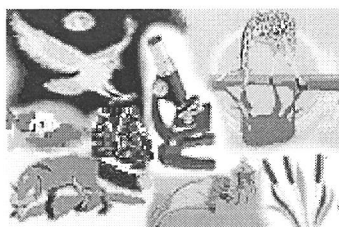3 0 JUIN 2000

P/494)277

**International Working Group
on Taxonomic Databases**

International Union of Biological Sciences
Taxonomic Database Working Group (TDWG)

Select a topic:

TDWG Newsletter

IUBS Taxonomic Databases Working Group	Number 10 / June 2000

TDWG Web page: http://www.tdwg.org

Communications

To keep up-to-date with what is happening within TDWG, please look at the web site http://www.tdwg.org for details.

There are also three email lists set up to help communications:

TDWG@USOBI.ORG

for matters of general interest to TDWG members.

See http://usobi.org/archives/tdwg.html on how to join the list.

TDWG-Proc@USOBI.ORG

to discuss the TDWG Standards Process.

See http://usobi.org/archives/tdwg-proc.html on how to join the list, and http://www.tdwg.org/process/tdwg99_blum.html on the reasons for the discussion.

TDWG-SDD@USOBI.ORG

to discuss the Structure of Descriptive Data, helping the TDWG SDD subgroup to analyse the requirements for a new standard for descriptive data based on XML.

See http://usobi.org/archives/tdwg-sdd.html on how to join the list, and read the archives.

Please use these resources to make your views known, and help make TDWG a more effective organisation.

TDWG Meetings

TDWG 2000

The annual meeting will be held at the Senckenberg Museum, Frankfurt, Germany from the 10th to the 12th November, 2000. It is being organised by Michael Türkay of the Senckenberg and Walter Berendsohn of the Botanic Garden and

Botanical Museum Berlin-Dahlem. The theme of this year's meeting is "Digitizing Biological Collections". For further infration, please see

http://www.bgbm.fu-berlin.de/tdwg/2000/ or the TDWG web site. Announcements will be also made through mailing lists such as the TDWG list and Taxacom.

TDWG 1999 - The Highlights

The meeting was held on 29th - 31st October, 1999 at the

Harvard Herbarium, Cambridge, USA. The full report can be seen at http://www.tdwg.org/rep1999.html .

The main aim of the meeting was to clarify some general standards issues, such as the role of standards in interoperability, when recommendations might be more appropriate, and TDWG's general role in facilitating the development of standards and recommendations. To begin this, Stan Blum, from the California Academy of Sciences, described the different types of standards and suggested ways in which TDWG could be useful in the development of more widely accepted standards, and how the organisation of TDWG could be improved.

John Rumble, from the National Institute of Standards and Technology, and currently President of CODATA, continued this theme presenting the need for standards, and the standards process. He described how CODATA fits into this process, illustrating how it might be useful for TDWG to be associated with CODATA, thus beginning the debate of where TDWG should "sit" institutionally, to be more visible and effective. Walter Berendsohn (Botanischer Garten und Botanisches Museum, Berlin) reported that the Global Plant Checklist project he was involved in was "sponsored" by CODATA and they had found it very useful in bringing people together. But he also wanted the meeting to consider possibilities with GBIF, the Global Biodiversity Information Facility. One of their prime areas of concern was the creation of checklists, and digitisation of databases, and Walter thought it should be possible for TDWG to get secretarial support, and financial help to attend standardisation meetings etc. He asked for permission to raise this with GBIF officials, and there was general agreement that he should go ahead with this.

Moving on to interoperability, Dave Vieglas from the Museum of Natural History, University of Kansas, illustrated the use of a general information retrieval standard to retrieve biological information from very different resources - the Z39.50 search and retrieval protocol, which now forms the basis of a number of projects focussed primarily on collections, and has led to the development of "The Species Analyst", see http://habanero.nhm.ukans.edu/SpeciesAnalyst. Using this tool, it is possible to search several databases at once, and merge the results into applications such as MS Excel.

