

Structuring Strain Data for Storage and Retrieval of Information on Fungi and Yeasts in MINE, the Microbial Information Network Europe

By W. GAMS,^{1*} G. L. HENNEBERT,² J. A. STALPERS,¹ D. JANSSENS,⁵
M. A. A. SCHIPPER,¹ J. SMITH,³ D. YARROW⁴ AND
D. L. HAWKSWORTH⁶

¹*Centraalbureau voor Schimmelcultures, PO Box 273, 3740 AG Baarn, The Netherlands*

²*Mycothèque de l'Université Catholique de Louvain, Place Croix du Sud 3,
B-1348 Louvain-la-Neuve, Belgium*

³*University of Amsterdam, Department for Medical Informatics, Academic Medical Centre,
Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands*

⁴*Centraalbureau voor Schimmelcultures, Yeast Division, Julianalaan 67, 2628 BC Delft,
The Netherlands*

⁵*Laboratorium voor Microbiologie en Microbiële Genetica, K. L. Ledeganckstr. 35,
B-9000 Gent, Belgium*

⁶*CAB International Mycological Institute, Ferry Lane, Kew, Surrey TW9 3AF, UK*

(Received 6 November 1987; revised 18 February 1988)

A distributed Microbial Information Network Europe (MINE) is being constructed by a number of major microbial culture collections in countries of the European Community, with the support of the Biotechnology Action Programme (BAP) of the Commission of the European Community. The representatives of the collections participating in MINE have agreed to adopt a general format for the computer storage and retrieval of strain data. This uniform format will facilitate the electronic combination and exchange of data from different collections in order to produce integrated catalogues and the use of identical commands to search the different databases. It is recommended to other collections who may wish to contribute data to the MINE network or between themselves.

Three kinds of records can be linked to the leading 'species records': strain records, synonym records, and alternative morphonym records. A minimum data set of 30 fields (similar to the fields used for producing catalogues) is defined that facilitates the exchange of data between the national nodes and serves as a directory to strains available at other nodes. It is suggested that the full strain record comprise 99 fields, grouped in 12 blocks: internal administration – name – strain administration – status – environment and history – biological interactions – sexuality – properties (cytology, biomolecular data) – genotype and genetics – growth conditions – chemistry and enzymes – practical applications. Several fields are divided into subfields of different ranks. Delimiters are used either to separate a range of entries that have to be indexed or to divide an entry from the reference to its source or remarks that should not be indexed. The contents and structure of the fields proposed for filamentous fungi and yeasts are described and in some cases illustrated by examples. Uniformity of input is essential for indexed fields and desirable for non-indexed fields. Seven thesaurus files are envisaged to ensure consistency.

INTRODUCTION

Data associated with strains of micro-organisms preserved in culture collections are of importance to systematic, ecological, physiological and biochemical research and teaching, and

to many facets of industry, particularly pharmaceuticals and biotechnology. The information on cultures is scattered amongst the various collections and in the literature, only a fraction being accessible through published catalogues. All collections see the desirability of computerizing the available data; curators, in particular, need access to the information on all strains in the collection from any approach, including the most recent data, to satisfy increasing numbers of requests. Access to up-to-date information, retrieved either at the collection site or externally, is also important to users of strains, particularly research workers in both industry and universities.

Investigations into the need for a European information system for micro-organisms carried out by Anonymous (1983) and Environmental Resources Ltd (1984) drew attention to the urgent requirement of European biotechnological industries to have access to micro-organisms and to data available in culture collections in the European Community (EC). One of the first documentation systems for microscopic fungi was developed a decade ago (Anderson *et al.*, 1976), with the limited scope of producing a compendium of 400 species of soil fungi (Domsch *et al.*, 1980). This is a representative example of a closed system for the computerization of a limited number of organisms and fields. Since then the possibilities for organizing strain data have increased considerably. A first small-scale project for storing and retrieving a limited amount of strain data was described by Bryant (1983). Since that time some local approaches to the computerization of microbial collections have been realized: the United Kingdom MiCIS system (Microbial Culture Information Service, using RAPPORT software and operated by the Laboratory of the Government Chemist, Department of Trade and Industry); the On-line Services system at NCYC (National Collection of Yeast Cultures, Agricultural and Food Research Council, Norwich, UK); the Canadian Fungal Collection Database (National Research Council of Canada, Halifax; formerly on a CYBER mainframe, now transferred to a VAX); and some other less sophisticated or less comprehensive computerized records (e.g. the Nordic Register in Scandinavian Countries). The MiCIS system is designed to make the resources of the British National Culture Collections more accessible to industry. It is a centralized system that aims at storing complete data sets, including physiological and biochemical activities, an approach that is not acceptable to many non-British collections, who are not in the position to cede all their information to databases outside their own country.

The MSDN (Microbial Strain Data Network) is a different approach, designed to act as a central directory of depositories for microbial strains and cell lines, and to disclose information on the sites where particular types of data are available, categorizing the types of available data on the strains and the means used to record and access these data. Its data structure (RKC Code; Rogosa, Krichevsky & Colwell, 1986) is discussed below. This system therefore supplements and does not duplicate the specific strain databases.

The needs of numerous collections for computerization are similar and can be coordinated in such a way that several databases become optimally compatible. Compatibility is necessary to facilitate the exchange and combination of data between different nodes, for producing integrated catalogues, and to search the different databases with identical commands. The establishment of an information network is considered the most satisfactory approach to the coordination of computerized strain data in the microbial culture collections in the EC. The EC is sponsoring this development through the Biotechnology Action Programme (BAP). The objective is that the databases of the major collections of each country will be connected in such a way as to allow them to retain the property of and the responsibility for the data in their keeping, while at the same time allowing optimal access to the data of all participants.

Most collections of micro-organisms are concerned with bacteria (including actinomycetes), filamentous fungi, and yeasts. Information on the features of strains of these organisms can in principle be stored in a general scheme, comprising data fields, each defined to contain a certain type of data. The fields may be either simple or divided into subfields. This contribution is concerned with the formats for filamentous fungi and yeasts only, agreed by participants in the MINE project. A number of features differ for bacteria and require the definition of distinct data fields. More fields will be required for data on algae, cell-lines and hybridomas. A corresponding paper on the format for bacteria is in preparation.

Different approaches to the storage of microbial data in a database can be envisaged. One possibility is that each item of information is represented as a single combination of digits and large numbers of entries must be accurately defined; in the case of an occurrence (or, more rarely, absence), a single symbol is entered. This option is that chosen by the RKC code (Rogosa *et al.*, 1986), which attempts to address all possible physiological and morphological properties of an organism (as well as their alternatives) using a flat-file structure. Depending on the number of items to be entered, more or less voluminous forms must be completed before coded data can be entered into the database.

In another option, a limited number of fields, possibly divided into a hierarchical sequence of subfields, is filled with simple or more complex data, represented either by a single symbol or expressed in a more or less natural language, that can be retrieved and reproduced in various ways. This is the option adopted by the MINE project. Several commercially available software packages are suited for this purpose (e.g. BASIS, a product of Battelle Software Products Center, Columbus, Ohio, USA, now Information Dimensions Inc., Dublin, Ohio).

GENERAL CHARACTERISTICS OF THE MINE DATABASE

Definition of system requirements

Criteria used in selecting hardware and software for microbial strain databases include: a large storage capacity, able to cope with many megabytes of data; the possibility of including variable field lengths; the possibility of incorporating different files or record types and linking them with each other; the possibility of indexing fields, subfields or other fractions of fields; flexible data management, with rapid and easy data retrieval; and a differential security system to prevent access by unauthorized users. When selecting software and defining the data structure it is vital to have an open-ended system with no limitation of space, so that fields can be modified or added at any time.

