Challenges to Taxonomic Information Management: How to Deal with Changes in Scientific Names*

R. Froese, **E. Capuli and M.C. Rañola**
International Center for Living Aquatic Resources Management (ICLARM),
MCPO Box 2631, 0718 Makati City, Philippines

Abstract
Scientific names in taxonomic works become invalid at a rate of about 5-10% per decade during the first fifty years after their publication. Because of these changes, and because of current indexing practices, searches in bibliographic databases find only a subset of relevant publications. Several solutions to these problems are presented, notably the Species 2000 initiative and a new indexing method for bibliographic services.

Introduction

The recent global focus on biodiversity issues has put the past and present work of taxonomists in the spotlight. While the importance of taxonomic work is recognized and calls are made to reverse the current decline in positions for taxonomists, there are also claims to make the results of taxonomic work more usable to non-taxonomists. For most people, scientific names are non-telling and difficult to pronounce and remember. So, why not use common names? Computer programmers tell us that integer numbers are processed much faster than combinations of words. So, why not use codes? And if we have to use Latin names, why do they keep changing all the time? Why is there no system that tells us what scientific names are currently valid and how they are correctly spelled? How are the many people dealing with organisms as part of their regular work supposed to keep track of name changes published in specialist bulletins? In this contribution we explore some of these issues and try to point out possible solutions.

Names keep changing

The 'two-Latin-words' scientific naming system for organisms introduced by Linnaeus in 1758 is still in use today. In this system, the scientific name of a species serves a dual purpose: 1) provide organisms with a unique and permanent label, and 2) indicate the nearest relatives of a species. The first objective is compromised by the second, because progress in understanding the relationship of species changes the scientific name. Such name changes may affect many users, especially when they involve widely known species such as the rainbow trout (renamed from *Salmo gairdneri* to *Oncorhynchus mykiss*), or the tilapias, which were split into three genera, *Tilapia*, *Sarotherodon*, and

* Rainer Froese, Emily Capuli and Catalina Ranola. Trieste5.DOC, ICLARM Contribution 1392.

Oreochromis (see discussion in Pullin 1982). Such changes, if not quickly adopted globally and linked to previous publications, severely hamper our ability to communicate and to access existing knowledge.

The example of fishes

Since Linnaeus (1758), about 53,000 species of fish have been described (Eschmeyer 1998) and about half of these are currently considered to be valid species (Nelson 1994). Figure 1 shows a chronology of original species descriptions, currently valid species, and valid species that still bear the name used in the original description, as contained in FishBase, a large biological database on fish covering about 4/5th of the known species (Froese and Pauly 1998; http://www.fishbase.org). The outstanding contributions of Linnaeus (1758), Bloch (1785 ff.), Lacepède (1798 ff.), Cuvier and Valenciennes (1828 ff.), Günther (1859 ff.) and Boulenger (1909 ff.) can easily be seen. Also, the increase in taxonomic work during the age of European expansion as well as its decrease during World Wars I and II are clearly visible. There is an interesting gap from 1880 to 1890, possibly caused by the fact that Cuvier, Valenciennes and Giinther had described most specimens available in the collections of the time. The graph also shows the high rate of duplicate descriptions from the early 19^{th} to the mid-20^{th} century, probably caused by a widespread rush to describe new species, coupled with inadequate access to published literature.

In Zoology, about 15,000 new species are described every year, about 4,500 known species are placed in a different genus (new combinations), and about 4,000 species are found to be identical with an already described species, their names thus becoming new synonyms (Judy Howcroft, Zoological Record, pers. comm.). For fish, the respective numbers are shown in Fig. 2. Note that these figures also include fossil fishes and therefore exceed the annual 200-300 new descriptions of fish estimated by Eschmeyer (1998). This continuous trend leads to a rapid 'aging' of taxonomic works, which obviously do not include species discovered after their date of publication, i.e., they become incomplete, and which continue to use scientific names that are no longer considered valid. For example, of the 416 fish species in the 10^{th} edition of Linnaeus' (1758) *Systema Naturae*, currently 89% are considered to be valid species, with only 18% still bearing their original name. Froese et al. (1996) made an effort to update all names in a widely used work on the marine fishes of Indonesia (Gloerfelt-Tarp and Kailola 1984). They found that 12 years after its publication, 105 (about 10 %) of the names were invalid (19 junior synonyms, 21 new combinations, 32 misspellings or emended specific epithets, 10 misidentifications, and 23 unnamed species which since have been identified).