There were some presentations representing the Taxon by

Character data "problem", experienced by users of packages such as DELTA, LucID, PAUP, etc. One aim of the meeting was that TDWG could facilitate a dialogue among the software developers. A subgroup was formed to discuss the "Structure of Descriptive Data" (SDD), and

there has been very active discussion on the list server set up (http://usobi.org/archives/tdwg-sdd.html).

Stuart Poss (Gulf Coast Research Laboratory, Ocean Springs, MS) reported on Database-related Activities Associated with ICZ XVIII [18th International Zoological Congress] - zoology was very fractionated and that there had not been an international conference for 40 years. Information on the conference can be

seen on the web site at
http://lionfish.ims.usm.edu/~musweb/icz_xviii/icz_home.html.

During the TDWG business session, there were reports from the economic botany
and geography subgroups, and on TL2. Francisco Pando (Paco) had sent in his
report as secretary as he was unable to attend the meeting. After being secretary
for 6 years, he was standing down. Everyone agreed that he should be thanked for
his efforts, without which TDWG might no longer exist. Peter Stevens reported
that he had approached Georgina MacKenzie to be the next secretary, and she had
agreed. There was general acceptance of this and she was voted in. Peter Stevens
presented the treasurer's report. Walter Berendsohn had agreed to organise the
next meeting, but suggested that it should be at Senckenberg Museum in
Frankfurt, which would accentuate the wider scope of TDWG.

Results of the BioCISE Project

This is an abbreivated version of the article by Walter Berendsohn, which can be
seen in full at http://www.tdwg.org/news10/biocise.html.

BioCISE (Resource Identification for a Biological Collection Information Service
in Europe) was a concerted action project financed under the 4th framework
programme of the European Commission in 1998 and 1999, with 17 participants
from 11 countries. The Concerted Action identified, analysed, and catalogued
biological collection information resources, recorded interdisciplinary
biodiversity database expertise, and analysed user needs for a European
Biological Collection Information Service. A concept for the implementation of
such a service was developed; and partnerships and proposals were initiated
based on the resources identified and in liaison with other organisations,
initiatives, and projects.

A large-scale survey of European collections was conducted throughout the
project period aimed at the identification of collections, collection databases and
biodiversity informatics expertise within the (initially) EU member states and
(lately) all potential participants in the 5th Framework Programme, i.e. including
the candidate countries for EU membership and the non EU-members among the
Northern and Central European countries.

The BioCISE World Wide Web Collection Catalogue provides access to the
survey's results. It can presently be queried by country, city, and collection
category, as well as by means of a free text search. For 60 % of the laboratories
which responded to the survey, the BioCISE Collection Catalogue is the first
representation of their collections on the World Wide Web. In addition to the
detailed survey results of respondents, all corroborated institutional addresses are
accessible.

To support follow-up projects, the BGBM continues to host the WWW site, the
database, and the questionnaires (in English, French, and German) for at least one
year after the project's conclusion. (see http://www.bgbm.fu-berlin.de/biocise/).

A group of members with experience in information modelling has finalised the
information model which is to provide a reference model useful for all biological
collections. The result was published in a 50-page paper in a major scientific
journal (Taxon) in 1999 (see http://www.bgbm.fu-
berlin.de/biodivinf/docs/CollectionModel/). In co-operation with TDWG, a
bibliography of collection information models and data resources was compiled
and published on the WWW (see http://www.bgbm.fu-berlin.de/TDWG/acc/).

The principal objective of the project was to develop a strategy for the
implementation of the Biological Collection Information Service and to instigate
project proposals within the 5th Framework Programme and national

programmes. It became clear that an all-out approach to enable access to collections in Europe on the level of the individual object (unit) is premature. A major challenge lies in the fact that long-term international funding for the service has to be kept low. The present thinking is therefore that three parallel approaches should be made: (1) foment collaboration of existing databases, e.g. by creating common interfaces to unit-level databases; (2) implementing an over all, flexible, metadata-driven access system to collections, which facilitates access to existing thematic, national, or regional networks and fills the gaps between them; and (3) to foment and instigate the creation and extension of such networks especially on the national level. The BioCISE project instigated and/or contributed to project proposals in all three areas.