In this paper only the general format is outlined to illustrate the possibilities of the MINE data banks now accessible or in an advanced stage of preparation. This format is largely independent of the choice of software. The implementation, however, takes advantage of the possibilities of subfield structures offered by certain software packages. To facilitate validation and differentiated indexing, a rather large number of fields has been defined. For displaying and reporting, however, fields can be combined by 'mapping' functions.

Each field is listed here with its label or prefix. The sequence and numbering of the fields need not be uniform in different systems to be compatible with other partners. Consequently no numbers are assigned to the fields here.

General structure of the data; record types

The basic unit of a database is the record. The data of each record are further distributed in fields. The contents of fields (or parts of fields) that have to be searched rapidly must be indexed.

Principally an individual record, a **strain record**, STR, is created for each microbial strain.

Instead of repeating the taxonomic name with each strain record, it was found to be more convenient to establish, besides the above-defined strain record, a separate record type, the **species record**, SP, containing taxonomic names of the species (or lower ranks in the taxonomic hierarchy), to which the strain records can be linked (Fig. 1). Moreover, not all data available in a collection apply individually to strains maintained in the collection. Data may pertain to the taxon in general and, as such, either be actually applicable to all strains or be presumably, but not necessarily, applicable to all the strains of the taxon (e.g. a property). These data are therefore preferably entered in the species record. Additional data are, however, explicitly tied to an individual strain kept in the collection; such data are entered in strain records.

Because of nomenclatural instability, several synonyms are often in common use for a given micro-organism. The user should be able to access a particular species from any generally known name. For fungi, the situation is further complicated by the existence of independent sexual and asexual sporulation stages (teleomorph and anamorph) in one species, each of which may have a

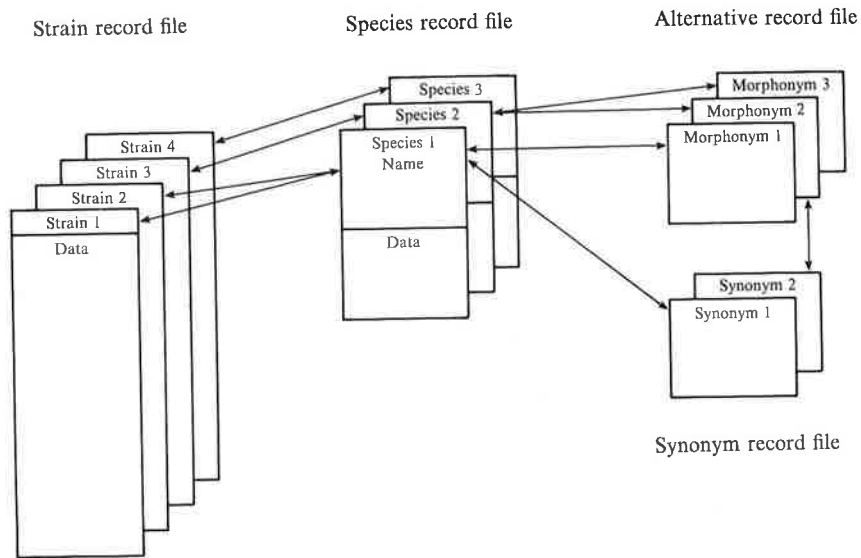


Fig. 1. Connection between species and strain records and between strain, synonym and alternate morphonym records.

separate name, for which G. L. Hennebert (unpublished) proposes, the term 'morphonym'. Two more record types have been defined to cope with these problems: ALT, the **alternative state record** for names of anamorphs and teleomorphs, and SYN, the **synonym record** for synonymous names, which can be linked to a particular species (or alternative state) record. A strict rule for the choice of a particular morphonym, viz. for the teleomorph or anamorph, cannot be formulated. If the teleomorph-anamorph relationship is equivocal, the name of the actual morph must be used. If it is unequivocal, the teleomorph name is to be preferred; otherwise, the asexual form of sporulation, being that most commonly observed, may justify the application of the anamorph name. Once a choice has been made for one of the available morphonyms, this name is to be used for all strains of that species subsequently accessioned in the collection.

For every record, one field, record type (RT), must be filled with one of these abbreviations (SP, STR, SYN or ALT). Each record has a unique accession number (ACCN). The connection between different record types can be established by different means according to the software used, e.g. by means of a further field, 'alternate posting' (APOS), containing the accession number of the associated species record.

It should be possible to retrieve and display both strain data and species data with a single search. While the field contents are normally indexed through the record number, they may also be indexed a second time through the APOS value, which links with the associated species record. Thus, while searching for a particular strain feature, not only can a particular strain be retrieved but at the same time also the associated species record. Conversely, the retrieval of strain records and strain data starting from the species record operates via the same link (Fig. 1).

While the strains of one taxon are tied to one species record through the same APOS number, several other SYN or ALT records may also be related to that taxon. Indeed, several other names, either synonyms or alternative morphonyms, or synonyms of alternative morphonyms, may be linked to one accepted species name.

Standardization of input

A standardized structure is crucial for indexed data. Either validation functions or a thesaurus can check the correctness of the input. The thesaurus facility can also switch from abbreviations to full names or from a synonymous name to the accepted name, both during input and during retrieval. Typist's input is checked by rules of validation, which can be defined for each field, before being entered into the database. The validation function ensures, for instance, that the

obligatory fields contain data and that the combination of text in the fields SP-SSP-VAR-F-FSP and the strain numbers are unique. With non-indexed fields or subfields more variation in the structure of data can be tolerated. However, standardization is also required in non-indexed fields, where names may vary according to language or where transliteration into Roman typography is needed.

Standardization particularly applies to fungal names, collection acronyms, scientific host names, chemical names, nutrient media, geographical names, and literature.

Fungal names. The chosen name and its spelling must be in accordance with the International Code of Botanical Nomenclature. Authors' names that are not indexed are either abbreviated according to Hawksworth (1980) or given in full. The names of the last two authors are connected by '&' (or 'et').

Collection acronyms. Standardized abbreviations are listed in Pridham (1981), McGowan & Skerman (1982) and Hawksworth & Kirsop (1988).

Host plants. The correct and valid scientific names are used in preference to vernacular names, and given without authors.

Nutrient media. Reference to culture media can be either by symbols or by a number. Storing full names was found to be impractical. The symbols or numbers are standardized according to a thesaurus file that also refers the user to the lists of media published by major culture collections or in handbooks. Abbreviations are to be chosen in such a way that each major component of the medium is, if possible, expressed by a single letter, sometimes followed by a figure indicating the concentration of an ingredient. Examples: MA 4 = 4% malt extract agar; GPYA = glucose/peptone/yeast extract agar.

Chemical compounds and enzymes. Abbreviations or numbers can be used for entering chemical compounds [CAS, Chemical Abstracts Service (the CAS numbers are listed in e.g. Merck, 1987/88); API, Analytical Products Inc.] and enzymes (Enzyme Nomenclature, 1984), but in the database the correct full name is stored. Synonyms are converted to the current chemical names.

Geographical names. Names of countries are spelt in English, those of cities according to the national usage in the local language. Names originally spelt in Cyrillic or other non-Roman letters are transliterated according to the ISO (International Standards Organization) norms. Current abbreviations of countries may be entered; they are converted by the thesaurus to the official names.

Literature references. Authors' names can be spelt using all characters available in the extended ASCII code (American Standard Code for Information Interchange), i.e. including letters with diacritic signs commonly used in European languages; concerning transliteration see above. Journal references are unabbreviated or abbreviated according to the *World List of Scientific Periodicals* or other standards.

The correspondences between names, synonyms, and abbreviations are also laid down in the thesaurus files. The following thesaurus files are envisaged: fungal taxa of the rank of genus and higher (based on Hawksworth *et al.*, 1983), substrata (including host plants/animals), geographical names, chemical substances, enzymes, culture media, and collection acronyms.

