Global Butterfly Names

(A)

A proposal to the ECAT programme of GBIF

B. Principal investigator:
James Mallet
Professor of Biological Diversity, University College London
Galton Laboratory, 4 Stephenson Way, London NW1 2HE, UK
tel: (+44)-(0)20-7679-7412, fax: (+44)-(0)20-7679-5052
http://abacus.gene.ucl.ac.uk/jim/
and
Honorary Research Fellow, The Natural History Museum
London, UK

C. Managing institution:
University College London, Department of Biology, Darwin Building
Gower Street, LONDON WC1E 5BT
Institutional contact person: **Dr. Jacques Gianino**
tel: (+44)-(0)20-7679-7421, fax: (+44) (0)20 7679 7096

Co-investigators:
Gerardo Lamas
Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima
George Beccaloni, Brian Pitkin, Campbell Smith and **Dick Vane-Wright**
The Natural History Museum, London

D. Partner institutions and organizations:
The Natural History Museum, London (NHM)
Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima
University College London (UCL)

Academia Sinica & Council of Agriculture, Taiwan
University of Alberta, Canada
National Museum of Natural History, Leiden, The Netherlands
Oregon State University (Corvallis), USA
Stockholm University
US National Museum of Natural History (Smithsonian Institution), Washington, USA

GloBIS
LepIndex/VIADOCS Card Index Digitization Project
The Lepidoptera Taxome Project

May 2004 (updated September 2004)

E. *Project summary*

A fundamental need for biodiversity-related studies is a freely available and complete catalogue of scientific names, together with the most up-to-date consensual classification placing those names in approximate phylogenetic context. We will undertake an international collaboration to produce, within 18 months, a >99% complete, online, freely-available database of butterflies (Rhopalocera), consisting of genus- and species-group names (including valid subspecies as well as complete data on available junior synonyms, invalid, and unavailable names) together with their classification.

We will merge data from a number of separate projects: (1) *LepIndex*, an online Lepidoptera name database based on the NHM card index, (2) modern global checklists, (3) Hemming's *The Generic Names of Butterflies*, and (4) Bridges' *Catalog of the Names of the Butterflies*, all of which we hold in electronic form. We will use high-efficiency, error-free data parsing, data merging and data validation as part of a world-wide collaboration. Our database will act as a Data Node for remote nameservers such as those envisaged by GBIF, Species 2000 or other programmes. *Global Butterfly Names* should do for butterfly taxonomy what GENBANK did for genetics, and will allow world data repatriation based on collections and information at the NHM and elsewhere.

F. *Project description*

a) The need for centralized information and its resolution

Biodiversity studies, and indeed any biological studies, require accessible specimen information and nomenclature databases to enable rapid assignment of individual specimens to particular names and classifications. An emphasis on online specimen databases is represented by GBIF's DIGIT programme, while online nomenclatural databases are represented by the ECAT programme of GBIF. However, the two programmes are much closer in practice than might be inferred from this division, because an ECAT database of names is essentially a DIGIT database of type specimens. These type specimens represent a shotgun scattering across the sum total of all specimens that make up global biodiversity, and it is these name-bearing type specimens, the only specimens that have been "identified" with certainty, which allow identification of all other specimens via comparison, or when comparison fails, demonstrate the need to erect new taxa and names. Ultimately other biological data such as ecological interactions, life-history data or DNA sequences relating to individual specimens, and geographic distribution data, will be attached to such online combined ECAT/DIGIT-type databases.

Current confusion about the taxonomic ranks of taxa calls into question our ability even to measure whether we can achieve, by 2010, a significant reduction of species loss, as mandated by the Convention on Biological Diversity. Recently, species counts have increased rapidly. Primate species counts have doubled since 1975, whereas mammals as a whole have increased 25%. Almost all this "taxonomic inflation" is due to changing concepts of species and, in particular, the elevation of already named subspecies to species rank, rather than due to actual discovery of new taxa (Isaac et al. 2004). Eventually, this catastrophic state of affairs might be resolved by international agreement, but it is clear that measuring biodiversity will remain uncertain for some time (Vane-Wright et al. 1991). To alleviate such problems, taxonomists must first document all scientific names in the form of complete and up-to-date global synonymic checklists. As these checklists become increasingly international, web-based and collaborative, and reflect world consensus rather than single-author opinion, they will form the only possible basis for stabilization of nomenclature and rank, for instance via Lists of Available Names in Zoology (ICZN art. 79).