Automatic compilation of accurate taxonomic databases from multiple non-computerised sources

Botany, like other mature sciences with a descriptive element, has extensive data locked in multiple overlapping natural language texts e.g. regional floras. The need to make these resources available in structured electronic form is urgent. However the problem of information extraction from text is amplified by the multiple sources and compounded by inconsistencies between them. For instance descriptions of the same plant species may be printed in several floras but using differing terminologies. Several parsing methods exist for extracting information from text but these have a number of limitations. This project aims to overcome these limitations by developing a tool for extracting data using techniques of electronic linguistic analysis that will exploit this redundancy in multiple parallel sources i.e. the several descriptions of the same taxon. It is expected that the union of the information derived from multiple sources will be both larger and more reliable than information from a single source. Whilst this tool is primarily aimed at biodiversity research once established it will hopefully be applicable in other sciences. The project is based at the University of Manchester, Department of Computer Sciences in collaboration with the Natural History Museum, London and is funded by the BBSRC EPSRC Bioinformatics Initiative . If you would like to know more about the project look at the website http://www.cs.man.ac.uk/ai/MultiFlora/.

International Plant Names Index is launched

The full version of this article by Eimear Nic Lughadh can be seen in full at http://www.tdwg.org/news10/ipni.html.

The Plant Names Project consortium has announced the Internet release of the International Plant Names Index (IPNI). This comprehensive listing of over 1.3 million scientific names for seed plants is the product of a collaboration between scientists and programmers at the Royal Botanic Gardens, Kew, U.K, the Harvard University Herbaria, USA, and the Centre for Plant Biodiversity Research, Canberra, Australia.

The Plant Names Project has adopted a number of the data standards endorsed by TDWG and new data added to IPNI will conform to these norms. However, retroactive implementation of these standards throughout the existing data is a mammoth undertaking, which is unlikely to be completed for some years. In order to encourage botanists worldwide to assist in this standardization process the PNP team is currently developing contributions software designed to enable users to make corrections to the data as they browse the index. For more information and access to the International Plant Names Index see http://www.ipni.org.

News on Standards

TL2

The following is a shortened version of an article written by Richard Pankhurst in March. Please refer to http://www.tdwg.org/news10/tl2.html for the full article.

A database of TL2 has been prepared, initially at the Missouri Botanical Garden and then at the Royal Botanic Garden Edinburgh, with financial support from the US Department of Agriculture (USDA). The accuracy of the data has been carefully checked. The data is being made freely available. The data fields that have been covered represent only a very small part of the information actually given in TL2. The database omits data on herbaria and types, the bibliography and biographical detail, and data on publishers and pagination. Data on various editions is not included unless TL2 gives them separate abbreviations.

In spite of its many virtues, TL2 is incomplete in various ways. The cutoff at 1940 means that many important publications are omitted. Even before that date, TL2 is does not include all references to publications containing new taxa by the authors listed, and there are certainly more minor authors who are omitted altogether. The supplements tend to contain rather a higher proportion of reprints, but they cover only up to letter G. The abbreviations for titles do have some consistency but they are not completely consistent, and need to be reworked to make them satisfactory.

In view of what has been said above, a number of improvements can be suggested

- Add the data from the last supplement

- Rework and reissue the abbreviations for titles

- For rare books, show which libraries have copies

- Add all the authors for multiple author works

- Add details of pagination with publication dates, for when a work did not appear all at one time

- Add details for all multiple editions

- Regularise the citation of reprints, with cross reference to the journal standard (BPH/S)

- Provide all references for existing authors, and add minor authors. This might amount to the same thing as completing the series of supplements.

- Add data for modern publications (and keep the catalogue continuously up to date). This might in practice amount to the same thing as cross-referencing TL2 to Index Kewensis (or its modern version, the International Plant Name Index).