Definition of different data sets

A practical but important distinction is made between the data fields according to their accessibility to the public. The **minimum data set** (MDS) contains the basic data of a collection. It may be exchanged between the national nodes of the MINE network and may serve as a directory to the information on strains available at other nodes. For fungi and yeasts it consists of 30 data fields. The **full data set** (FDS) contains all data of a collection. It consists of fields that are generally accessible (the minimum data set) together with many others that are not. About 115 fields (of which 99) have been adopted to accommodate the information of the full data set. The printed catalogue of a collection generally contains the data pertaining to the 30 MDS fields where available. A collection may also wish to add more information in its printed catalogue, such as conditions of maintenance, references to published literature, etc.

Field structures

Particular attention has been devoted to the definition of the fields so that they are neither ambiguous nor overlapping. Duplication of input is to be avoided. Although subfields can be used, even at several levels in a hierarchical system, it was usually found to be preferable to distribute the data over separate fields. The various kinds of fields can be categorized by their function and content.

With respect to their function, some fields differ from others by being obligatorily required by the internal structure of the data bank, with contents automatically generated or not. These fields are grouped under 'internal administration'. The other fields are data fields, three of which are obligatory, otherwise a record cannot be entered: the record type, the species name (in SP, SYN and ALT records), and the strain number (in STR records).

Data fields can be grouped according to their subjects: those concerning the taxonomic name, strain administration, status, environment and history, biological interactions, sexuality, properties, genetics, growth conditions, chemistry and enzymes, and practical applications. Such a grouping is not of absolute value but aids the rational ordering of the data fields.

Some fields pertain only to species records (SP), some to strain records (STR), and some to both. Fields pertaining either to SP records or to STR records may be independent of each other (e.g. species name and strain number), but in other cases either the same field or a pair of separate fields may be defined for comparable data on species and on strains to facilitate indexing. In these cases the SP record contains a general feature, while in the STR record actual observations are given. For example, TEMPR in species records is indexed, whereas TEGR in strain records is not; as in the pairs PHR-PHC and SALR-SALC. When data in fields of the ecology, property, or chemistry blocks are entered in a SP record rather than in a STR record, this implies that either: (a) a characteristic is known to apply to all strains of a species (e.g. TEMPR: thermophilic, PHR: acidophilic); or (b) an important feature has not yet been associated with any particular strain held in the present collection and need not apply to all strains of the species. Similar data may therefore appear in both SP and STR records, although factual data are preferentially entered in the STR record, with the risk of repetitions.

Subfields are defined to allow several entries in a field. This structure also allows the subfield contents to be indexed individually (sequential entries at equal rank, unless free-text indexing is required); parts of the contents can also be excluded from indexing (truncating function, subfields of lower rank). Any delimited part of a field may be indexed and/or used for printing of catalogues. (The facility of BASIS software to distinguish up to 12 ranks of subfields is therefore a great advantage.)

In the present format, as implemented in BASIS, several subfield delimiters have been adopted to provide a structured system. The criteria of choosing delimiters are summarized as follows.

comma, ',' – used to separate authors' names from organism name in SP (truncation); used to separate state from country in LOC (sub-subfield).

semicolon, ';' – used to separate sequential entries in many fields; also used to separate indexed information from non-indexed data in LOC and SSTR (truncation).

colon, ':' – used to separate unequivalent but related parts of information (to be indexed separately at the same rank), such as reaction and symptoms from target organisms in PATH, ALLER, MPAR, TOX and ANT; also used in HEATR to separate temperature from time of treatment.

less than, '<' (interpreted as an arrow pointing left to the main information) – used to separate remarks and data on origin from indexed information (truncating function).

parenthesis, '(' – similar to '<', used to separate minor details from rest of non-indexed information (in COLL, ISOL, HIS).

slash, '/' – used to separate values for minimum, optimum and maximum temperatures and substrate concentrations.

The database will mainly comprise data available from published or unpublished sources on the particular strains held in the collection. It is generally not intended to administer stocks and orders with this database, although the system allows this to be done. The strain database need

not exhaustively cover all the literature on taxonomy, nomenclature, ecology, chemical properties, etc., but only as far as the particular strains are concerned. It is also not the purpose of this strain database that it be used for identification. Therefore it will contain relatively few phenotypic data, in contrast with the RKC system. Identification of yeasts might, however, be possible using patterns of growth and fermentation from a selection of fields where these data are recorded, but other specialized software more suitable for this purpose is available.

Reference to the sources of data

Literature references or references to other sources of data can be entered at three levels.

(1) For species and strains in general: in the field LIT designed for references on a species or strain in general.

(2) For a group of fields: in group fields ECOLIT, PATHLIT, PROPLIT, GENLIT, CHLIT and APLIT.

(3) For any particular field or subfield. This information is entered after a delimiter symbol (e.g. '<'), the source of information being indicated by 'I' for work done at the same institute, and 'D' for information supplied by the depositor. Otherwise, literature references (possibly abbreviated) can be entered here. Information after the delimiter is not indexed and can also easily be skipped in catalogue printing. To avoid the duplication of entries, full references should be given only once at level 1 or 2, and an abbreviated citation at the lower level(s). By mapping, all relevant references can be displayed in one command.

Another option in referencing literature is to refer by number to an entry given in full in a bibliographic record or database.

THE PROPOSED FIELDS

In the following list, all fields are grouped into 12 blocks so that fields of SP records and related fields of STR records appear adjacent to each other. Each field is characterized by the field tag (label, prefix) and a short circumscription. When subfields are to be used, this is indicated. The fact that a field is indexed is indicated with a bold I, its obligate presence with a bold O. A bold A indicates automatic generation of the contents. Fields included in the minimum data set (MDS) for exchange between national nodes are marked with an asterisk (*).

A. INTERNAL ADMINISTRATION

ACCN Accession number of the record: **O I**

Occurs automatically in all records, with a unique number.

APOS Alternate posting: **O I**

Must occur in all STR, SYN and ALT records to connect them with the name of the species and other species data. Contains the accession number (ACCN) of the connected SP record, but a SYN record can also refer to a connected ALT record.

RT Record type: **O I**

Must occur in all records. There are four possibilities: SP = species; STR = strain; SYN = synonym; ALT = alternate.

EDI Date of input: **I A**

Occurs in all records.

EDM Date of update: **I A**

Occurs in all records after an update. Date of last update.

- SEC Security code:
May occur in any record. Prevents publication of a record in a hard copy catalogue or external access to it. No entry = open access, otherwise the following three categories are distinguished: 2 = not catalogued but open to access by anyone; 3 = access only by privileged users; 4 = only accessible by privileged staff members.
- ID Typist's identification: A
Occurs in all records.
- OT Organism type: O A
Occurs in species records. Indicates type of databank. B = Bacteria, F = Filamentous Fungi, Y = Yeasts.

B. NAME

- SP Species name, authors: O I *
Occurs in all SP, SYN and ALT records. Contains the genus name and species epithet, author(s). Authors' names are separated from the species name by a comma delimiter. The part of the field after the delimiter is not indexed and can be printed in a different type face to the name.
Authors' initials are used when necessary to distinguish several authors with the same name. Authors' names are preferably not to be abbreviated. If an abbreviation is used, it should be formed according to the list of Hawksworth (1980). If more than two authors are connected to a name, the first one is preferably cited followed by "et al." rather than citing all, separated with comma, and the last two with "&". Collections are free to add the year of publication after the author's name, with or without the bibliographic reference to the place of first publication of the name in a subfield. The thesaurus of mycological names provides a check for synonyms and allows a taxonomic orientation in the higher ranks.
Examples: Abortiporus biennis, (Bull.:Fr.) Singer
Fusarium oxysporum, Schlecht.:Fr.
- SSP Subspecies name, authors: I *
Occurs in SP, SYN and ALT. Name, author(s), and delimiter as under SP. Contains subspecies epithet and (if different from species epithet) author(s).
- VAR Variety name, authors: I *
Occurs in SP, SYN and ALT. Name, author(s), and delimiter as under SP. Contains variety epithet and (if different from species epithet) author(s).
- F Form name, authors: I *
Occurs in SP, SYN and ALT. Name, author(s), and delimiter as under SP. Contains forma epithet and (if different from species epithet) author(s).
- FSP Special form name, authors: I *
Occurs in SP, SYN and ALT. Name, author(s), and delimiter as under SP. Contains epithet of forma specialis and author(s).
- RACE Race name, authors: I *
Only in STR. Only the number is entered. The author(s) may be indicated after "<".

