessary to ensure that the samples are maintaining viability satisfactorily. For most species such tests probably need not be more frequent than, say every 20 years or so; though initially tests every five years may be necessary to inspire confidence. In order to maintain genetic integrity, the aim should be to keep viability above 90%.

At present it is not possible to recommend any generally suitable conditions for the storage of recalcitrant seeds which cannot be dried.

References

- Harrington, J.F. (1972). Seed storage and longevity. In T.T. Koslowski (ed.) Seed Biology, Vol III, 145-245. Academic Press, New York and London.
- Roberts, E.H. (1972). Storage environment and the control of viability. In E.H. Roberts (ed.) Viability of Seeds, 14-58. Chapman & Hall, London.
- Roberts, E.H. (1973). Loss of seed viability: chromosomal and genetic aspects. Seed Sci. & Tech. 1.
- Roberts, E.H., and Roberts, D.L. (1972). Viability nomographs. In Viability of Seeds. In E.H. Roberts (ed.) Viability of Seeds, 417-23. Chapman & Hall, London.

Recommended Conditions for Seed Storage
at Genetic Resources Centres.

E.H. Rober

Table 2. The capital cost of storing various numbers of seed samples at -20°C.

Capacity of "off-the-peg" storage unit		Basic cost in U.K. *		15% addition to basic cost for stand-by refrig. unit.		Useful storage space		No. of 1 Kg samples	Capital C per sampl
ft ³	m ³	£	\$	£	\$	m ³	% of total space		£
3,000	85	5,200	11,960	5,980	13,750	60	70	22,800	0.26
4,000	113	6,000	13,800	6,900	15,870	84	74	31,900	0.22
6,500	183	8,000	18,400	9,200	21,160	148	80	56,200	0.16
10,000	282	10,400	23,920	11,960	27,500	240	85	91,200	0.13

* At December 1973 prices and exchange rates (£1.00 = \$2.30 U.S.)

Proposed Standards for seed installations
used for long-term conservation of
'Base Collections'

(1) Standard germination tests and moisture determinations, where possible using procedures identical with or similar to those prescribed by the International Seed Testing Association (Proc. Int. Seed Test Ass., 31, 1966), should be carried out on all samples received for inclusion in base collections.

(2) Providing the viability is 80% or greater, the whole sample will be accepted for storage. If the viability is less than 80%, unless there are special circumstances (e.g., it is difficult or impossible to obtain samples showing greater viability), a proportion of the sample should be sent for multiplication to a centre concerned with active collections, so that the original sample can be replaced by one in better condition.

(3) The aim should be to store seed at $5 \pm 1\%$ moisture content (wet weight basis). If the sample received is greater than this, it should be dried by a method which is unlikely to affect viability; either of the following methods would be suitable: (a) Forced ventilation with warmed air not exceeding 40°C . (or initially 30°C . if the moisture content exceeds 18%); (b) equilibrating the seeds on trays at room temperature in a relative humidity controlled at approximately 25% R.H.

(4) When seeds are sufficiently dry (i.e., $5 \pm 1\%$), they should be stored in sealed containers in rooms or cabinets held at a temperature of -18°C . or less.

(5) Arrangements should be made to maintain such a temperature, or return it rapidly, should there be a power failure or a breakdown in the refrigeration equipment. It is therefore recommended that refrigeration plant should be duplicated and that a stand-by alternative power source should be provided.

(6) Arrangements should be made to carry out germination tests on sub-samples of 400 seeds on each acquisition at suitable intervals - say five years; though this period may ultimately be extended, at least for some species, in the light of experience.

(7) When such tests show that viability has dropped 5% below the initial value on acquisition, arrangements should be made to send a proportion of the seeds for multiplication to a centre where active collections are maintained. The remainder of the sample will be kept, however, until the replacement seed has been received from the 'active centre'.

(8) Full records will be kept of each acquisition.