- Since so many of the original publications are rare or obscure works, make photo images of the text and distribute in electronic form. This might be seen as a modern way of implementing the microfilm libraries which already exist

- Make TL2 available on CDROM (or DVD-ROM) and/or on the Internet

HISCOM/HISPID

Report on HISCOM2000:

The Herbarium Information Systems Committee (HISCOM) recently met in Darwin, Australia (2-5 May 2000). With representatives from all major State and Territory Australian mainland herbaria and invited guests from the Australian Biological Resources Study, CSIRO Publishing, KE Software (all Australia), and LandCare Research (New Zealand).

The major issues discussed by the meeting included:

1. The revision of HISPID3 (as HISPID4). A draft version of HISPID4 is available on the Internet as a searchable database (http://plantnet.rbgsyd.gov.au/Hispid4) or visit the HISCOM website (http://www.rbgsyd.gov.au/HISCOM/). Note: the HISPID4 database does not give examples or more detailed discussion (as present in HISPID3) at this stage. Furthermore, several corrections are yet to be completed.

2. The revision of the prototype Virtual Australian Herbarium (VAH) website. The VAH is a searchable distributed database system, that currently searches the accession databases of the major Australian herbaria and presents the information as a species distribution map. Currently, there is only restricted access to this site. The future development of the VAH includes the development of an external view to the data and the inclusion of other flora information, such as descriptions and images.

Barry Conn (Past-Coordinator of HISCOM)

BPH

This is an abbreviated form of the report by Richard Pankhurst, which can be seen in full at http://www.tdwg.org/news10/bph.html.

The Botanico-Periodicum Huntianum (or BPH for short) is a TDWG standard. To quote from the introduction it 'is a compendium on all periodical publications that regularly contain articles dealing with botanical literature'. BPH was first published by the Hunt Botanical Library, Pittsburgh, in 1968, and the Supplementum (BPH/S) in 1991. These two books need to be used together, since the supplement does not repeat some of the data already contained in the first volume, while it also lists all the old titles and adds many new ones as well, and presents new spellings for transliterated titles for several countries. BPH is the obvious standard to use for the titles and abbreviations of journals when compiling plant taxonomic databases. Experience in using BPH with the PANDORA taxonomic database suggests that it is very comprehensive. It is very unusual not to be able to find there the title of a journal containing published plant scientific names. As in the case of TL2 (for botanical books) it is the publications that constitute the standard, and NOT the corresponding database.

The database:

The BPH/S publication was word-processed, and the text file was kindly made available by Bob Kiger at the Hunt Institute. Manual editing to delimit the text into database fields and to remove superfluous periods at the ends of sentences was carried out at the Berlin Botanical Garden (Walter Berendsohn), the Harvard University Herbaria (Jim Beach and Nora Murphy) and at the Royal Botanic Garden, Edinburgh. The database has a single table and each record corresponds to a journal title. It turns out that a few of the titles used for preceding or superceding journals lack their own entries in the table, but these are usually very obscure titles. The main deficiency of the current database is that much of the series data which is included in the 1968 volume has not been entered.

Membership

Use the form included, or contact John Wiersema, TDWG Treasurer, c/o USDA, ARS, SBML, Building 011A, Room 304, BARC-West, Beltsville, MD 20705-2350 USA, or jwiersema@ars-grin.gov for details.

Newsletter

Please send your contributions and ideas to the Newsletter Editor: Georgina MacKenzie,

54 Micklegate,YORK, YO1 6WF, UK

Fax: +44 1904 612793

Email: gmackenzie@york.biosis.org

1999 - 2000 Executive

Chairperson:
Peter Stevens, Missouri, USA
peter.stevens@mobot.org
Secretary & Newsletter Editor:
Georgina MacKenzie, York, UK
gmackenzie@york.biosis.org
Treasurer:
John Wiersema, Beltsville, MD, USA
jwiersema@ars-grin.gov