Here we propose to fund a collaboration between developed and developing countries, backed by resources and archives at The Natural History Museum, London (NHM), to provide an open, online, complete and up-to-date database of all ~80,000 names applied to ~17,500 butterfly species. Butterflies represent around 1% of described species, are important in conservation, and are well known to the public as well as to biologists. Their wide appeal has led to a vast and chaotic taxonomic literature (Vane-Wright 2003), but the world collections held by the NHM include well over 50% of all butterfly type specimens. Butterflies thus represent a particularly high-profile and difficult challenge to the GBIF enterprise, but which is of major importance for biodiversity (Boggs et al. 2003), and one with which the NHM is particularly suited to deal. Butterflies will in future act as a model for similar databases for larger and less well-known groups of insects, such as plans afoot in the NHM for the remaining Lepidoptera (moths) and Coleoptera (beetles) global databases – around 450,000 species in all, 25-30% of the world's named biological species.

b) The importance of The Natural History Museum, London in global butterfly taxonomy
The NHM contains many important butterfly collections, particularly those amassed by the earlier British Museum, combined with those of Charles Oberthür and Walter Rothschild. This butterfly collection of ca. 2,800,000 specimens is curated by a team of three staff, under the guidance of P.R. Ackery. Other NHM staff members, postdocs, and scientific associates with leading skills in butterfly taxonomy include H. Gaonkar, J.D. Holloway, T. Larsen, D.C. Lees, A. Neild, C.R. Smith, J. Tennent, R.I. Vane-Wright and K. Willmott, representing a critical mass of professional expertise unmatched elsewhere. We estimate that the NHM contains over 50% of all butterfly type specimens worldwide, and over 70% in some groups.

c) Project description
(i) The proposed product. We here offer GBIF a flexible online database, >99% complete for all butterfly scientific names (including synonyms, homonyms, infrasubspecific and other unavailable names), together with their currently accepted status and rank based on worldwide taxonomic expertise, as well as information on type specimens and other important data.

We will incorporate global butterfly checklist information into our existing Lepidoptera names catalogue, *LepIndex* (http://www.nhm.ac.uk/entomology/lepindex/), created at the NHM under the leadership of Malcolm Scoble. *LepIndex* is an electronic version of the NHM card index of all lepidopteran scientific names, including unavailable as well as available names. Already, every one of the 290,000 typewritten cards have been databased as digital images; of these, 77% (223,000) are online to date, including all 68,900 butterfly names (24% of all Lepidoptera names). Fully synonymic checklists at any validated taxonomic rank are also available. For example, a prototype synonymic checklist generation can be obtained using: http://avalanche.nhm.ac.uk/jdsml/perth/lepindex/checklist.dsml?Current_subfamily=Ithomiinae for the subfamily Ithomiinae (Nymphalidae). n.b.: Not all genus and family-group names have yet been validated. For these taxa, checklists cannot be produced. Also, many species group names still return "rank not recorded". For a fully validated list (Tineodidae), try: http://avalanche.nhm.ac.uk/jdsml/perth/lepindex/checklist.dsml?Current_family=Tineodidae. By August 2004, all 290,000 Lepidoptera names will be online for fields of scientific name, current valid genus name, author, year of publication, and higher classification. *LepIndex*, however, contains information mainly up to the early 1980s, when NHM staff shortages halted further upkeep of the original card index; many new names (~2-5%) have been coined since then. In addition, the current status of most names has yet to be entered; in most genera, only a single species-group name per genus has been validated.