SVA

THE TAXONOMIC INFORMATION RETRIEVAL SYSTEM
WITHOUT THE UNIQUE DESCRIPTOR STATE REQUIREMENT

WILLIAM J. WEIDA

15 APRIL 1974

In the latter half of the 1960's a small number of binary information storage and retrieval systems were developed to deal with data storage problems of a taxonomical nature. These systems, variously called STIRS, TAXIR or MEDLARS, were based on the principles of Boolean algebra and set theory, and they functioned by being able to assign a unique "descriptor state" to each item stored in a data file and then, by issuing a Boolean query, being able to call up and generate the data which was stored in a condensed state in the form of "characteristic functions". The concept of the unique descriptor state allowed the calculation of these "characteristic functions" which are actually sets of binary data, the set boundaries of which are determined by the binary number of the descriptor states. Since these characteristic functions expressed the data involved in an ever increasing number of descriptor states as the size of the data file increased in depth, and since all data could be generated from the characteristic functions based on a Boolean query, the storage space required for the data was greatly reduced and hence these systems could offer not only speed in data retrieval (generation) but also a great savings in required storage space.¹

However, it is the concept of the unique descriptor state for each item that greatly limits the usefulness of the STIRS-type system, for although this restriction fits well in a taxonomic scheme, it is obvious that most data is of a type that would need two or more descriptor states to be adequately stored. This paper is the result of an attempt to overcome this restriction and it will be presented

¹It is not the intent of this writer to elaborate on the basic STIRS-type systems in this paper since they are currently well known and widely applied. An excellent paper on this subject is "The Theory of the TAXIR Accessioner," by G. Estabrook and R. Brill, Mathematical Biosciences, 5, 1969, pp 327-340.

in the following format: Various "successful" methods for overcoming the unique descriptor state restriction will be presented in the order in which they were evolved, and since all of these methods involve restrictions of their own, these limitations to their implementation will also be discussed in the hope that one or more of these methods will provide the impetus for the development of a full fledged STIRS-type system which will accept all forms of data.

The problem with more than one entry in a descriptor state can be reduced to two important segments: (1) identifying the multiple entries in the descriptor state in a manner which allows the normal calculations necessary to produce the characteristic function to be performed, and (2) preserving the identity of the multiple entries throughout the calculations necessary to generate the data from the characteristic functions. To allow these segments to be concisely dealt with, the following rules and definitions were developed to deal with the multiple entry descriptor state situation:

(A) A number is defined as a set of all the binary ones in any column in the calculation of a characteristic function and the number is the set size. Therefore, $2 = (1\ 1)$.

(B) The compliment of a number is the compliment of its binary elements. Therefore, the compliment of $2 = \text{comp } (1\ 1) = (0\ 0) = 0$.

(C) In the data generation process, using the characteristic functions, if the total of binary ones in any column resulting from the summation of the numbers in that column is equal to or greater than the number of characteristic functions employed in the generation process the result is a binary one in the summation row (generated descriptor state).

(D) Multiple descriptor state entries must occur in rows, the binary numbers of which possess elements in the same column. i.e., if ones appear in the same column of rows i, j, \dots, n , of the state table then ones must appear in these same rows in some column of the binary row number matrix and these ones must ALL be in the SAME column. However, duplication that has occurred because of the binary ones in a certain column of the row numbers must not be repeated on the basis of ones in another column (see Example A, point 1) and thus when determining which rows are suitable for duplications, one must proceed in the direction of increasing row numbers at all times.

(E) No rows in the data matrix may be composed entirely of zeroes. This implies that every binary number must be used in sequence in numbering the descriptor state rows.

Example A demonstrates how these rules and definitions are applied to allow multiple descriptor state entries.

EXAMPLE A

Let the data matrix be

CHARACTERISTIC STATES	DATA ITEM						ROW NO.
	1	2	3	4	5	6	
A	1	0	1	0	1	0	001
B	0	0	0	1	0	1	010
C	1	0	1	1	1	1	011
D	0	1	0	0	0	0	100
E	1	1	1	0	1	0	101

In this case multiple descriptor state entries occur in every column but the following points should be noted:

(1) A one could not be placed in row 5, column 6 of the data matrix because this duplication would be based on ones which

were repeated as a result of the second column of the binary row numbers, and this column is empty for row 5.