Members:
Stan Blum, California, USA
sblum@calacademy.org

Gail Kampmeier, Illinois, USA
gkamp@uiuc.edu

Regional Secretaries:
Asia:
Junko Shimura, Saitama, JAPAN
junko@ulmus.riken.go.jp
Europe:
Walter Berendsohn, Berlin, GERMANY
wgb@zedat.fu-berlin.de
Latin America:
[position vacant]
North America:
David Boufford, Cambridge, MA, USA
boufford@oeb.harvard.edu
Oceania:
Barry Conn, Sydney, AUSTRALIA
barry@rbgsyd.gov.au

Results of the BioCISE project

BioCISE (Resource Identification for a Biological Collection Information Service in Europe) was a concerted action project financed under the 4th framework programme of the European Commission in 1998 and 1999, with 17 participants from 11 countries. The Concerted Action identified, analysed, and catalogued biological collection information resources, recorded interdisciplinary biodiversity database expertise, and analysed user needs for a European Biological Collection Information Service. A concept for the implementation of such a service was developed; and partnerships and proposals were initiated based on the resources identified and in liaison with other organisations, initiatives, and projects.

A large-scale **survey of European collections** was conducted throughout the project period aimed at the identification of collections, collection databases and biodiversity informatics expertise within the (initially) EU member states and (lately) all potential participants in the 5th Framework Programme, i.e. including the candidate countries for EU membership and the non EU-members among the Northern and Central European countries. By the end of the project period, 2550 laboratories (institutes and other collection holders) had been contacted by mail. In addition, 371 professional societies and organisations were asked to support the survey, either by notifying their membership, by linking to the project's site through the WWW, or by publishing our call for co-operation in their newsletter or journal. 483 laboratories (or 19 % of those contacted) replied to the survey. Of the respondents, 292 (61%) do maintain one or more biological collection databases. The total number of collection units (including survey records) catalogued in these 448 databases exceeds 42 million.

The **BioCISE World Wide Web Collection Catalogue** provides access to the survey's results. It can presently be queried by country, city, and collection category, as well as by means of a free text search. For 60 % of the laboratories which responded to the survey, the BioCISE Collection Catalogue is the first representation of their collections on the World Wide Web. In addition to the detailed survey results of respondents, all corroborated institutional addresses are accessible.

The Collection Catalogue also provides access to a set of **linked networks**. In the survey, the project had tried to limit its efforts to collections which have not already been covered by a functioning network. Such networks may be restricted to a certain kind of collection resource ("thematic networks", e.g. genetic resources, microbial culture collections, zoological gardens) or to a certain geographic region or country, or both. The BioCISE does not aim at replacing these networks, which are often community-driven, provide high quality data and – in their area – are wider in scope as envisioned for the general collection information service. Rather, the general system should provide access to linked subsets of the information held by these networks, so that users can be passed on to them where required. The Collection Catalogue provides a rough

prototype to demonstrate this idea of **interoperability and access** on a broader scale. Collaboration was agreed with two national networks (BIODIV in Belgium and NatureWeb in Austria). Tools were implemented which map collection categories used within these systems to those used by the BioCISE collection catalogue. Choosing, e.g., a Belgian laboratory within the BioCISE collection catalogue will now dynamically create a link to the referring entry in BIODIV. Three thematic networks, Index Herbariorum (IH), Common Access to Biotechnological Resources (CABRI), and the Directory of Germplasm Collections (IPGRI) and a national-thematic network (Polish Herbaria) were similarly linked.

To support follow-up projects, the BGBM continues to host the **WWW site,** the database, and the questionnaires (in English, French, and German) for at least one year after the project's conclusion. (see http://www.bgbm.fu-berlin.de/biocise/).

Throughout the project, BioCISE collected information on **users and potential users of biological collections**. Because of the broad, interdisciplinary approach of the project, the identification of user groups and user interests had to follow an informal procedure – BioCISE used meetings, discussions and interviews rather than questionnaires to gather information. The approach was justified by the results: Users are only rarely interested in a single category of collections. To provide a framework for the results, the project focussed at the following questions: Can we categorize users into broad groups according to their primary interests in collection information? Can we categorize uses of collections? What are the common denominators among different user groups and different uses of collections, i.e. what are the most important tasks a collection information service should tackle? A project subgroup concerned with the **environmental sector** discussed user needs in the context of governmental agencies, local government, and private enterprises conducting surveys or environmental impact assessments in BioCISE workshops and during congresses and other community events. It became clear that at least in this field, access based on geographic data is the prevailing demand. The absence of high quality data from biological collections often greatly complicates decision-making in landscape planning, conservation and related activities. Other groups focussed on user interests in systematics and in biotechnology. A paper with the results is in preparation.