- SERO** Serotype, serovar
Only in STR.
Example: Serotype A.
- MIS** Misapplied name, authors: I *
Occurs in STR records to indicate names that have been applied to the strain but which are not synonymous with the adopted name (same structure as field SP).
Example: [in Acremonium chrysogenum CBS 779.69]
Cephalosporium acremonium, Corda
- TAX** Taxonomy
Occurs only in SP. Contains names of suprageneric taxa. Eventually this field will be superseded by a taxonomic thesaurus based on Hawksworth et al. (1983), which is available in BASIS, and therefore will not be entered in this database.
Examples: Basidiomycotina, Agaricales, Pluteaceae
Ascomycotina, Pezizales, Pezizaceae
- MOR** Morph
Occurs in SP, SYN and ALT. Indicates the morph represented by a particular name but not the morph actually developed in the culture. Used only for ascomycetes, basidiomycetes and deuteromycetes. There are only two possibilities: A = anamorph; T = teleomorph. May be generated automatically from the taxonomic thesaurus or from field TAX.
- ILL** Illustrations
Occurs in STR and SP. Reference number to an envisaged graphic database in which figures may be stored, or to illustrations in literature from a bibliographic database.
- TAXREM** Taxonomic remarks
In SP. Refers, e.g. to
-- alternate names, if no specific epithet is available to create an unequivocal ALT record,
-- synonyms or morphonyms, if the relation is equivocal.
-- Cross-references referring to other databases are also indicated here.
Indications that the taxonomic position of a particular strain is dubious, are entered in REM.
Examples: [Geotrichum in Yeasts] see Filamentous Fungi
[Preussia multispora] anam. = Phoma sp.
[Eupenicillium javanicum] cf. also Penicillium simplicissimum.
- LIT** Taxonomic and other literature related either to a species or a strain in general (also possible for entry to be the record number of a full reference in a bibliographic database). In SP or STR.
Examples: Boidin, J. et al. Mycotaxon 9: 45-98. 1984; Lanquetin, P. Bull. Soc. mycol. Fr. 87: 387-412. 1982.

C. STRAIN ADMINISTRATION

- STN** Strain number: O I *
Must occur in all STR records. The number only has to be entered (for strains kept in affiliated collections, also the collection acronym or initials). But in retrieval, display, or editing, the number is always

preceded by the collection acronym appearing as field label. The STN must be unique.

Examples: 376.85 (for filamentous fungi at CBS)
1422 (for yeasts at CBS)

- OCC Numbers in other collections: I *
- Occurs in STR. Collection acronym or initials and number. Numerous entries in subfields possible, delimited by ";".
- Examples: ATCC 12564;IMI 127309
- Collector's numbers are to be entered in field COLL or ISOL; if numbers of personal collections are entered at the end of field OCC, then they should be separated from the remainder by a delimiter to exclude them from the index (the full names, or initials, each followed by a full stop, should then be used).
- EDA Date of accession: I *
- Occurs in STR. Format Mmm YYYY (day not normally entered).
- Example: Mar 1982
- EDR Date of receipt: I
- Occurs in STR. Day, month and year to be entered.
- CORR Correspondence
- Occurs in STR records. Refers to letters in archive(s).
- Examples: Blakeslee, 15 Mar 1904.
 letter 415/03111985
 letter 322/17 Mar 1987.
- SPINF Special information available: A *
- This is a field of STR records, automatically generated, which has been added to the Minimum Data Set in order to indicate that special information is available, particularly on chemical properties. An indication, identical with or similar to the field tag, is introduced automatically during VALIDATE, if one of the data fields of blocks F, I, J, K or L contains information.
- REM General remarks
- Only in STR. The dubious taxonomic position of a particular strain is also to be entered here.
- Examples: unusually large conidia
 abundant production of ascospores
 deviating from the diagnosis by larger ascospores.
- CHK Internal check of identity at collection
- Only in STR. Initials, date. The check implies that the strain is living, pure, and of correct identity.
- Example: JAS, 5 Jun 1972; JAS, 18 Dec 1986.
- PRE Mode of preservation / storage: *
- In STR (also in SP if a technique not yet or only partially applied in the actual collection is found suitable elsewhere). The date of preservation is not entered in this database, but in the one dealing with stock management. AG = agar slants (interval of transfers in parentheses); LY = lyophilized (the cryoprotectant in parentheses); LN = liquid nitrogen (the cryoprotectant in parentheses); CR = cryopreservation at -135°C (the cryoprotectant in parentheses); FR = frozen at -70 to -80°C (the cryoprotectant in parentheses); LD = L-drying; MO = sealed with mineral oil; WA = under water; SG = silica gel.
- Examples: AG (6 mo), LY, LN (10 % glycerol), MO
 WA, LN (DMSO)

PREREM Remarks on preservation

In STR (in SP when results of other workers with other strains are reported). Special care needed, results of preservation experiments, or remarks on the general ability to survive, sporulate etc. after preservation. The field is, however, not intended to keep a regular record of the checking of the maintained stock of strains.

Examples: cannot be preserved in LN on ChA

does not endure storage at 4°C

requires DMSO for LN

LY- [does not survive lyophilization]

RESTR Restrictions, care and precautions

In SP (rarely in STR). Concerns strains but, when all strains of a taxon are involved, is entered in SP.

-- precautions for internal policy and distribution of cultures

Example: Transfer only in biohazard laminar flow cabinet

-- restrictions for distribution and export of cultures

(a) Pathogenicity class (WHO standard and/or national standard), recombinant category, and handling precautions

(b) distribution and transportation restrictions and precautions
Example: Not to be released to unauthorized customers.

(c) import regulations

restrictions imposed by laws of the country where the collection is located, e.g. in relation to plant quarantine

(d) export regulations, restrictions imposed by laws of other countries, e.g. required permits in the addressee's country and the kind of wrapping required in connection with WHO classification.

D. STATUS**STAT** Status (indicating type or authentic strains): I *

Occurs only in STR records. As living fungal cultures cannot have the formal nomenclatural status of a type specimen, the authenticity of any isolate derived from a type specimen or yielding a dried type culture (ex-type) is indicated as follows: T = ex type strain (in general); HT = ex holotype strain (if the holotype status of the associated specimen is to be indicated explicitly); NT = ex neotype strain; LT = ex lectotype strain; IT = ex isotype; ST = ex syntype; PT = ex paratype; PTT = pathotype; AUT = authentic strain; OR = original strain; REFS = reference strain.

Status either refers to the name mentioned in the NAME fields, and then only one of the above abbreviations is given, or it refers to another name of the taxon, then that other name with author(s) is given, preceded by "of". References to literature on typification may be added after a delimiter ("<").

Examples: T of Penicillium psittacinum Thom

[under Arthrotrrys entomopaga] NT and T of Arthrotrrys pauca.

E. ENVIRONMENT and HISTORY**SUSP** Substratum specificity: I

Only in SP. Indicates restriction to certain substrata, when consistent for a particular species. Names of host plants/animals according to thesaurus for substrata. Several entries in subfields possible,

delimited by ";". Literature references to be entered after a delimiter ("<>").