(2) A one could not be placed in row 4, column 5 since the entry being duplicated here could only reside in row 5, column 5 (the only place where ones exist in the same binary row number column), and this duplication would not be in the direction of increasing row numbers.

(3) A column (in this case, column 2) must be reserved for the fourth descriptor state row to enter the data matrix because, since this row starts a new binary number column, it is implied that no duplication would be allowed in the row.

Now the characteristic functions are calculated by implementing definition A:

Characteristic function #1:	<u>#2</u>	<u>#3</u>
1 0 1 0 1 0	0 0 0 1 0 1	0 1 0 0 0 0
1 0 1 1 1 1	1 0 1 1 1 1	1 1 1 0 1 0
<u>1 1 1 0 1 0</u>	<u>1 0 1 2 1 2</u>	<u>1 2 1 0 1 0</u>
3 1 3 1 3 1		

Remembering that the numbers represented in the characteristic functions are actually sets of binary ones, one can then generate the rows of the data matrix from the characteristic functions by employing rules B and C:

Generate:	<u>Row 1</u>	<u>Row 2</u>	<u>Row 3</u>
	3 1 3 1 3 1	0 0 0 0 0 0	3 1 3 1 3 1
	0 1 0 0 0 0	1 0 1 2 1 2	1 0 1 2 1 2
	<u>0 0 0 1 0 1</u>	<u>0 0 0 1 0 1</u>	<u>0 0 0 1 0 1</u>
	<u>1 0 1 0 1 0</u>	<u>0 0 0 1 0 1</u>	<u>1 0 1 1 1 1</u>
	<u>Row 4</u>	<u>Row 5</u>	
	0 0 0 0 0 0	3 1 3 1 3 1	
	0 1 0 0 0 0	0 1 0 0 0 0	
	<u>1 2 1 0 1 0</u>	<u>1 2 1 0 1 0</u>	
	0 1 0 0 0 0	1 1 1 0 1 0	

Reference to the data matrix will show that this method does indeed overcome the restriction of unique descriptor state entries while still allowing a STIRS-type calculation and storage of data in characteristic functions. However, this method also incorporates a large number of restrictions and problem areas which would make its implementation quite difficult. These are:

(1) The highly restrictive form in which duplication of data is allowed. This problem is extremely critical in the first 5 rows of the data matrix, and although it would be alleviated somewhat as more rows are added, it still remains a very critical factor.

(2) The form of the first few rows determines the allowable duplications in the rows that follow, and hence these rows must be packed in the most efficient manner possible to allow maximum duplication in the rest of the matrix. This implies that these first few rows would probably have to be constructed of dummy data to allow the rest of the matrix to function efficiently.

(3) Every time a new column of binary row numbers is started (i.e., at row numbers 1,2,4,8,16,32,etc.) a data row with no duplication of descriptor states must be entered in the data matrix. This requires a double waste of space since the row must be virtually blank, and the one entry in the row must reside in a column that has been blank from row 1 up to the row in question.

(4) Even assuming that these problems could yield a tractable solution, one major problem area still remains; namely, that the whole system is extremely size dependent. The problem here is really the binary number system itself. Up to and including the number 6, the binary system allows a maximum of three binary ones to be involved in the calculation of any element of a characteristic function.

However, after this point the binary system always allows a greater number of binary ones to be involved in the calculation of the characteristic functions than the number of the characteristic functions in being, and this problem gets worse and worse as the binary numbers grow larger. In essence, this situation is tantamount to the insolvable problem of more unknowns than **equations**, and therefore to further extend the STIRS-type approach it is first necessary to develop a new number system to replace the binary system.