A group of members with experience in **information modelling** has finalised the information model which is to provide a reference model useful for all biological collections. The result was published in a 50-page paper in a major scientific journal (Taxon) in 1999 (see http://www.bgbm.fu-berlin.de/biodivinf/docs/CollectionModel/). In co-operation with TDWG, a **bibliography** of collection information models and data resources was compiled and published on the WWW (see http://www.bgbm.fu-berlin.de/TDWG/acc/).

The rather fractured structure of the biological collection community has been a major **obstacle** for the survey. Holders of biological object information resources strongly emphasise taxonomic boundaries, which must be traversed if a user seeks information from a geographical, ecological, or biotechnological point of view. Another strong distinction encountered is that between survey and "physical" collection data. BioCISE aims at both physical collections (museums, botanical & zoological gardens, culture collections etc.) and surveys ("virtual collections": sighting, mapping projects, etc.), the findings of the information model group clearly indicate that the differences between both data areas are not justifying a distinction from the informaticians' point of view. Again, a user seeking general information, e.g. about the occurrence of organisms in a certain area, will not care for a distinction between data generated through an observation or such data found connected to an object in a collection (although only information confirmed by a voucher in a collection can be verified!).

The principal objective of the project was to develop a **strategy for the implementation of the Biological Collection Information Service** and to instigate project proposals within the 5th

Framework Programme and national programmes. It became clear that an all-out approach to enable access to collections in Europe on the level of the individual object (unit) is premature. A major challenge lies in the fact that **long-term international funding for the service** has to be kept low. The present thinking is therefore that three parallel approaches should be made: (1) foment collaboration of existing databases, e.g. by creating common interfaces to unit-level databases; (2) implementing an over all, flexible, metadata-driven access system to collections, which facilitates access to existing thematic, national, or regional networks and fills the gaps between them; and (3) to foment and instigate the creation and extension of such networks especially on the national level. The BioCISE project instigated and/or contributed to project proposals in all three areas.

Automatic compilation of accurate taxonomic databases from multiple non-computerised sources

Botany, like other mature sciences with a descriptive element, has extensive data locked in multiple overlapping natural language texts e.g. regional floras. The need to make these resources available in structured electronic form is urgent. However the problem of information extraction from text is amplified by the multiple sources and compounded by inconsistencies between them. For instance descriptions of the same plant species may be printed in several floras but using differing terminologies. Several parsing methods exist for extracting information from text but these have a number of limitations. This project aims to overcome these limitations by developing a tool for extracting data using techniques of electronic linguistic analysis that will exploit this redundancy in multiple parallel sources i.e. the several descriptions of the same taxon. It is expected that the union of the information derived from multiple sources will be both larger and more reliable than information from a single source. Whilst this tool is primarily aimed at biodiversity research once established it will hopefully be applicable in other sciences. The project is based at the University of Manchester, Department of Computer Sciences in collaboration with the Natural History Museum, London and is funded by the BBSRC EPSRC Bioinformatics Initiative . IF you would like to know more about the project look at the website www.cs.man.ac.uk/ai/MultiFlora/

Dr Rob Huxley

Head of Collections Division

Department of Botany,

The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Phone (44) 020 7942 5118

Fax (44) 020 7942 5529

E-mail: r.huxley@nhm.ac.uk

Museum Home Page URL: http://www.nhm.ac.uk/

International Plant Names Index is launched

The Plant Names Project consortium has announced the Internet release of the International Plant Names Index (IPNI). This comprehensive listing of over 1.3 million scientific names for seed plants is the product of a collaboration between scientists and programmers at the Royal Botanic Gardens, Kew, U.K, the Harvard University Herbaria, USA, and the Centre for Plant Biodiversity Research, Canberra, Australia.