Global Butterfly Names will add new data and upgrade existing information in *LepIndex* to provide the major butterfly name source for GBIF and its users. We will update *LepIndex* fields with all more recent butterfly names, as well as with currently accepted nomenclature and classification information not so far entered in *LepIndex*, such as type locality, original genus, current status, rank, higher classification and geographic distribution. Modern, synonymic checklists of butterflies exist for most regions. Two are particularly useful: the *Atlas of Neotropical Lepidoptera Checklist* (Lamas et al. 2004a; see Fig. 1 below) treats 7,784 species considered valid, 44% of world species, and ~30,000 names (including all known available and unavailable published scientific names), 38% of global names; *Carcasson's African Butterflies* (PR Ackery et al. 1995, updated by MC Williams 2003) contains a further

3,900 species, 22% of the global total, and >10,500 names, 13% of the global total. We hold electronic, word-processed copies of both lists. Similar checklists exist or are under construction (mostly word-processed, but in some cases fully databased) for the remaining 49% of world butterfly names (see many refs. below). Fully databased, updated versions of CA Bridges' (1988-1994) *Catalogue of the Names of* [world] *Butterflies*, and of *The Generic Names of Butterflies and Moths*, an online database for all 31,000 Lepidoptera genus-group names (http://www.nhm.ac.uk/entomology/butmoth/) based in part on Hemming's (1967) catalogue, are both also held by the NHM (Pitkin & Jenkins 2004): these will provide essential name validation, as well as adding data to many fields in *LepIndex*. Classification will follow GloBIS higher classification (Lamas et al. 2004b).

Fig. 1. Fragment from the Atlas of Neotropical Lepidoptera Checklist (from Lamas et al. 2004)

PODOTRICHA Michener, 1942
2367 **judith** (Guérin-Méneville, [1844]) (*Cethosia*)
a) judith (Guérin-Méneville, [1844]) (*Cethosia*) Colombia
euchroia (Doubleday, [1847]) (*Colaenis*) Colombia
b) caucana (Riley, 1926) (*Colaenis*) Colombia
c) mellosa (Stichel, 1906) (*Colaenis*) Ecuador
d) straminea (Riley, 1926) (*Colaenis*) Ecuador
2368 **telesiphe** (Hewitson, 1867) (*Colaenis*)
a) telesiphe (Hewitson, 1867) (*Colaenis*) Ecuador
tithraustes var. *diaphana* (Niepelt, 1915) (*Colaenis*) Peru
b) tithraustes (Salvin, 1871) (*Colaenis*)
Ecuador

DRYADULA Michener, 1942
2369 **phaetusa** (Linnaeus, 1758) (*Papilio*) "Indiis"
f. *stupenda* (Stichel, [1908]) (*Colaenis*) Panama
f. *lutulenta* (Stichel, [1908]) (*Colaenis*) Paraguay
f. *deleta* (Stichel, [1908]) (*Colaenis*) Paraguay

To import word-processed lists into a database like *LepIndex*, data must first be "parsed" to identify each field uniquely. This is a relatively trivial programming task, since well-structured checklists flag many pieces of information using unique text styles. For example, in Fig. 1 (Lamas et al. 2004a), genera are capitalized and in bold; species are always in bold lower-case text and flagged with 4-digit numbers; subspecies considered valid are preceded by a letter; while species-group or infrasubspecific names considered not valid are italicised. Original generic combination, type locality information, and many other conventions are also shown in Fig. 1. Programs will be written to parse the majority of data from this and similar checklists into appropriate *LepIndex* fields. Other information will also be encountered, such as notes indicating the reason for use of the name [e.g. in the genus *Actinote*, Lamas et al. (2004a) write "2336 discrepans d'Almeida, 1958 (ICZN, Art. 45.6.2) Brazil (SC)"]. Here, the "ICZN, Art. 45.6.2" information would be placed in *LepIndex*'s comment field for later individual attention.

Merging pre-existing databases presents few problems. For instance, importing *Bridges' Catalog* or *Generic Names* databases into *LepIndex* requires matching unique combinations of fields in both databases (we will use scientific name, original generic name, author, and year of publication) to import information from other fields into each name record. The labour-requiring step, for which we need the taxonomic expertise of Dr. Lamas, is to assess taxa not currently in *LepIndex*, and mismatches of name, author or date of publication.