Toward this end three alternative number systems are presented, all three of which will function properly under the rules already enumerated, and each will also insure that at any point in the calculations there are always as many characteristic functions as there are possible numbers of binary ones which can be summed up in any column. Needless to say, all three of these systems are less efficient than the binary system itself. In order of increasing efficiency these are:

The System of Fours							
Number	Column Value						
	16	12	8	4	3	2	1
1							1
2						1	
3					1		
4				1			
5				1			1
6				1		1	
7				1	1		
8			1				
9			1				1
10			1			1	
11			1		1		
12		1					
13		1					1
14		1				1	
15		1			1		
16	1						

This system stabilizes in sets of four which are connected by elements in the number 1, 2, and 3 columns. It also stabilizes at approximately a 25% efficiency level, which implies that each characteristic function

will represent four lines of data after the system gets into moderately large numbers.

The System of Fives

Number	Column Value						
	18	13	8	4	3	2	1
1							1
2						1	
3					1		
4				1			
5				1			1
6				1		1	
7				1	1		
8			1				
9			1				1
10			1			1	
11			1		1		
12			1	1			
13		1					
14		1					1
15		1				1	

This system stabilizes in sets of five and hence, it approaches an efficiency of 20% so that each characteristic function represents five lines of data after the matrix gets sufficiently large.

A departure from the last two approaches is one which is less efficient than the binary number system but which increases geometrically in a binary number fashion.

The Binary X System

Number	Column Value					
	D	C	B	A	2	1
1						1
2					1	
3					1	1
4				1		
5				1		1
6				1	1	
7			1			
8			1			1
9			1		1	
10			1	1		
11		1				
12		1				1
13		1			1	
14		1		1		
15		1	1			
16		1				

In this system the numbers A,B,C,D take on different values as the row numbers increase, but a coherence is maintained within the system that allows each number set to be tied to all other sets. In addition, this system also allows a geometric progression that gives greater and greater efficiency as the data matrix size increases.

SIIRS-type systems built around these new number systems work well when rules A through E are employed. For example, when the system of fours is employed we have the following data matrix:

Descriptor State Number	EXAMPLE B							Row Number							
	Data Item							16	12	8	4	3	2	1	
1	0	0	0	0	1	0	0								1
2	0	0	0	1	0	0	0								1
3	0	0	1	0	0	0	0								1
4	0	1	0	0	0	0	0								1
5	0	1	0	0	1	0	0								1
6	0	1	0	1	1	0	0								1
7	0	1	1	1	1	0	0								1
8	0	0	0	0	0	1	0								1
9	0	1	0	0	1	1	0								1
10	0	1	0	1	1	1	0								1
11	0	1	1	1	1	1	0								1
12	0	0	0	0	0	0	1								1
13	0	1	0	0	1	1	1								1
14	0	1	0	1	1	1	1								1
15	0	1	1	1	1	1	1								1
16	1	0	0	0	0	0	0	1							1
17	1	0	0	1	1	1	1	1							1
18	1	1	0	1	1	1	1	1							1
19	1	1	1	1	1	1	1	1							1

Calculation of characteristic functions from this data proceeds as in Example A to yield

Characteristic Function #1	#2	#3
0 0 0 0 1 0 0 0	0 0 0 1 0 0 0 0	0 0 1 0 0 0 0 0
0 1 0 0 1 0 0 0	0 1 0 1 1 0 0 0	0 1 1 1 1 0 0 0
0 1 0 0 1 1 0 0	0 1 0 1 1 1 0 0	0 1 1 1 1 1 0 0
0 1 0 0 1 1 1 0	0 1 0 1 1 1 1 0	0 1 1 1 1 1 1 0
1 1 0 0 1 1 1 1	1 1 0 1 1 1 1 1	1 1 1 1 1 1 1 1
<u>1 4 0 0 5 3 2</u>	<u>1 4 0 5 4 3 2</u>	<u>1 4 5 4 4 3 2</u>
#4	#5	#6
0 1 0 0 0 0 0 0	0 0 0 0 0 1 0 0	0 0 0 0 0 0 0 1
0 1 0 0 1 0 0 0	0 1 0 0 1 1 0 0	1 0 0 0 0 0 0 0
0 1 0 1 1 0 0 0	0 1 0 1 1 1 0 0	1 1 0 0 1 1 1 1
0 1 1 1 1 0 0 0	0 1 1 1 1 1 0 0	1 1 0 1 1 1 1 1
<u>0 4 1 2 3 0 0</u>	<u>0 3 1 2 3 4 0</u>	<u>4 3 1 2 3 3 3</u>