Although many scientific organisations have databases which hold records of plant names and bibliographic sources until now there has been no freely available, comprehensive, global system. This million dollar project, jointly funded by the US National Science Foundation, the US Geological Survey, the Headley Trust and Reuters Foundation, represents a major step towards making accurate information on plant resources available to scientists throughout the world. The IPNI will be used by biologists, conservationists, ecologists and ethnobotanists in order to check the names and bibliographic details of the plants with which they are working. It is anticipated that it will also have a broader appeal as a ' plant name spell checker' for anyone whose interests bring them into contact with scientific plant names e.g. herbalists, journalists and even keen gardeners. Eventually, links will lead the user from the name itself to other electronic sources of information on the tree, shrub or herb in question.

The need for a complete listing of all the world's plants by name has long been recognised. In 1881, just a few months before his death, Charles Darwin arranged a legacy of £250 a year for five years so that the Royal Botanic Gardens, Kew could employ someone to produce just such a list. The eventual product, entitled *Index Kewensis*, contained some 400,000 names. However, new names continue to be published at the rate of up to 6,000 a year as new species are described and scientific understanding of relationships between organisms improves. Additions to the list over the past century have resulted in a total database of over 1,000,000 names. When these are combined with data contributed by American and Australian botanists, from the Australian Plant Names Index and the Gray Herbarium Index respectively, the result is a listing which is unrivalled in size and coverage.

The Plant Names Project has adopted a number of the data standards endorsed by TDWG and new data added to IPNI will conform to these norms. However, retroactive implementation of these standards throughout the existing data is a mammoth undertaking, which is unlikely to be completed for some years. In order to encourage botanists worldwide to assist in this standardization process the PNP team is currently developing contributions software designed to enable users to make corrections to the data as they browse the index.

For more information and access to the International Plant Names Index see http://www.ipni.org

TL2

Introduction

Taxonomic Literature edition 2, edited by Frans Stafleu, consists of 7 volumes and 5 supplements (letters A to G), and is known to its friends as TL2. In the form of these published volumes (but not as a database), TL2 is a TDWG standard. At the current time, all the contributing authors are deceased and the publication of TL2 in book form would appear to be at an end. TL2 catalogues botanical taxonomic books and reprints in which plant taxa have been published. Strictly speaking, an author is included if he or she has published 4 or more genera. All plants are included, meaning cryptogams and fungi as well as angiosperms, and fossils. Publications are not included after 1940, so that many modern major references are omitted. TL2 is comprehensive as far as the classical authors are concerned (e.g.Linnaeus) but less so for more recent authors. If an author has written both books and articles in journals, then no claim is made that all references to journals have been covered.

The compilers of TL2 went to great lengths to provide accurate information about exact dates of publication. In the 19th century, there was evidently a great ferment of interest in scientific discovery which stimulated the activities of many local scientific societies. Articles on new plants were preprinted, published and reprinted, sometimes with corrections and alterations and in translation, often with different titles and pagination, reissued in collected editions, and in journals which are now extinct and only to be found in a few specialist libraries. The difficulties involved in compiling exact data about such complicated publisheing procedures are considerable. Further back in time, plagiarism was more frequent and TL2 gives some detail – but not in such fashion as to give any currency to the pirated publications.

Database

A database of TL2 has been prepared, initially at the Missouri Botanical Garden and then at the Royal Botanic Garden Edinburgh, with financial support from the US Department of Agriculture (USDA). The accuracy of the data has been carefully checked. The data is being made freely available. The data fields that have been covered are:

- TL2 number (not quite always unique, nor exactly sequential)
- Expanded title (including accents), usually corresponding to the italicised portion of the text in TL2 i.e. not necessarily the full title. Occasionally the italicised section is excessively brief, and has been expanded a little in order to make more sense.
- Abbreviated title. Reprints do not have abbreviations for titles.
- Author abbreviation. Spaces after initials are suppressed, as in Brummitt and Powell. As in TL2, for multiple author works, only the first author is given.
- Date or range of dates from earliest to latest year (actual dates are given, if different from the published dates)
- Key for the Brummitt & Powell author abbreviations (foreign key to separate database). Some authors in the supplements are not in B & P.
- Notes; for most reprints, the journal citation has been keyed. There is some data on editions.