The extent of the problem is shown in Fig. 3, which displays, for some taxonomic works on fishes, the percentages of names that have changed as a function of the year of publication. Note that the slope of the hand-fitted line between 1950 and 2000 suggests that about 5-10% of scientific names became outdated within a decade after publication. This number can be expected to be significantly higher in non-taxonomic works that tend to place less emphasis on the validity of scientific names.

Impact of name changes on bibliographic databases

One of the tasks of taxonomy is to provide a functional system of unique names (labels) for living organisms. This functionality can be tested by using bibliographic indexing services such as the Zoological Record (ZR) for publications in Zoology, or Aquatic Sciences and Fisheries Abstracts (ASFA) for aquaculture and fisheries. The most time-consuming task of these services is to assign appropriate keywords to publications. Scientific names are regularly used as key words, in the spelling that was used in the publication, and sometimes introducing new misspellings. In addition, ASFA and ZR use a taxonomic thesaurus to assign families and higher taxa, based on the provided generic name (TRITON, the thesaurus used by ZR is available on the Internet at http://www.york.biosis.org/triton/). Names that are not mentioned in the title, abstract, or provided key words are either ignored (ASFA) or are used only when they constitute new names, synonyms or combinations (ZR). The indexing services normally do not correct typographic errors in the scientific names that were brought to their attention, neither those contained in the original papers nor those committed by themselves.

In 1989, the scientific name of the well-known rainbow trout was changed (Smith and Steady 1989). Figure 4 shows the results of four searches in ASFA for *Salmo gairdneri*, *Oncorhynchus mykiss*, rainbow trout, and a combination of these three terms, respectively. As can be seen, some publications are found under the new name before 1989, indicating that the new name had been added to the keywords of some references by ASFA staff, probably for pre-1989 publications that were processed after 1989. After that date, the use of the new name increased steadily. It took about three years for most researchers to accept the new name. The common name is used steadily throughout this period; however, it usually finds fewer publications than the scientific names. A combination of the three names finds 10-30% more publications than the respective scientific names. The same searches were also conducted in the Zoological Record and produced very similar results. However, several important publications are not found at all, because neither of the three names was used in the title, key words, or abstract. This is typically the case if a publication treats more than 10-20 species.

Coding systems have failed

Coding systems have been suggested to deal effectively with nomenclatural systems. Codes are especially popular with system analysts, probably because they fit well with programming languages such as Assembler, FORTRAN or C, and operating systems such as Unix. The suggested advantages of codes are:

- they are shorter than scientific names;
- less storage space is needed, resulting in faster retrieval and faster data entry;
- they can be easier sorted or grouped, e.g., at the family level; and
- they are more stable than scientific names.

However, none of these benefits stood the test of time. Coding systems that started with 3-5 digits have grown to 8-12 digits and more. A numbering system for all taxa is estimated to need codes of 40 or more digits (Pinborg and Paule 1990). The advent of fast computers, large storage capacity, and modern relational database software has made the listed advantages largely irrelevant. Also, working with codes is more prone to errors (Hureau, pers. comm.), and it is very difficult to detect typographical errors (Eschmeyer, pers. comm.).

Coding systems are created as a snapshot of taxonomy at a certain point in time. However, scientific names change. To accommodate such change, coding systems have to keep track of former and current codes (for examples, see Smith and Heemstra 1986 or Hardy 1993). Depending on the degree to which a coding system tries to reflect taxonomy, it might even need changes where scientific names remain unchanged, as is the case when a genus is transferred to another family. To avoid this problem, the recent Australian coding system for aquatic biota (Yearsley et al. 1997) decided to continue the family classification of Greenwood et al. (1966), thus ignoring 30 years of taxonomic research (Nelson 1984, 1994, Eschmeyer 1990, 1998). It is therefore incompatible, at the family level, for several groups.