Generation of data from the characteristic functions yields (for the purpose of space conservation only selected rows of the data matrix are generated):

Generate:#5 Row	#11Row	#14 Row
1 4 0 0 5 3 2	0 0 1 1 0 0 0	0 0 1 1 0 0 0
0 0 1 0 0 0 0	0 0 1 0 0 0 0	1 4 0 5 4 3 2
0 0 0 0 0 0 0	1 4 5 4 4 3 2	0 0 0 0 0 0 0
0 4 1 2 3 0 0	1 0 0 0 0 1 1	1 0 0 0 0 1 1
1 0 0 0 0 0 1	0 3 1 2 3 4 0	1 0 0 0 0 0 1
1 0 0 0 0 0 0	1 0 0 0 0 0 0	0 3 1 2 3 3 4
0 0 0 0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0
0 1 0 0 1 0 0	0 1 1 1 1 1 0	0 1 0 1 1 1 1

Concerning this example, the following points should be noted:

(1) The previous problems with the rows of data at and beyond the binary number seven have been overcome and any size data matrix may now be used.

(2) All of the problems associated with the strict structuring of Example A are still present, except that now the mandatory blank rows which accompany each new number column (i.e. rows 8,12,16) consume a set percentage of the useable data space.

(3) This system is still somewhat size dependent and as larger data matrix sizes are approached confusion may result in the data generation process. This condition can occur when the matrix is nearly fully packed with binary ones in all available spaces and the data row to be generated contains a zero in a space which is eligible to receive a duplicated one. In this situation the sheer numerical weight of the sets used in the generation of this specific row of data may lead to the generation of a one in the space where the zero should reside.

To overcome this problem a code system was developed which would indicate the actual condition of the data set in each column of data used in the calculation of the characteristic functions. This code system, composed of set designators, is employed as follows:

Referring to Example B, in the calculation of characteristic function number 1, five descriptor states are used in this calculation:

0	0	0	0	1	0	0
0	1	0	0	1	0	0
0	1	0	0	1	1	0
0	1	0	0	1	1	1
1	1	0	0	1	1	1
1	4	0	0	5	3	2

The set designators refer to the data sets in each column of this calculation using the following code:

C = complete set - a column of five ones in this case.

E = empty set - a column of zeroes.

P+ = a set composed of more than one-half binary ones.

P- = a set composed of less than one-half binary ones.

Thus for characteristic function number one we would have

P- P+ E E C P+ P-

In order to use these signs in the data generation process they must have compliments. These are as follows:

Comp C = E
Comp E = P-
Comp P- = P+
Comp P+ = P-

In using these in the data generation process, the following rules apply:

(1) A C in any column of the characteristic functions used in the generation of the descriptor state is generated as a 1 no matter what other set designators are present in the column.

(2) An E in any column of the characteristic function used in the generation of the descriptor state is generated as a 0 unless:

- (a) There are two or more P+'s in the same column for every E.
- (b) There is a C in the same column.

If (a) or (b) occur, the column is generated as a 1.

(3) Columns composed of all P-'s, all P+'s or a mixture of both are generated as 1's.

As an example of the use of set designators, the calculations already completed in Example B, are repeated in set designator form.