Examples: scale insects
wood of Abies

SSTR Original substratum: I *

Only in STR. Indicates substratum/host on which the fungus was found. Names of host plants/animals according to thesaurus of substrata. If information on the biotope is associated with a particular substratum on which a fungus is found, then it should be entered here, otherwise in field HAB; the location should be entered in field LOC. Main information (what is to be published in catalogue) first, to be separated by ";" from less important information; the latter is not indexed and catalogued.

Examples: wood of Quercus robur
mosses; along a ditch
wheat-field soil; 10 cm below surface

HAB Habitat (original biotope)

In SP and STR. Indicates the biotope where the fungus is found, particularly of interest for the biotope of fruiting-bodies of macromycetes.

Examples: beech forest
limestone grassland

LOC

Origin / location of original material (country, locality): I *

Only in STR. Sequence: Country, state; local details. A major delimiter ";" may separate subfields in such a way that only the first part is indexed and published in catalogues. Spelling of country names in English, other names according to national spelling (thus "Hannover, München, Lisboa, Warszawa, Moskva"), cyrillic names are transliterated according to ISO norm 833-1974. A thesaurus for geographical names helps to select the correct names and permits the use of abbreviations.

Examples: Canada, Quebec; Montreal, La Salle Woods, 400 m alt.
Netherlands; Baarn, Estate Groeneveld, near ditch.

COLL Collector

Only in STR. Gives name of the collector with initials, original number of specimen, date of collecting. Information separated by commas (or parentheses). Collector's initials, when used in the number, are followed by dots, no spaces.

Example: F.F.Lombard, F.F.L. 1298, 18 Sep 1965.

HERB Herbarium where specimen is deposited

Only in STR. Gives herbarium acronym (without dots) and number (no comma or period to separate thousands). The kind of specimen may be entered between parentheses as "N" = specimen (or microscopic slide) from nature, or "C" dried culture (or slide made from it) or other text.

Examples: L 153876-342
IMI 2367145
CBS 1857 (C; original material)

ISOL Isolated by: *

Only in STR. Name of isolator, original number of isolate, date. Information separated by commas (or parentheses). Mycologist's initials, when used in the number, are followed by dots, no spaces (see COLL).

Example: I.S.Nogood, I.S.N. 1548, 16 Oct 1982.

ISOM Source and method of isolation

Only in STR. Source (single spore, mycelium), technique, and medium (in parenthesis).

The field contains

a) the kind of inoculum (source):

SS = single sexual spore; MS = mass sexual spores; SA = single asexual spore; MA = mass asexual spores; MY = mycelium; RH = rhizomorphs; SCL = sclerotia; STR = stromatic tissue; AS = ascoma; BA = tissue of basidioma; or generally CARP = carpophore.

b) the method:

DP = dilution plate; DC = damp chamber; SU = surface sterilization; DT = direct transfer; EM = enrichment method; BAIT = baiting (indicate substrate); PAST = pasteurization.

Several entries are separated by commas.

The isolation medium (abbreviated, as in field CONDS) may be given in parentheses.

Examples: SS (Hagem)

BA, DT (MEA + benomyl)

DET Identified by: *

Only in STR. Name; if a second mycologist ("redet. NN") reached a different identification, also the date (year) is to be entered between parentheses. After the delimiter "<" (a) the natural/cultural material on which the identification was based, in particular NAT = material collected from nature, CULT = cultured material, TELEO = teleomorph, ANAM = anamorph; (b) the method used, e.g. DNA hybridization, crossing, morphology, and physiology can be entered. Determination is meant to imply a critical taxonomic study of the organism and not just a renaming according to the literature (the latter to be entered only in NAMCH).

Examples: V. Demoulin <NAT TELEO

T. Hoogenboom, redet. R.A. Samson (1983) <CULT

DEP Depositor: I *

Only in STR. Individual name or Institution, abbreviated (see McGowan & Skerman, 1982; Pridham, 1981; CBS List of Cultures, 1987, p. 10; Hawksworth & Kirsop, 1988) without period. Two depositors may be cited, separated with "and". A date is not given here but in field EDA. Mycologist's initials need not be separated from the name by a space, they are not indexed. After a delimiter ("<") the depositor's strain number and/or depositor's institute may be entered.

HIS History (between isolation and deposit): *

Only in STR. Fate of the isolate between isolation and deposit in the present collection. The forward sequence of deposits is used: each entry thus may contain the name of collection, (month and) year of acquisition and, between parenthesis, possibly another name, if the name was changed in that collection, the delimiter ">" meaning "sent to".

Example: [Fusarium chlamyosporum, CBS 124.73] S.I.Ahmed > 1971
CMI (Fusarium fusarioides) > 1973 CBS

NAMCH Name changes / previous names: *

In (SP or) STR. Name changes introduced at the present collection.

Examples: received as Fomes ignarius

Until 1977 preserved as Trichoderma viride

ECOLIT Literature on ecology

In SP and STR.

ECOREM Remarks on ecology and environment
In SP and STR.

F. BIOLOGICAL INTERACTIONS

SYMB Symbiosis

In SP and STR. Only for mutualistic or commensalistic symbioses. Type of symbiosis (lichen, ecto- or endomycorrhiza), scientific name of symbiont or of group of symbionts, separated by "with".

Examples: Endomycorrhiza with Neottia
Ectomycorrhiza with Abies alba, Picea abies
Lichenized with Coccomyxa

MPAR Mycoparasitism: I

In SP and STR. Qualification of the parasite or symptoms, and host (indicated by scientific name), separated by ":". The host used for growing obligate mycoparasites is to be entered in field CONDS.

Example: Destructive mycoparasite: Rhizoctonia solani

PATH Pathogenicity: I *

In SP and STR. Name and/or symptoms of disease, susceptible group of organisms, separated by ":". If possible use scientific names of organisms. The names are organized through the thesaurus file of substrata. Use singular for the organism.

The expression "causing adiaspiromycosis in mice" is entered as: "adiaspiromycosis: mice". Several entries are possible in subfields. Further example: leaf spot: Hordeum; sheath rot: Triticum

ALLER Allergenicity: I

In SP and STR. Type of allergy and susceptible group of organisms delimited by ":", (usually "man", which is not entered). Several entries possible in subfields, separated by ";".

Examples: [Cladosporium herbarum:] asthma
skin irritation: rabbit

TOX Toxicity to other organisms: I

In SP and STR. Symptoms, susceptible group of organisms, organs and degree of toxicity. Toxic metabolites are normally listed under MET but the toxin involved may be given here between parenthesis besides field MET).

The word "causing" need not be entered but is indicated by the delimiter ":", as in PATH. Use singular for the organism. After the susceptible organism a "<" delimiter can be placed so that the further field contents are not indexed. Several entries are possible in subfields, separated by ";".

Example: Skin lesions:rabbit < at 0.01 % (HT-2 toxin).

ANT Antagonistic activities against other organisms: I

In SP and STR. Action and name of susceptible organism, separated by ":", responsible mechanism, separated by "<". Nematophagous activities are also entered here.

Examples: Growth inhibition: Heterobasidion annosum
capturing nematodes <adhesive network

PATHLIT Literature on pathogenicity and other biological interactions
In SP and STR.

PATHREM Remarks on pathogenicity
In SP and STR. Including special conditions for pathogenicity, notes on severity, etc.

G. SEXUALITY

SEX Sexual behaviour: *
Occurs only in SP. Indicates the type of sexual behaviour. HO = homothallic; A = amphithallic; HE = heterothallic; HE1 = heterothallic unifactorial (bipolar); HE2 = heterothallic bifactorial (tetrapolar). Two indications may occur in one record.
Example: HO, HE.

SEXST Sexual state: *
Only in STR. Indicates the actual sexual condition of the strain. MK = monokaryotic; HOK = homokaryotic; HA = haploid; HEK = heterokaryotic; DK = dikaryon; DP = diploid; MT = mating type and code of mating type (male, female, +, -, a, A, alpha, A1), mating partners used to assess the reaction to be entered in REM; ± = crossed mating types (parental strains may be entered here). "HO" (homothallic) may occur in this field when a particular strain of a normally heterothallic species is homothallic. The indication "MK" is superfluous when the mating type is specified.
Examples: [MK,] MT A1B1.
MT +
± (CBS 112.08 x 116.16).