The attempt to provide a stable coding system for a continuously changing taxonomy is bound to fail. It will either perpetuate outdated knowledge, including known mistakes such as misidentifications, or will require the creation and maintenance of extensive synonymies of code numbers, a pointless exercise. Therefore, we strongly support the view that scientific binomina with their established rules and synonymies are the `coding system' that should be used globally.

The FishBase approach

FishBase is a biological database on fish, available in part on the Internet (http://ww.fishbase.org/search.cfm) and in annual updates on CD-ROM (Froese and Pauly 1998). FishBase maintains a SYNONYMS table, which contains about 60,000 scientific names of fishes that have been used in publications, the current status of each name (original combination, new combination, junior synonym, misspelling, misidentification, other), a yes/no statement whether the name is valid, and the reference that was used to determine the status of the name. Attached to the SYNONYMS table is the BIBLIO table, which contains, for each name and publication, the page number where the species is dealt with, and the reference (code) number of the publication. The BIBLIO table is linked to the REFERENCES table, which contains, for each reference number, the full bibliographic details. The SYNONYMS table is also linked to the SPECIES table, which contains only valid species, and which is linked to a COMMON NAMES table, which currently holds more than 70,000 common names in more than 100 languages (see Fig. 5). This relatively complex structure is hidden from the data encoders. Whenever a new reference is added for a species, a dialog box opens which asks for the reference (code) number, the page number, and the scientific name that was used as valid. This name can be selected from a list of available synonyms. If it does not exist yet, it has to be entered in the SYNONYMS table.

This structure finds all references for a species for all recorded synonyms and common names. A change in the validity of a scientific name will not affect the attachment to the BIBLIO table, and thus all references continue to be attached to the current name as well as to the name they used as valid. This allows the creation of errata lists for nomenclatural changes for any reference.

The Species 2000 **approach**

Species 2000 is an initiative to provide a global standard of scientific names (including synonyms and common names) for all known organisms (Bisby and Smith 1996). The names will be made available in annual updates on CD-ROM and on the Internet (http://www.species2000.org/). The structure of the Species 2000 index is very similar to the one described above. Obviously, such an index would be of use to bibliographic indexing services such as ASFA and the ZR, to replace the taxonomic thesauruses they are currently using, and which, as in the case of ASFA (Fagetti et al. 1986), are incomplete and largely outdated.

Species 2000 has adopted a policy never to discard a name that was ever listed, even if it was misspelled or misapplied. This continuity will ensure that links that were once established with the Species 2000 index will continue to work with future upgrades, even if the scientific name has changed. Thus, for example, the bibliographic indexing services will not only improve the completeness and correctness of the hits found in a search for a scientific name, they will also be able to offer a new service to users; that is, to obtain recent lists of nomenclatural changes for important taxonomic works.

Ultimately, users of scientific names need a system where they can verify that the names that they are about to use in a publication are valid. The Species 2000 initiative will provide such a system.

Conclusion

From a user's point of view, the current naming system for organisms suffers from several shortcomings, mainly that there is no easy way to find out what name and spelling is currently valid and that it is currently impossible to find all relevant information on a species, either in bibliographic services or in the Internet. The Species 2000 Initiative seems to offer a solution to these problems, but it depends on the collaboration of the taxonomic community to create and continuously update an index of valid and synonymous scientific names that can be used as a standard. The museums of the world will have to accept that fulfilling their mandate in the 21st century will include the creation and maintenance of global species databases for at least some of the groups they are working with. For fishes, the *Catalog of Fishes* database (Eschmeyer 1998) and FishBase (Froese and Pauly 1998) have shown the feasibility of this task.

Acknowledgments

We thank Judith Howcroft of the Zoological Record for explaining their mode of operation and providing some of the data presented here. We thank Daniel Pauly, Roger Pullin and Meryl Williams for comments on the manuscript.

References

Bisby, F. and P. Smith. 1996. Species 2000: indexing the world's known species. Project Plan. University of Southampton, UK. 44 p.

Bloch, M.E. 1785-95. Naturgeschichte der ausländischen Fische. Berlin. Parts 1-9 [v. 1 (1785), v. 2 (1786), c. 3 (1787), v. 4 (1790), v. 5 (1791), v. 6 (1792), v. 7 (1793), v. 8 (1794), v. 9 (1795)].