These data fields represent only a very small part of the information actually given in TL2. The database omits data on herbaria and types, the bibliography and biographical detail, and data on publishers and pagination. Data on various editions is not included unless TL2 gives them separate abbreviations.

In spite of its many virtues, TL2 is incomplete in various ways. The cutoff at 1940 means that many important publications are omitted. Even before that date, TL2 is does not include all references to publications containing new taxa by the authors listed, and there are certainly more minor authors who are omitted altogether. The supplements tend to contain rather a higher proportion of reprints, but they cover only up to letter G. The abbreviations for titles do have some consistency but they are not completely consistent, and need to be reworked to make them satisfactory.

Database Improvements

In view of what has been said above, a number of improvements can be suggested

- Add the data from the last supplement

- Rework and reissue the abbreviations for titles

- For rare books, show which libraries have copies
- Add all the authors for multiple author works
- Add details of pagination with publication dates, for when a work did not appear all at one time
- Add details for all multiple editions
- Regularise the citation of reprints, with cross reference to the journal standard (BPH/S)
- Provide all references for existing authors, and add minor authors. This might amount to the same thing as completing the series of supplements.
- Add data for modern publications (and keep the catalogue continuously up to date). This might in practice amount to the same thing as cross-referencing TL2 to Index Kewensis (or its modern version, the International Plant Name Index).
- Since so many of the original publications are rare or obscure works, make photo images of the text and distribute in electronic form. This might be seen as a modern way of implementing the microfilm libraries which already exist
- Make TL2 available on CDROM (or DVD-ROM) and/or on the Internet

R.J.Pankhurst

9th **March 2000**

BPH

Introduction

The Botanico-Periodicum Huntianum (or BPH for short) is a TDWG standard. To quote from the introduction 'is a compendium on all periodical publications that regularly contain articles dealing with botanical literature'. BPH was first published by the Hunt Botanical Library, Pittsburgh, in 1968, and the Supplementum (BPH/S) in 1991. These two books need to be used together, since the supplement does not repeat some of the data already contained in the first volume, while it also lists all the old titles and adds many new ones as well, and presents new spellings for transliterated titles for several countries. BPH is the obvious standard to use for the titles and abbreviations of journals when compiling plant taxonomic databases. Experience in using BPH with the PANDORA taxonomic database suggests that it is very comprehensive. It is very unusual not to be able to find there the title of a journal containing published plant scientific names. As in the case of TL2 (for botanical books) it is the publications that constitute the standard, and NOT the corresponding database.

Database

The BPH/S publication was word-processed, and the text file was kindly made available by Bob Kiger at the Hunt Institute. Manual editing to delimit the text into database fields and to remove superfluous periods at the ends of sentences was carried out at the Berlin Botanical Garden (Walter Berendsohn), the Harvard University Herbaria (Jim Beach and Nora Murphy) and at the Royal Botanic Garden, Edinburgh. The database has a single table and each record corresponds to a journal title. The data fields are as follows:

- Full title
- Abbreviated title
- Publisher
- Town/city where published
- Series, with dates and volume numbers
- Pointer to name of preceding journal, if any
- Pointer to name of superceding journal, if any
- BPH code, the page and line number of the 1968 publication

- Comments field for preceding and superseding titles, publication information, etc.

It turns out that a few of the titles used for preceding or superceding journals lack their own entries in the table, but these are usually very obscure titles. The main deficiency of the current database is that much of the series data which is included in the 1968 volume has not been entered.

The database fields of the Harvard publications records currently differ slightly and include additionally

- ISSN number
- ISBN number
- Library of Congress Number

R.J.Pankhurst

10th March 2000