H. PROPERTIES (cytology, biomolecular data)

NUCL Number of nuclei
Only in SP and STR. Nuclear behaviour, according to Boidin's (1964) classification (in SP) or number of nuclei per cell (in STR).
Examples: 2 ("nuclei per cell" not entered).
3-15
coenocytic
astatocoenocytic

FSTR Fine structure data
In SP and STR. E.g. wall-layering, septal pores.
Examples: cupulate parentheses<D
TEM ascosporeogenesis< van Brummelen, 1985

WCONS Wall constituents
In SP and STR. Name of component, percentage, (technique used to be entered in parentheses; specify, whether whole-cell preparations or purified cell walls were used). Names of compounds standardized according to chemical thesaurus. Several entries possible in subfields, separated by ";".
Examples: glucose 60;mannose 15 (whole cell hydrolysate).

CCONT Cell contents
In SP and STR. Contents including lipids, reserve substances, quinones and other substances. Names of compounds standardized according to chemical thesaurus. Several entries possible in subfields, separated by ";".
Example: glycogen;lipid

- COQ** Coenzyme-Q system: I
In SP and STR. The number of the dominant Co-Q is entered, with percentage, if available, between parentheses (omit "%"). The other Co-Q detected, with percentage in subfields separated by ";" .
Example: Q8 (85); Q10 (15)
- STAIN** Staining reactions
In SP (and STR). Enter reagent and reaction with +, - or v. This field also includes cell wall reactions with Diazonium Blue B.
- PIGMENT** Pigment production and autofluorescence
In SP and STR. Colour of pigment, name (if known, also to be entered in MET), (non-)diffusing, fluorescence (wave length), medium. Several pigments separated by ";" .
Example: red, cynodontin, dif, OA.
- GRWTH** Growth characteristics
In STR, rarely in SP. In particular, radial growth on agar measured in millimetres in relation ("/") to time, temperature and other conditions. Data on quantitation of biomass (dry weight, duplication time etc.) in submerged culture also to be entered here.
Example: 15 /5 d, 20 C, PDA, dark; ...
- TKIT** Results of test-kit: I
In STR. Test-kits as opposed to single tests (entered in fields CSOR, NSOR, ENZ, NUGR; e.g. the API test (Dermóumi, 1979; Bridge & Hawksworth, 1984; Wasfy *et al.*, 1987). Strip indication - colon (":") - code. But individual results from API 20 C can also be entered in field CSOR and results from API ZYM in ENZ. Several tests make subfields delimited by ";" .
Example: API 20 C: 2 764 774.
- PROPLIT** Literature on morphological, taxonomic, fine structure characteristics.
In SP and STR.

I. GENOTYPE AND GENETICS

- GENOT** Genotype
Only in STR. The word "genotype" is to be entered.
Example: [Saccharomyces cerevisiae] genotype MAL4,ade1
- GC** Guanine-cytosine content of DNA
Usually only in STR. Percentage of GC in DNA (without "%").
Example: 43.2.
- DNAD** Hybridization DNA-DNA with other strains
Only in STR. Strain number, percentage binding; different data in subfields separated with ";" (omit "%").
Example: CBS 324.78, 93; CBS 752.82, 98
- DNAR** Hybridization DNA-r.RNA with other strains
Only in STR. Strain number, type of RNA, temperature of denaturation [Tm(e) in °C]; different data in subfields separated with ";" .
Example: CBS 1241, 5sRNA, 65.

- RNAD** Hybridization r.RNA-DNA with other strains
Only in STR. Type of RNA, strain number, temperature of denaturation; [Tm(e) in °C]; different data in subfields separated with ";".
Example: 5rRNA, CBS 1241, 75.
- MUT** Mutants: *
Only in STR. The word "mutant" or equivalents (like mutated characteristics) are to be entered. May also indicate the number and/or name of the parent strain. Mutant strains are accessioned with strain numbers distinct from those of the parent strains. After the delimiter "<" author and date can be included.
Examples: white mutant of CBS 324.74
[Phycomyces blakesleeanus] mut. gracilis Burgeff
- MUTMET** Method by which mutation was induced
Only in STR. May also indicate the author of the mutation and date.
Examples: UV 30 min
10 % isothiocyanate 10 min
- HYBR** Hybrids
Only in STR. In SP records, the hybrid nature appears in field SP (with an "x") according to the International Code of Botanical Nomenclature. Strain numbers and/or names of parents with "x". Author, also date and methodology may be entered.
Examples: CBS 393.36 x U. levis CBS 158.22 -.*
U. levis mut. buff-smut x U. avenae
- PLS** Plasmid out/in the host
Only in STR. Subfields contain information about
a. plasmid name
b. plasmid type
OP = original plasmid; CV = cloning vehicle; RP = recombinant plasmid
c. derived from plasmid (name)
d. unique restriction sites
e. relevant genes
OR = origin of replication; TP = transposons; PM = promoters; TM = terminators; SG = structural genes
- KIL** Killer properties of yeast: *
Only in STR.
Examples: killer
flocculant killer
killer-sensitive
- GENREM** Remarks on genetics, DNA data, mutations, plasmids
Only in STR.
- GENLIT** Literature on genetics and mutants
Normally only in STR. References to literature containing data on genetics and mutants.

J. GROWTH CONDITIONS

Remarks, together with source of information, may be entered after the delimiter "<" in each field. Negative (or uncertain or variable) test results may be entered too in the appropriate field with a negative sign "-" (or "v"), or under PREREM if maintenance is concerned. A positive reaction is default.

- CONDS** Conditions for growth and maintenance on solid media: *
Mainly in SP, but also in STR. In SP media to be used for all strains of a species.
Media (+ additives), separated by ",", then temperature and light conditions, separated by ";". Media can be referred to by means of abbreviations (acronyms) or numbers. Abbreviations for media used in the first subfield are organized through the thesaurus of media.
- CONDL** Conditions for growth in liquid media
In SP (and STR). Media (+ additives); temperature, light, shaking. Abbreviations or numbers as in CONDS.
- CONDSP** Conditions for fruiting or sporulation
In SP (and STR). Media (+ additives); temperature, light, gas phase. Abbreviations or numbers like in CONDS.
- CONDGER** Conditions for germination
In SP (and STR). Media (+ additives); temperature, light, gas phase. Abbreviations or numbers like in CONDS.
- CSOR** Carbon sources that have been tested for growth: I
Only in STR. Name of compound and result of test (+, -, v). The full names of the compounds are entered. Abbreviations to facilitate input or CAS numbers are to be organized through the thesaurus for chemicals. Several entries form subfields, separated by ";".
Example: glucose; galactose; sorbose/L-; sucrose; maltose; mannitol; citric acid -.
- NSOR** Nitrogen sources that have been tested for growth: I
Only in STR. Name of compound and result of test (+, -, v). The full names of the compounds are entered. Several entries form subfields, separated by ";". Hydrolysis of urea as such should be entered in ENZ. AM = ammonium; NA = nitrate; NI = nitrite; AA = amino acids; ON = organic nitrogen compounds. Further abbreviations to facilitate input or CAS numbers are to be organized through the thesaurus for chemicals.
- CNSOR** Single compounds tested as sole carbon and nitrogen source for growth: I
Only in STR. As in CSOR and NSOR.
- NUGR** Nutritional requirements and growth factors required: I
In SP and STR. Amino acids, trace elements, growth factors (vitamins), blood, serum. The ten vitamin tests commonly used for the characterization of yeasts are to be entered here. Several entries form subfields. Abbreviations to facilitate input are to be organized through the thesaurus for chemicals.
- DEF** Deficiencies: I
In SP and STR. Like NUT.
Name of compound with concentration between parentheses. To be used for incidental or mutational deficiencies, other than the regular growth factor requirements in field NUGR; characteristics of deliberately induced mutants to be entered in field MUT. Abbreviations for input to be organized through the thesaurus for chemicals. Several entries form subfields, delimited by ";".