EXAMPLE C

<u>Characteristic Function No.</u>	<u>Normal</u>	<u>Compliment</u>
1	P- P+ E E C P+ P-	P+ P- P- P- E P- P+
2	P- P+ E C P+ P+ P-	P+ P- P- E P P- P+
3	P- P+ C P+ P+ P+ P-	P+ P- E P- P- P- P+
4	E C P- P- P+ E E	P- E P+ P+ P P P-
5	E P+ P- P- P+ C E	P- P- P+ P+ P- E P-
6	E P+ P- P- P+ P+ C	P- P- P+ P+ P- P- E
7	C P+ P- P- P+ P+ P+ E P-	P+ P+ P- P P P-

Then, Generate:

<u># 5 Row</u>	<u># 11 Row</u>	<u>#14 Row</u>
P- P+ E E C P+ P-	P+ P- P- P- E P- P+	P+ P- P- P- E P- P+
P+ P- P- E P- P- P+	P+ P- P- E P- P- P+	P- P+ E C P+ P+ P-
P+ P- E P- P- P- P+	P- P+ C P+ P+ P+ P-	P+ P- E P- P- P- P+
E C P- P- P+ E E	P- E P+ P+ P- P- P-	P- E P+ P+ P- P- P-
P- P- P+ P+ P- E E	E P+ P- P- P+ C E	P- P- P+ P+ P- E P-
P- P- P+ P+ P- P- E	P- P- P+ P+ P- P- E	E P+ P- P- P+ P+ C
<u>E P- P+ P+ P- P- P-</u>	<u>E P- P+ P+ P+ P- P-</u>	<u>E P- P+ P+ P- P- P-</u>
0 1 0 0 1 0 0	0 1 1 1 1 1 0	0 1 0 1 1 1 1

Although the use of set designators does overcome some of the discrimination problems previously mentioned, this method still leaves something to be desired since:

(1) The highly structured set of rules incorporated in Example A must still be followed for data storage.

(2) An added restriction is now imposed in that the partial set designators (P+ and P-) will only function on data sets of an odd size. Obviously, if six descriptor states are used in the calculation of a characteristic function, a condition could occur where 3 ones and 3 zeroes are present in any column. This injects an indeterminacy into

the generation calculations that renders the set designators ineffective. Although this restriction is critical, it is generally no more than a nuisance, since it would only require the storage of extra dummy descriptor states to insure an odd number of descriptor states in the calculation of characteristic functions.

At this point the basic problem of altering the STIRS-type systems to allow multiple descriptor state entries has been solved, but due to the highly structured nature of the data required to perform these calculations, this solution is of little practical use. The efficiency of STIRS is, to a great extent, based on its ability to remain within the binary domain during all calculations and the set designator concept makes this difficult. Moreover, it is impractical to assume that anyone would be willing to structure the data into the necessary format prior to submitting it for storage. Therefore, this problem is further investigated with the effort now being concentrated in two areas to develop:

- (1) the ability to accept data without restriction in format and in any density from no entries to a data matrix that is 100% ones.
- (2) the ability to remain within the binary domain.

An investigation of the descriptor states which must be left blank except for a solitary binary entry when a new column is entered in the binary row number table (see rows 4, 8, 12, 16, Example B) shows that these descriptor states basically perform a place-keeping function in the data matrix. If the data matrix is restructured as follows:

EXAMPLE D

<u>PLACE-KEEPERS</u>				<u>DATA</u>			<u>System of Fives</u>
D'	C'	B'	A'	A	B	C	
0	0	0	0	0	0	1	<u>8 4 3 2 1</u> 1
0	0	0	0	0	1	0	1
0	0	0	0	1	0	0	1
0	0	0	1	0	0	0	1
0	0	0	1	0	0	1	1 1
0	0	0	1	0	1	1	1 1
0	0	0	1	1	1	1	1 1
0	0	1	0	0	0	0	1
0	0	1	0	0	0	1	1 1
0	0	1	0	0	1	1	1 1
0	0	1	0	1	1	1	1 1
0	0	1	0	1	1	1	1 1

It becomes obvious that these place-keeping columns exhibit a pattern which is dependent only on the number system used, and which is independent of the data stored in the descriptor states. Therefore, the place-keeping columns no longer need to be carried in the data storage at all, they can be automatically tacked on by the computer during calculation, and since they will be constant, the descriptor state number or characteristic function number will be sufficient to indicate which unique pattern should be used. This fact in itself will greatly reduce the lateral storage space required for the characteristic functions (by over 50% in most cases), but it also has other implications.