Finally, some manually typed data entry (we estimate <10%) will be necessary for regions of the world where electronic checklists are lacking, and to tidy up existing data where mechanized data entry fails (e.g. for homonyms published in the same year).

ii) Project milestones and timeline. Once suitable programmes are written, parsing word-processed lists into database format and merging databases is virtually instantaneous. We therefore propose to complete the majority of the project within 1 year, with the final, taxonomically validated version available after 18 months. Even if all names were processed

manually, there should be enough time to hand-type data fields for all ~80,000 butterfly names into the database. However, manual typing is error-prone and to be avoided. Rechecking and retyping data already correct in checklists is wasteful of taxonomists' time. We will therefore use automated data entry wherever possible to free scientists' time (e.g. for making decisions on taxa new to the database, taxon circumscription and mismatches or researching new data in the form of previously undiscovered literature or type specimens). As well as complete data entry via existing checklists, it should be possible to revalidate >90% of the names at the end of the 18 months, with the unparalleled library resources available at the NHM. Dr. Lamas will coordinate contacts with our collaborators around the world for lists in their regions and expertise in particular groups.

Milestone date	**Milestone**	**Cumulative validated names**
before 1 Oct 04	• NHM card index databasing complete (1758-1981) • Bridges (1979-94) databased • Generic Names online (Pitkin & Jenkins 2004) complete • Neotropical checklist complete (Lamas et al. 2004a) • GloBIS butterfly higher classification complete (Lamas et al. 2004b) • Afrotropical list completed (Ackery et al. 1995, Williams 2003) • Other electronic checklists completed (see refs. below) • *LepIndex* online, most Lepidoptera, all butterfly names entered (Beccaloni et al. 2004) • Checks and additions using Generic Names database (3,988 butterfly genus-group names)(Beccaloni et al. 2004) • Facility for online checklist generation added to *LepIndex* web portal	(69,000 card records basic data only, online) 4000 genera checked 0 name records complete
1 Jan 05	• Neotropical checklist parsed and online (30,000 names)	30,000
1 Apr 05	• Afrotropical butterfly names parsed and online (11,000 names)	41,000
1 July 05	• *LepIndex* – Species 2000/GBIF link goes live	41,000
1 Dec 05	• Checks and additions using Bridges (1979-1994) database (excludes Nymphalidae) (~ 15,000 new names)	56,000
1 Apr 06	• Checks and additions using all other checklist sources (>24,000 new names)	~80,000

iii) Human and institutional infrastructure. The project will involve global taxonomic collaboration between personnel in London, Lima, Alberta, Leiden, Oregon State, Stockholm, Washington DC, Taipei, etc. GBIF funds are chiefly required to bring Gerardo Lamas, lead author of the Neotropical checklist, to London for 5 months. A 5-month salary is necessary to support his stay in London because of the low salary he currently receives in Peru and the very high cost of living in London. A GloBIS member and ICZN commissioner, Dr. Lamas brings the needed world expertise to the project. His taxonomic and nomenclatural knowledge is essential both in resolving mismatch conflicts, and also in locating suitable checklists and expert help for further additions to *Global Butterfly Names*.

Additional taxonomic expertise is provided by GloBIS, as well as taxonomists at Alberta, Oregon State, Stockholm, Smithsonian Institution, Leiden and Taiwan. James Mallet, working closely with checklist authors Gerardo Lamas and Dick Vane-Wright, and with database author George Beccaloni, undertakes to write the parsing programs. Responsibility for database integration and additional manual data entry rests with James Mallet and Fraser Simpson, at UCL, and Gerardo Lamas, assisted by George Beccaloni, designer of *LepIndex*, Brian Pitkin, designer of *Generic Names* and Bridges' databases, as well as Dick Vane-Wright and Campbell Smith, co-authors of the Afrotropical checklist.