- TOL** Tolerances and sensitivities: I
 In SP and STR. Agent: concentration (give unit), growth response. Several entries form subfields, delimited by ";". The delimiter ":" after the agent has the effect that only the first part of the (sub-) field is indexed.
 Tests commonly used for yeasts include the tolerance to cycloheximide, acetic acid. Use chemical names, not commercial names (cf. thesaurus of chemicals).
 Examples: Benomyl: 10 ppm, +
 PCNB: 0.01 %, + but atypical growth.
- TEMPR** Temperature relationships for growth: I *
 Only in SP. PS = psychrophilic (optimum < 20°C); TH = thermophilic (minimum > 15°C, optimum 35-45°C); ET = extremely thermophilic (minimum > 20°C, optimum 40-45°C); ME = mesophilic (remaining range).
 The definitions of the terms may deviate for other organism types.
- TEGR** Cardinal temperatures for growth
 Only in STR. Single value or range (no "degree C") in three subfields for minimum, optimum and maximum temperature, separated by "/". But recommended temperature for preservation in field CONDS. If a subfield contains no data, care must be taken to use the proper number of slashes before entering further data. If growth has been tested at a particular temperature without establishing the range, the temperature is indicated with the response (+, -, v), different values delimited by ", ". Remarks, together with source, may be entered after "<".
 Examples: 12.5/ 20-24/ 30 <D
 /12-14 [= optimum 12-14°C]
 //44.5 [= maximum 44.5°C]
 37+,40- [= growth occurring at 37°, but not at 40°C]
- TESP** Temperature optimum range for sporulation
 In SP and STR. Subfields may be used for minimum, optimum and maximum, separated by "/". If sporulation has been tested at a particular temperature without establishing the range, the temperature is indicated with the response (+, -, v), different values delimited by ", ". Remarks, together with source, may be entered after "<".
 Examples as under TEGR.
- HEATR** Heat resistance
 In SP and STR. A "+" in the SP record indicates heat resistance in general (the capacity of a species to survive at 60°C for at least 15 min). In the STR record, temperature/time are entered (separated by ":"), growth response: to survive with "+", to kill with "-"; between parentheses the survival structure may be entered.
 Examples: 60: 30 min + (conidia and ascospores); 70: 20 min + (ascospores); 70: 15 min - (conidia)
- PHR** pH requirements for growth: I
 Only in SP. EA = extremely acidophilic (growth below pH 3.0); AC = acidophilic (optimum pH 3.0-4.5); NE = neutrophilic (optimum pH 5.5-7.5); AL = alkalophilic (optimum pH 7.5-9.0); EAL = extremely alkalophilic (growth above pH 9).
 The definitions of the terms may deviate for other organism types.
- PHC** pH conditions
 Only in STR. Value or range in three subfields for minimum, optimum, and maximum pH, separated by "/". If a subfield contains no data, care

must be taken to use the proper number of slashes before entering further data. If growth has been tested at a particular pH without establishing the range, the pH is indicated with the response (+, -, v), different values delimited by ",". Remarks, together with source, may be entered after "<".

Examples: 4.5/5.5-6.5/7.3

//8.0

7.8+, 8.5-

ETHC Ethanol conditions

In SP and STR. Optimum and maximum concentrations of ethanol for growth, separated by "/". If a subfield contains no data, care must be taken to use the proper number of slashes before entering further data. If growth has been tested at a particular concentration without establishing the range, the concentration is indicated with the response (+, -, v), different values delimited by ",". Remarks, together with source, may be entered after "<".

SALR Salinity requirements for growth: I

In SP. M = marine (0.35%); H = halophilic (> 0.4%); EH = extremely halophilic (> 20%).

SALC NaCl concentration (specified other salts may also be entered)

In STR, values or ranges in subfields for minimum, optimum and maximum for growth, delimited by "/". If a subfield contains no data, care must be taken to use the proper number of slashes before entering further data. Concentrations in percent (without entering "%", but with indication of w:w, v:v or w:v between parentheses). If growth has been tested at a particular concentration without establishing the range, the concentration is indicated with the response (+, -, v), different concentrations separated by ",". Entries for different salts are to be separated by ";". Remarks, together with source, may be entered after "<".

Examples: 2/5-8.5/14 (w:v) [NaCl minimum concentration 2 %, optimum 5-8.5 %, maximum 14 %].

KCl //5 [maximum concentration of KCl 5%].

SUGC Optimum and maximum sugar concentrations

In STR. Names of sugars and optimum and maximum concentration for growth in % (without "%", but with indication w:w, v:v or w:v between parentheses). Optimum and maximum are separated by "/". If a subfield contains no data, care must be taken to use the proper number of slashes before entering further data. If growth has been observed at a particular concentration without establishing the range, the compound and concentration are indicated with the response (+, -, v), different concentrations separated by ",". Entries for different sugars are to be separated by ";". Remarks, together with source, may be entered after "<".

Examples: saccharose 5-20/40 (w:v)

glucose 20, glycerol /25 (w:w). [optimum glucose concentration 20 %, maximum glycerol concentration 25 %].

saccharose 3/4.5 w:v/5- [optimum 3 %, maximum 4.5 %, no growth at 5 %].

saccharose 3/5 (no sporulation) [optimum 3 %, maximum 5 %, but at 5 % no sporulation].

OSM Osmophily and xerophily: I

In SP. XE = xerophilic (minimum water activity for growth < 0.9), EX = extremely xerophilic (minimum water activity below 0.72).

WATC Conditions of water activity
In SP and STR. Optimum and minimum a_w for growth and, between parentheses, osmoticum used. Specify units used.
Example: optimum a_w 0.95, minimum a_w 0.85 (PGS) <Northolt *et al.*, 1976. [The abbreviation PGS stands for a peptone-glucose-salts medium used by the authors, viz. NaCl:KCl:Na₂SO₄ = 5:3:2].

LC Light conditions
Only in STR. In this field specific values of Brightness [in L (lux) or E (Einstein)] and wavelength are entered in three subfields for minimum, optimum and maximum (separated by a slash). If a subfield contains no data, care must be taken to use the proper number of slashes before the data. The use of day light or black light for inducing sporulation is entered in CONDSP (or CONDS).

K. CHEMISTRY and ENZYMES

ENZ Enzymes produced: I
In SP and STR. Name (or EC number). Several entries form subfields, delimited by ";". Remarks after the delimiter "<" may also include optimal conditions for production and the information whether the enzyme is inducible. Prefixes (1-, 2-, D-, L-, alpha-, beta-) to be entered after the name, separated by "/". "alpha" for "α", "beta" for "β" to be spelled fully.
Note that urea hydrolysis should not be entered here but in NSOR. The official name will be entered in the database but EC numbers and synonymous names are acceptable and conversion of the input will be organized through the thesaurus of enzymes (based on Enzyme Nomenclature, 1984).
Example: an input of 3.5.1.5;3.4.21.37;3.2.1.20;3.2.1.26;2.4.1.34 would be translated by the thesaurus into
urease;elastase;glucosidase/alpha-;invertase;UDP-1,3-beta-D-glucan glucosyltransferase

DEC Decomposition and deteriorating capacities: I
In SP and STR. Several entries possible as subfields, delimited by ";". Polymeric substrates, like starch, cellulose, xylan, lignin, gelatin, chitin, keratin, arbutin, fat (simpler, chemically defined compounds in CSOR and NSOR, specific kinds of wood generally to be entered in SUSP). Chemically defined compounds that are degraded but not used for growth can be entered here too. Terms like "amylolytic", "cellulolytic" are not to be used but the respective substrates, such as starch, cellulose etc.