Boulenger, G.A. 1909.Catalogue of the fresh-water fishes of Africa in the British Museum (Natural History). London. V. 1, p. i-ix + 1-373

Cuvier, G. and A. Valenciennes. 1828-49. Histoire naturelle des Poissons. Paris-Strasburg. Volumes 1-22 [vols. 1 & 2 (1828), vols. 3 & 4 (1829), vols. 5 & 6 (1830), v. 7 (1831), v. 8 (1832), v. 9 (1833), v. 10 (1835), v. 11 (1836), v. 12 (1837), v. 13 (1839), vols. 14 & 15 (1840), v. 16 (1842), v. 17 (1844), v. 18 (1846), vols. 19 & 20 (1847), v. 21 (1848), v. 22 (1849)].

Eschmeyer, W.N. 1990. Catalog of the genera of recent fishes. California Academy of Sciences, San Francisco. 697 p.

Eschmeyer, W.N. 1998. Editor. Catalog of fishes. Special Publication, California Academy of Sciences, San Francisco. 3 vols. 2905 p.

Fagetti, E., D.W. Privett, and J.R.L. Sears (Compilers). 1986. Aquatic sciences and fisheries thesaurus: descriptors used in the Aquatic Sciences and Fisheries Information System (ASFA).

Froese. R., S. Luna, and E. Capuli 1996. Checklist of marine fishes of Indonesia, compiled from published literature. pp. 217-275 *In* D. Pauly and P. Martosubroto (eds.) Baseline studies of biodiversity: the fish resources of Western Indonesia. ICLARM Stud. Rev. 23, 321 p.

Froese, R. and D. Pauly. 1998. Editors. FishBase 1998: concepts, design and data sources. ICLARM, Manila, Philippines. 293 p.

Gloerfelt-Tarp, T. and P.J. Kailola. 1984. Trawled fishes of southern Indonesia and northwestern Australia. Australian Development Assistance Bureau, Australia, Directorate General of Fishes, Indonesia, and German Agency for Technical Cooperation, Federal Republic of Germany. 407 p.

Greenwood, P.H., D.E. Rosen, S.H. Weitzman and G.S. Myers. 1966. Phyletic studies of teleostean fishes with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. 131(4):339-455.

Günther, A. 1859-70. Catalogue of the fishes in the British Museum. London. Volumes 1-8 [v. 1 (1859), v. 2 (1860), v. 3 (1861), v. 4 (1.862), v. 5 (1864), v. 6 (1866), v. 7 (1868), v. 8 (1870)].

Hardy, J.D. 1993. NODC taxonomic code links biology and computerized data processing. Earth System Monitor, 4(2): 1-2.

Lacepède, B. 1798-1803. Histoire naturelle des poissons. Paris. Volumes 1-5 [v. 1 (1798), v. 2 (1800), v. 3 (1801), v. 4 (1802), v. 5 (1803)].

Linnaeus, C. 1758. Systema Naturae, Ed. X. (Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentis, synonymis, locis. Tomus I. Editio decima, reformata) Holmiae. v. 1:824. [Nantes and Pisces in Tom. 1, pp. 230-338. Date fixed by ICZN, Code Article 3].

Nelson, J.S. 1984. Fishes of the world. 2nd edition. John Wiley and Sons, New York. 523 p.

Nelson, J.S. 1994. Fishes of the World. 3rd edition. John Wiley and Sons, New York. 600 p.

Pinborg, U. and T. Paule. 1990. NCC Coding System: a presentation. The Nordic Centre, Stockholm, Sweden.

Pullin, R.S.V. 1982. *Tilapia, Sarotherodon* or *Oreochromis*? ICLARM Newsletter 5(1):19.

Smith, M.M. and P.C. Heemstra, Editors. 1986. Smith's sea fishes. Springer Verlag, Berlin. 1047 p.

Smith, G.R. and R.F. Stearly. 1989. The classification and scientific names of rainbow and cutthroat trouts. Fisheries 14(1):4-10

Yearsley, G.K., P.R. Last and G.B. Morris. 1997. Codes for Australian Aquatic Biota (CAAB): an upgraded and expanded species coding system for Australian fisheries databases. CSIRO Marine Laboratories, Report 224. CSIRO, Australia.