iv) Global Butterfly Names and GBIF philosophy of free and open access to data. i) International initiatives. The organization behind "*Global Butterfly Names*", GloBIS, is an international federation of butterfly taxonomists. We seek to provide global standards for butterfly classification (e.g. Lamas et al. 2004a,b) to the world community. Our work will help to stabilize names, and allow measurement of diversity changes by 2010, as mandated by the Convention on Biological Diversity (see above). *ii) Collaboration*. Our partnerships involve world collaboration. For example, the Neotropical checklist (Lamas et al. 2004a) included collaborators from across the Americas. *iii) Targeting.* Our work fills an important gap because no up-to-date global butterfly checklist exists.

v) How Global Butterfly Names will be made publicly available. LepIndex is currently available via its own website. Data will always be available in the short term, immediately they are entered (indeed complete butterfly name information, as yet unvalidated, is already available). Work is already funded to link *LepIndex* to the Species 2000 database, itself a Data Node for GBIF. Alternatively, assuming GBIF finance is received, direct access to *LepIndex* as a "GBIF data node" can be provided via GBIF-supported XML tools.

These efforts will allow access to all 290,400 Lepidoptera names, as well as the ~80,000 complete, checked and validated butterfly names from this project. Plans are also underway at the NHM to use the *LepIndex* database to index all 500,000 cards for Coleoptera in the near future; thus any programming and GBIF data delivery operations funded by *Global Butterfly Names* will be available seamlessly for nearly 800,000 names. Funding this project would make *LepIndex* and its successors the world's largest, and perhaps the largest ever ECAT provider.

A major long-term aim for the completed butterfly names project, after appropriate worldwide consultation, is to serve as an adopted List of Available Names in Zoology (ICZN art. 79) for the world's butterflies. This should help nomenclature stabilization for the long-term future.

Meanwhile, a number of initiatives are underway to provide better access to NHM data overall. The most far-reaching project for integrated web delivery of NHM data is the Museum Information Locator System (MILS), which will be implemented in the next few weeks. MILS allows a fully distributed search on any set of web-enabled NHM databases and will provide standard Dublin Core metadata through an OAI (Open Archives Initiative) repository, created through a harvested search. MILS will include science databases, the NHM Picture library and main NHM Library catalogue. It is also being used to provide NHM specimen data to BioCASE via Berlin-written wrapper software to access data using the MILS harvested search techniques. All data provided by the 31 countries active in BioCASE is then being made available to GBIF through a separate mechanism. Therefore, NHM data, including *LepIndex*, will be provided to GBIF anyway, but via BioCASE.

A second likely initiative is to convert all NHM collections, archive and library data to the Ke-EMU museum-oriented database system, giving integrated access to NHM data under a single database management system. In doing so, NHM will adopt standards used by other major world museums (such as the Smithsonian Institution, Washington DC), which should greatly simplify data exchange between institutions, including with GBIF.

It is still unclear precisely how and which of these Museum-wide initiatives will impact specifically on *LepIndex*, but they demonstrate continuing NHM commitment to database integration and open access to the Museum's data generally. In particular, NHM's Interactive Media Services department are committed to providing data to BioCASE, Species 2000, and GBIF for all names and specimen databases, including *LepIndex*, at the earliest opportunity.

vi) Global Butterfly Names and data compatibility, interoperability, and community standards. *LepIndex* already uses most GBIF-recommended data fields: accepted scientific names and related fields, synonyms, status, source of data, comment field, and family name. It currently lacks provision for two recommended groups of GBIF fields: "common names (including references)" and "latest taxonomic scrutiny". To rectify the first problem, common name fields are already being added to enable integration of *LepIndex* with Species 2000 (and GBIF) as a separate table with a field for language, linked to the main scientific name table using a "many-to-many" relationship. However, it will often be unclear to which scientific name a particular common name refers, since common names are not linked to types. Common name provision is not intended as a primary output of a global synonymic catalogue such as *LepIndex*; rather, *LepIndex* will supply the key reference for finding the currently valid taxonomic name of any scientific name linked to a common name in any other publication.