MET Metabolites produced (including antibiotics, toxins, also industrial fermentation products): I
In SP and STR. Several entries form subfields, delimited by ";". The names of the compound are entered in the database without abbreviations. Abbreviations or CAS numbers to facilitate input and retrieval are to be organized through the thesaurus of chemicals. After the delimiter "<" the substrate utilized (with "from"), optimal conditions for production and/or origin of information can be entered in each subfield. The fermentation substrate is also to be entered in DEC and, if used at large scale, in APPL. Chemical transformations that do not involve degradation are entered in BIOTR.
Prefixes (1-, 2-, D-, L-, alpha-, beta-) are to be entered after the name, separated by "/"; "alpha" for "α", "beta" for "β" to be spelled fully.

Examples: ethanol <from glucose
lactic acid <homofermentation
lactic acid < from lactose, heterofermentation

BIOTR Biotransformations

In STR (and SP). Several entries form subfields, delimited by ";".

Examples: C-1,2-dehydrogenation of 5-beta-pregnane-3,11,20-trione
detoxification of aflatoxin

CHREM Remarks on chemical data

In SP and STR

CHLIT Literature on the impact of chemical and physical factors

In SP and STR

Reference or reference number for literature database containing data on metabolites, enzymes and constituents

L. PRACTICAL APPLICATIONS

APPL Industrial and general applications (including fermented foods): I *

In SP and STR. Keyword, description. Several entries may form subfields, delimited by ";".

Examples: production of ontjom
transformation of hydrocortisone
production of penicillin in submerged culture
amylo process in starch saccharification

TEST Applications in testing (assay): I *

In SP and STR. Several entries may form subfields, delimited by ";".
Assay for amino acids, antibiotics, vitamins, trace elements, fungicides, toxicity, testing materials. The expression "used for" applying to all entries need not be entered.

Example: biotin assay.

BIOM Production of biomass, edibility: I

In SP and STR. Conditions, growth, yield, by-products. Several entries may form subfields, delimited by ";".

Examples: cultivated for consumption
single-cell protein from molasses
digestion of straw as cattle feed
protein production from paddy straw

PATENT Patent(s): I *

In STR. This kind of information is rarely made public, only upon instruction by the patent holder.

Patented application, country and number. Several entries concerning the same or different applications may form subfields, delimited by ";".

The word "patent" is to be entered together with the country.
Example: cellulase production, US patent 2,087,456; British patent...

CTR Control of the microorganism in question

In SP and STR. Chemical control (agent, type of control, technique) - physical control (agent, type of control, technique) - biological control (organism, type of control, technique). Application of the organism for controlling others is entered in ANT (possibly also APPL).

Several entries may form subfields, delimited by ";".
 Examples: Prochloraz, foliar application, at tillering
 water of 50°C, seed treatment, 30 min.

APLIT Literature on application, patents, etc.
 In STR.

The development of databases for culture collections in Europe with common formats would not have been possible without the financial support of national resources and of the European Commission's Biotechnology Action Programme (BAP contracts 0004-UK, 0005-NL, and 134-B). The Dutch MINE node is indebted to the Dutch Ministry of Education and Sciences, the Royal Dutch Academy of Sciences, and the Innovation-oriented Programme in Biotechnology (Ministry of Economic Affairs); the Belgian MINE node to the Belgian Science Policy Office (project 'Belgian Coordinated Collections of Microorganisms') and to the Bureau M. van Dijk, particularly Mr L. Caeymex; and the UK MINE node to CAB International, the CAB International Mycological Institute and the UK Department of Trade and Industry.

The format presented here was developed in collaboration between the fungal culture collections participating in MINE. The staff members of the Centraalbureau voor Schimmelcultures are indebted to Mr C. Brinkers (Control Data) for assisting in the first implementation of the format in BASIS, and to Mr G. J. Stegehuis for perfecting the CBS database. Mr J.-M. Falisse and Mr C. Meessen implemented a similar system at the MUCL collection on a Macintosh computer.

This is MINE Technical Communication no. 1.

REFERENCES

- ANDERSON, T.-H., BODENSTEIN, J. & DOMSCH, K. H. (1976). A partially computerized documentation system for microscopic soil fungi. *Canadian Journal of Botany* **54**, 1709-1713.
- ANONYMOUS (1983). *Microbiological Resource Database, MIRDAB*. Workshop 24 May 1983, Brussels. Report of the European Communities. Amsterdam: Elsevier, Excerpta Medica.
- BOIDIN, J. (1964). Valeur des caractères culturaux et cytologiques pour la taxonomie des Thelephoraceae résupinés et étalés-réfléchis (Basidiomycètes). *Bulletin de la Société botanique de France* **111**, 309-315.
- BRIDGE, P. D. & HAWKSWORTH, D. L. (1984). The API ZYM enzyme testing system as an aid to the rapid identification of *Penicillium* isolates. *Microbiological Sciences* **1**, 232-234.
- BRYANT, T. N. (1983). A microcomputer-based information storage and retrieval system for the maintenance of records for a culture collection. *Journal of Applied Bacteriology* **54**, 101-107.
- CENTRAALBUREAU VOOR SCHIMMELCULTURES (1987). *List of Cultures*. Baarn: Centraalbureau voor Schimmelcultures.
- DERMOUMI, H. (1979). Differentiation of yeast-like fungi isolated from clinical specimens with the API 20 C auxanogram. *Ärztliches Laboratorium* **25**, 289-291.
- DOMSCH, K. H., GAMS, W. & ANDERSON, T.-H. (1980). *Compendium of Soil Fungi*. London: Academic Press.
- ENVIRONMENTAL RESOURCES LTD (1984). *EEC Information System for Microorganisms*. Report prepared for the Commission of the European Communities.
- ENZYME NOMENCLATURE (1984). Recommendations (1984) of the Nomenclature Committee of the International Union of Biochemistry. Orlando: Academic Press.
- HAWKSWORTH, D. L. (1980). Recommended abbreviations for the names of some commonly cited authors of fungi. *Review of Plant Pathology* **59**, 473-480.
- HAWKSWORTH, D. L. & KIRSOP, B. E. (editors) (1988). *Living Resources for Biotechnology: Filamentous Fungi*. Cambridge: Cambridge University Press.
- HAWKSWORTH, D. L., SUTTON, B. C. & AINSWORTH, G. C. (1983). *Ainsworth & Bisby's Dictionary of the Fungi*, 7th edn. Kew: Commonwealth Mycological Institute.
- MCGOWAN, V. F. & SKERMAN, V. B. D. (editors) (1982). *World Directory of Collections of Cultures of Microorganisms*, 2nd edn. Brisbane: University of Queensland. Issued by World Data Center of Microorganisms.
- MERCK (1987/88). *Reagenzien, Diagnostica, Chemikalien*. Darmstadt: E. Merck AG.
- PRIDHAM, T. G. (1981). Microorganisms culture collections: acronyms and abbreviations. *Agricultural Reviews and Manuals ARM-NC-14*. Peoria: US Department of Agriculture.
- NORTHOLT, M. D., VERHULSDONK, C. A. H., SOENTORO, P. S. S. & PAULSCH, W. E. (1976). Effect of water activity and temperature on aflatoxin production by *Aspergillus parasiticus*. *Journal of Milk and Food Technology* **39**, 170-174.
- ROGOSA, M., KRICHEVSKY, M. I. & COLWELL, R. R. (1986). *Coding Microbiological Data for Computers*. New York: Springer.
- WASFY, E. H., BRIDGE, P. D. & BRAYFORD, D. (1987). Preliminary studies on the use of biochemical and physiological tests for the characterization of *Fusarium* isolates. *Mycopathologia* **99**, 9-13.