Figure Legends

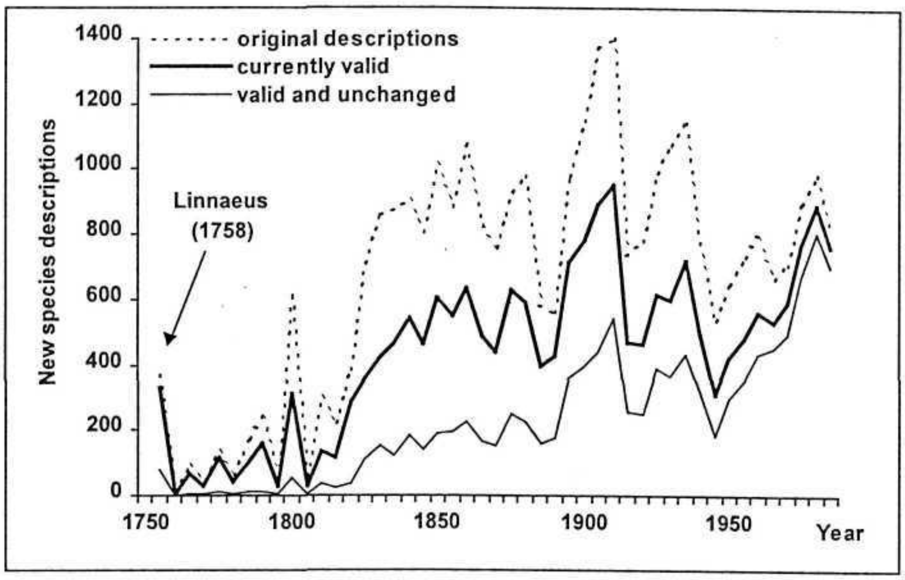
Fig. 1. Species descriptions per 5 year intervals over time, as contained in FishBase (Froese and Pauly 1998), covering about 4/5 of the known species.

Fig. 2. New fish names in the Zoological Record from 1978 to 1994 (fossil fishes included). New combination refers to species transferred into another genus, new synonyms refers to species that are found to have been described previously.

Fig. 3. Changed fish names (in percentage) in selected taxonomic works. Note that the slope of the hand-fitted line between 1950 and 2000 suggests that about 5 - 10% of the names will be outdated within ten years after publication.

Fig. 4. Results of four ASFA searches for rainbow trout, using different search terms.

Fig. 5. Proposed database structure that will assign a publication to the name used and to the currently valid name.



original descriptions
currently valid
valid and unchanged
New species descriptions
Linnaeus
(1758)
1400
1200
1000
800
600
400
200
0
1750
1800
1850
1900
1950
Year