Fields for latest taxonomic scrutiny are already to some extent covered in the following way: firstly, a complete updating history is accessible in the underlying database, and the most recent updater for each name is already displayed for each *LepIndex* name; secondly, fields for new and old combinations of names store current and previous classification information; finally, the introductory material in *LepIndex* is already explicit about the suprageneric classification it uses – for the butterflies, this will be based on global opinion of butterfly taxonomists (GloBIS) (Lamas et al. 2004b). Nonetheless, it is clearly advisable, in addition, to add fields for checklists or other secondary taxonomic literature (as opposed to references for the primary combination) that are the immediate sources of classification adopted by GloBIS; these fields will therefore be added as part of a GBIF-funded *Global Butterfly Names* project.

Global Butterfly Names conforms as far as possible to community standards imposed by the current ICZN code, particularly since Dr. Lamas is a commissioner of the ICZN. Some reviewers have expressed concerns about the controversy over gender agreement between species group and genus group names. This problem is made more complex by databasing exigencies – to incorporate the rules of Latin grammar into databases so that gender agreements are preserved through reclassifications or new combinations would be difficult to achieve. The solution adopted here is to use original orthography for species group names, except where modified by ICZN rules other than gender agreement. In practice gender-neutral or "fuzzy" database searches are relatively straightforward to implement; partial name searches are already implemented in LepIndex, so that the controversy affects the major function of *LepIndex* very little.

vii) Institutional support and partnerships. GloBIS will collaborate closely with specialists by major taxonomic group for each of the six butterfly families, and also at regional level. For example, the African Butterfly Research Institute in Nairobi, under the direction of Steve Collins, would advise on the Afrotropical taxa. For North America, GloBIS would seek guidance from, among others, Paul Opler (see Opler et al. 1995, 2002), and Jonathan Pelham, currently engaged in developing a new synonymical checklist for the Nearctic fauna, and working in close collaboration with G. Lamas. For South East Asia and the Far East, we will collaborate with scientists such as Osamu Yata (Kyushu University), Frank Yu-Feng Hsu (National Taiwan Normal University, Taipei) and Shen-Horn Yen (Department of Biological Sciences, National Sun Yat-Sen University & Academia Sinica, Taipei). For the Indian Region, Harish Gaonkar would advise. For the Australian region, we would contact Michael Braby, Ted Edwards and Mike Parsons. In other words, the necessary network is in place to provide peer review and quality control.

viii) Technical and human resources required, project management. GBIF funds are chiefly required to bring Gerardo Lamas, lead author of the Neotropical checklist, to London for 5 months. A GloBIS member and ICZN commissioner, he adds world expertise to the project, and his taxonomic and nomenclatural knowledge is essential both in determining which database name or other field data should be accepted whenever there are conflicts, and also in locating other suitable checklists and expert help for further additions to *LepIndex*. Dr. Lamas has been budgeted for a 5-month salary of UK Senior Lecturer/Reader equivalent, concomitant with his status as senior researcher and professor in Peru and with his international status as a major world taxonomist. Once the information from the Neotropical checklist is entered, Dr. Lamas will decide on the taxonomic placement of all non-Neotropical taxa in *LepIndex* currently assigned to genera present in the Neotropical region (taxonomic concepts will often differ between the checklist and the NHM card index on which *LepIndex* is based). Dr. Lamas' expertise, in collaboration with other GloBIS members, is also necessary to enter 'taxonomic rank' and 'taxonomic status' fields into *LepIndex* for all taxa not present in any of the other electronic databases or checklists used.