If this technique can be used for the place-keeper columns, it can also be applied to the descriptor rows which must be left blank except for the solitary binary 1 which indicates the entry of a new descriptor state number column. Thus, taking the Binary X number system described previously we have:

EXAMPLE E

<u>Descriptor State Row</u>	<u>Binary X Row Number</u>
1	0 1*
2	1 0*
3	1 1
4	1 0 0*
5	1 0 1
6	1 1 0
7	1 0 0 0*
8	1 0 0 1
9	1 0 1 0
10	1 1 0 0

Of these descriptor state rows, numbers 1, 2, 4, and 7 (marked by *) are rows which are not eligible for duplication and hence have a predetermined format (row 1 also has this predetermined format since the data in this row must be as close to a 100% density of binary ones as possible to allow maximum duplication in the rows that follow.) Since this format is predetermined and hence, constant within any given situation, it is implied that the only descriptor states that really need to be considered are those which can have a varied format, i.e., rows 3, 5, 6, 8, 9, and 10, since these rows are the only ones which could conceivably change the calculations. This in turn leads to the technique illustrated in Example F.

EXAMPLE F

Given the Binary X number system (the same as the binary number system in this case) and six descriptor states:

<u>DATA</u>	<u>DESCRIPTOR STATE NUMBER</u>
A	0 0 1
B	0 1 0
X ₁	0 1 1
C	1 0 0
X ₂	1 0 1
X ₃	1 1 0

Now, with one column of data for an example (if a method works for a one-column width of data it will also work for an n-column width) we remove from further consideration rows 1, 2, and 4 leaving

<u>DATA</u>	<u>DESCRIPTOR STATE NUMBER</u>
X ₁	0 1 1
X ₂	1 0 1
X ₃	1 1 0

If we now modify the set designator rules to the following format

Set Size: 2 = C		Comp P = P
1 = P	and	Comp E = P
0 = E		Comp C = E

We can now calculate for the following situations:

(a) $X_1 = X_2 = X_3$

Characteristic Function:

#1 = 2 = C
 #2 = 2 = C
 #3 = 2 = C

Generate:

#3 = C	#5 = C	#6 = E
C	E	C
$\frac{E}{1}$	$\frac{C}{1}$	$\frac{C}{1}$

(b) $X_1 = 0, X_2 = X_3 = 1$

Characteristic Function:

#1 = 1 = P
 #2 = 1 = P
 #3 = 2 = C

Generate:

#3 = P	#5 = P	#6 = P
P	P	P
$\frac{E}{0}$	$\frac{C}{1}$	$\frac{C}{1}$

(c) $X_1 = X_3 = 1, X_2 = 0$

Characteristic Function:

#1 = 1 = P
 #2 = 2 = C
 #3 = 1 = P

Generate:

#3 = P	#5 = P	#6 = P
C	E	C
$\frac{P}{1}$	$\frac{P}{0}$	$\frac{P}{1}$

(d) $X_1 = X_2 = 0, X_3 = 1$

Characteristic Function:

#1 = 0 = E
 #2 = 1 = P
 #3 = 1 = P

Generate:

#3 = E	#5 = E	#6 = P
P	P	P
$\frac{P}{0}$	$\frac{P}{0}$	$\frac{P}{1}$

(e) $X_1 = 1, X_2 = X_3 = 0$

Characteristic Function:

#1 = 1 = P

#2 = 1 = P

#3 = 0 = E

Generate:

#3 = P	#5 = P	#6 = P
P	P	P
P	E	E
I	O	O

(f) $X_1 = X_2 = X_3 = 0$

Characteristic Function:

#1 = 0 = E

#2 = 0 = E

#3 = 0 = E

Generate:

#3 = E	#5 = E	#6 = P
E	P	E
P	E	E
O	O	O

This system has several important advantages over the systems previously investigated:

- (1) Complete freedom of data entry is possible with no rules or restrictions.
- (2) A matrix of 100% binary ones is possible meaning that no space is wasted with mandatory zero entries.
- (3) Any amount of data could be recorded simply by numbering the consecutive sets of three. i.e., to generate a descriptor state the Boolean query would be 1 - 3 for set one, row three.