Fig. 1

Fig. 2

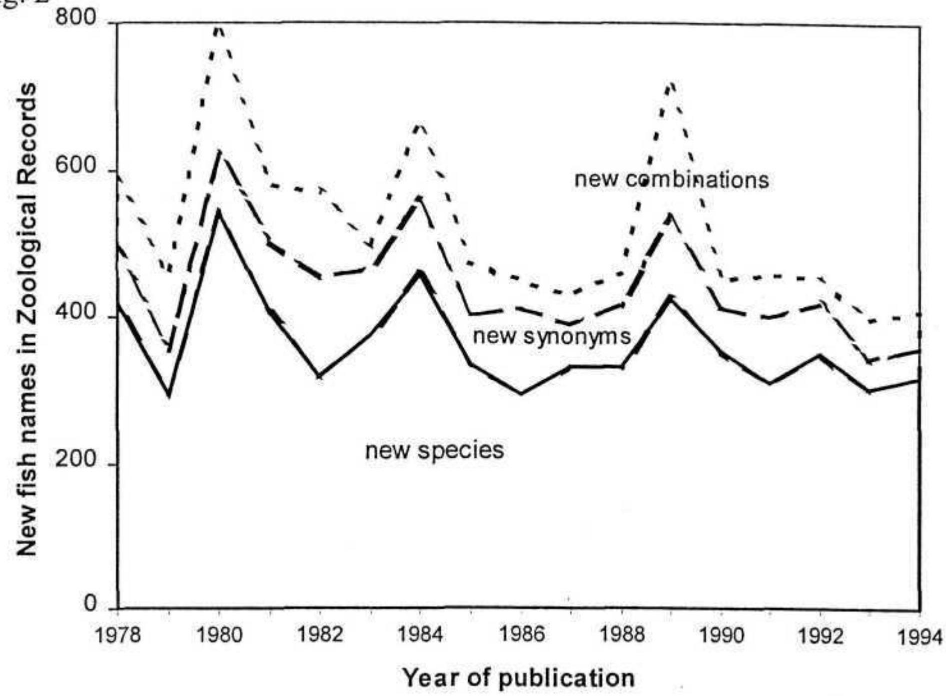
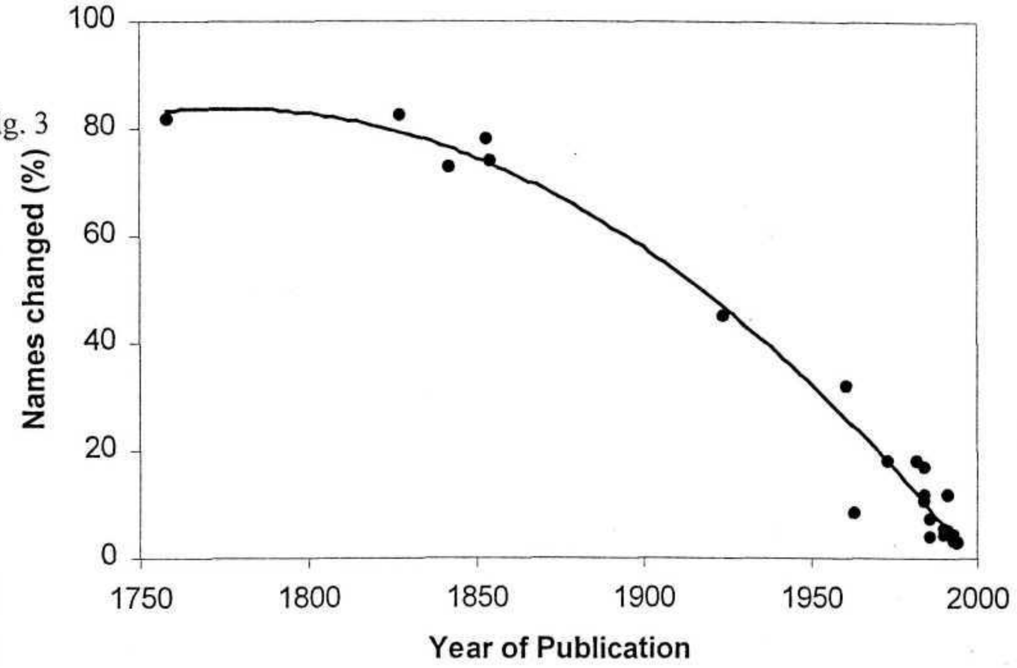

800
600
400
200
0
New fish names in Zoological Records
new combinations
new synonyms
new species
1978
1980
1982
1984
1986
1988
1990
1992
1994
Year of publication

Fig. 3



100
80
60
40
20
0
Names changed (%)
1750
1800
1850
1900
1950
2000
Year of Publication