Note that the actual row numbers must be used in all cases. And note also that although this does solve the problem of structuring the data, in order to convert the set designators to the binary language two columns of binary numbers would be required for every column of the characteristic functions to be stored and thus, in this case where there is one characteristic function for every row of data we would end up requiring twice the storage space as would be required to store the data alone.

To attack this problem we expand the size of the data set (using the Binary X number system) to ten rows. This yields, after rows 1, 2, 4, and 7 have been deleted in manner shown in Example F:

(d) $X_1, X_3, X_5 = 0, X_2, X_4, X_6 = 1$

Characteristic Functions:

#1 = 2 = P+
 #2 = 0 = E
 #3 = 2 = P+
 #4 = 2 = P+

Generate:

#3 = P+	#5 = P+	#6 = E
E	P+	E
E	P+	P+
$\frac{E}{O}$	$\frac{E}{1}$	$\frac{E}{O}$
#8 = P+	#9 = E	#10 = E
P+	E	P+
E	E	P+
$\frac{P+}{1}$	$\frac{P+}{O}$	$\frac{P+}{1}$

(e) $X_2, X_6 = 1, X_1, X_3, X_4, X_5 = 0$

Characteristic Functions:

#1 = 1 = P-
 #2 = 0 = E
 #3 = 2 = P+
 #4 = 1 = P-

Generate:

#3 = P-	#5 = P-	#6 = P
E	P+	E
E	P+	P+
$\frac{P-}{O}$	$\frac{P-}{1}$	$\frac{P}{1}$
#8 = P-	#9 = P-	#10 = P-
P+	E	P+
E	E	P+
$\frac{P-}{O}$	$\frac{P-}{O}$	$\frac{P-}{1}$

(f) $X_2=1, X_1, X_3$ through $X_6 = 0$

Characteristic Functions:

#1 = 1 = P-
 #2 = 0 = E
 #3 = 1 = P-
 #4 = 0 = E

Generate:

#3 = P-	#5 = P-	#6 = P-
E	P+	E
P-	P-	P-
$\frac{P+}{O}$	$\frac{P+}{1}$	$\frac{P+}{O}$
#8 = P-	#9 = P-	#10 = P-
P+	E	P+
P-	P-	P-
$\frac{E}{O}$	$\frac{E}{O}$	$\frac{E}{O}$

This system gives four characteristic functions for every six lines of data. Converting the set designators to binary gives a resultant efficiency of 75%. This means a loss of 25% storage space gains all of the efficiency of a SPIRS-type system - plus having a complete freedom of entry and format in the data stored in the system. The only conceivable data combination that can cause any indeterminacy in this system is if

binary ones were stored in rows 3 and 10, or 5 and 9, or 6 and 8 only, and this condition could be simply screened and corrected during the filing of data, or it could be eliminated by structuring these pairs of descriptor states so that the pairs were mutually exclusive within themselves.

Since the matrix of data no longer has to abide by the rules of mutually exclusive descriptor states, the transpose of the matrix can now be considered as well as the original matrix, leading to the possibility of treating the items of data as descriptor states themselves. This in itself opens up another avenue of possibilities for the STIRS-type format.

In summary, this paper has provided an evolutionary development of a method to extend the capabilities of the current STIRS-type systems. This method, like all others, has its drawbacks but it is felt that these are greatly outweighed by the possibilities of reduced file search time and the other accompanying benefits of the STIRS-type format. Further research will be directed toward the end of incorporating more data in each set of characteristic functions (stopping the previous example at the Binary X number 10 was not done by chance - above this number of rows the set designators currently cannot handle the entire scope of possibilities associated with the greater number of descriptor states), and in refining the rules for set designator calculations. It is hoped that this initial exploration of the STIRS-type system without the unique descriptor state requirements will be of aid in furthering the development of this important area of data management.