Fig. 4

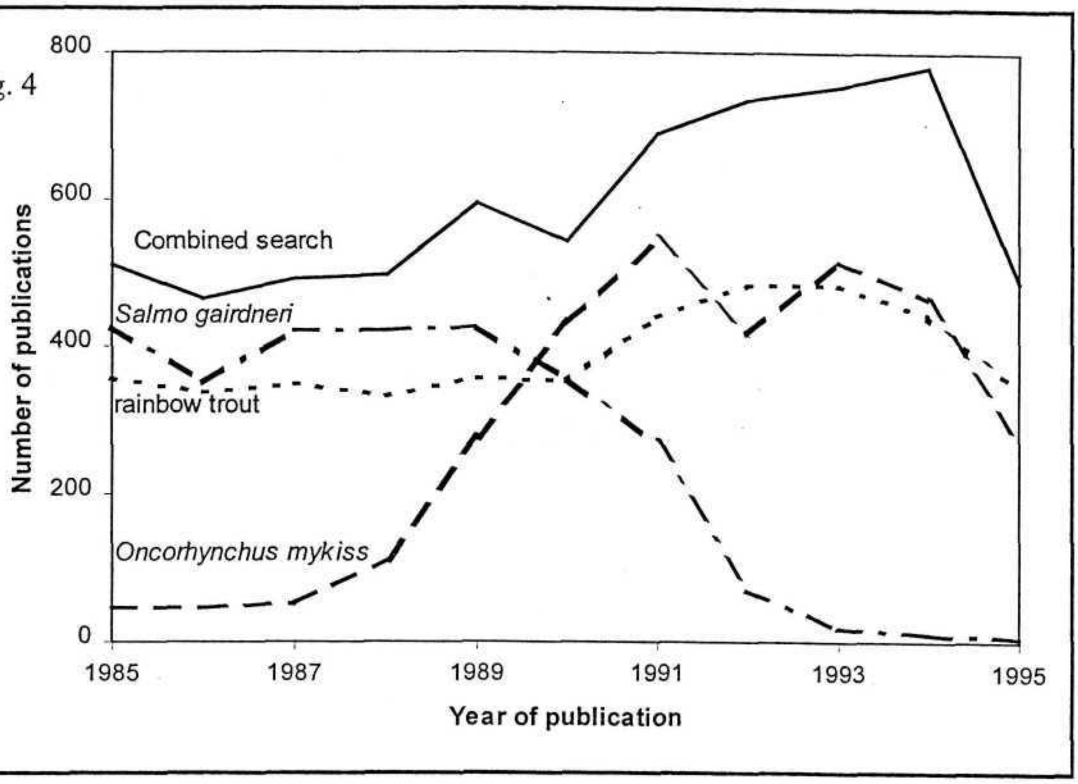
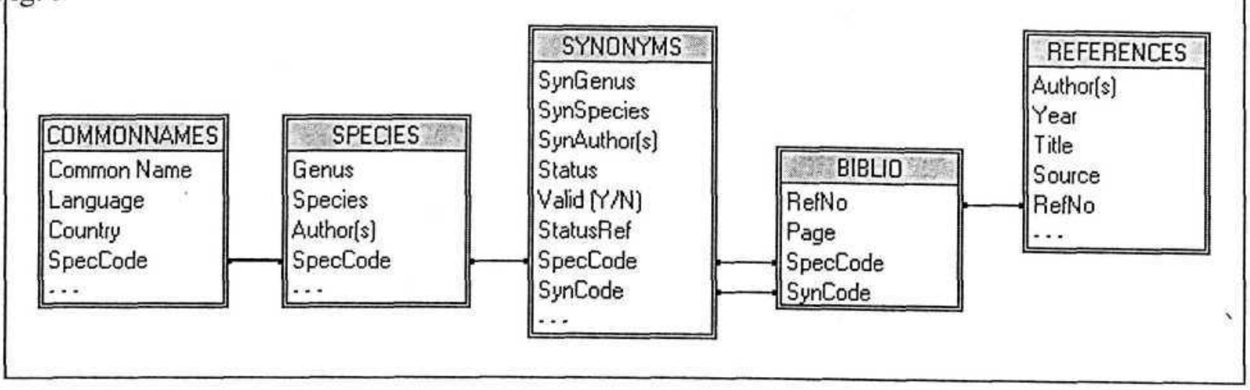

800
600
400
200
0
Number of publications
Combined search
Salmo gairdneri
rainbow trout
Oncorhynchus mykiss
1985
1987
1989
1991
1993
1995
Year of publication

Fig. 5


COMMONNAMES
Common Name
Language
Country
SpecCode
...
SPECIES
Genus
Species
Author(s)
SpecCode
...
SYNONYMS
SynGenus
SynSpecies
SynAuthor(s)
Status
Valid (Y/N)
StatusRef
SpecCode
SynCode
...
BIBLIO
RefNo
Page
SpecCode
SynCode
REFERENCES
Author(s)
Year
Title
Source
RefNo
...