James Mallet, working closely with checklist authors Gerardo Lamas, Dick Vane-Wright, Phil Ackery, and Campbell Smith, undertakes to manage the project and to write the parsing programs. Programs and merge queries are simple to adapt to other large-scale checklists already available online (for instance the GART/GloBIS Papilionidae list), or as word-processed files (many of the checklists in the references below). Additional programming will also be need to be carried out, funded by GBIF for two weeks' work, to ensure WWW checklist delivery to GBIF. Responsibility for database integration and any manual data entry will lie with James Mallet, Gerardo Lamas and Fraser Simpson, assisted by George Beccaloni, designer of *LepIndex*, Brian Pitkin, designer of the *Generic Names* and Bridges' databases, as well as Dick Vane-Wright and Campbell Smith, co-authors of the Afrotropical checklist.

ix) Efficiency and metrics: numbers of species-group names existing, to be added, and costs. Some idea of the relative scope and coverage of the existing *LepIndex* database and the *Atlas of Neotropical Lepidoptera Checklist* (Lamas et al. 2004a) can be gauged using a sampling approach. For *Dione*, *Heliconius* and related genera (today classified as sub-tribe Heliconiina, but formerly treated as tribe Heliconiini, and currently listed as subfamily

Heliconiinae in *LepIndex*), 1,149 names are included. In *LepIndex*, only 926 cards currently exist for this group. Since few names were added since the early 1980s (90 in the checklist), and since the heliconians were of particular interest to Museum staff, it seems that *LepIndex* could be only about 80% complete. The deficit of names in the current *LepIndex* may be due largely to weaker coverage of less important unavailable names, as well as new names. However, detailed research by Dr. Gerardo Lamas (sole author of the Heliconiinae and many other parts of the Neotropical list) over the last 20 years has also revealed many taxonomic and nomenclatural problems present among Neotropical butterflies before 1981, when the card index was discontinued due to NHM staff shortages. Without completing the GBIF project outlined here, it is hard to gauge the overall completeness of *LepIndex* for the rest of the butterflies, but it is clear that butterfly scientific names will be considerably improved, maybe by as much as 25% to a total of well over 80,000 names, if this project is funded.

Global Butterfly Names will provide complete taxonomic information to GBIF according to the highest world taxonomic standards for some 80,000 scientific names in 17,500 species at a cost of $0.63¢/name, or $2.86¢/species. However, this work will automatically and immediately entail full access to the complete *LepIndex* database of 290,000 names (223,200 currently online) for all 150,000 Lepidoptera species, resulting in a cost to GBIF of $0.17¢/name, or $0.33¢/ species. In addition, a new plan at the NHM has just extended the *LepIndex* model to the Coleoptera, which will ultimately result in 500,000 names in the Coleoptera catalogue. When this data comes online, it will be straightforward to provide GBIF access to virtually complete taxonomic names for about 25-30% of the world's biological species at a cost of $0.06¢/name and $0.11¢/species (assuming $50,000 GBIF funding for the butterflies only).

x) Strategy for maintenance of the product after GBIF funding ends (see also under *v* above). *Global Butterfly Names* employs a single classification approved by a world-wide panel (Lamas et al. 2004b), but the classification is easy to update as other experts begin to work on particular groups. *LepIndex* users can already log in to add corrections to *LepIndex*, such that the data reflects world taxonomic opinion. Because all data is shared openly, we envisage that others will make use of our data as a first step in taxonomic work on a group of butterflies; then, having carried out new research, they will reciprocate by contributing revised data to future versions of the database. NHM is committed to long-term maintenance of online data (see *v* above) into the foreseeable future, and *LepIndex* is one of its major projects: the highest worldwide standards for long-term data curation and open access will be in operation. We also welcome the possibility of mirroring this database at other sites to ensure maximum long-term data security. In case of conflict between different taxonomists, a system of reviewing will be introduced in order that *LepIndex* can provide an accessible basis for a global consensus of butterfly taxonomic information for the future, in accord with GBIF philosophy.

Although not part of this project, it is intended soon to extend *LepIndex* in various ways. An obvious and very important extension would be to provide photographs of type specimens in the same way that images of index cards are currently delivered. The database resulting from this project will form the taxonomic core, for the butterflies, of the overall Lepidoptera Taxome Project (www.ucl.ac.uk/taxome/), in which complete and global taxonomic, biological,

bibliographical and distribution information, together with photographs of types and other representative specimens, as well as links to molecular genealogy information for genera (coverage is already ~ 25% for some mtDNA and nuclear loci), and an NSF proposal to complete a generic phylogenetic classification of the butterflies and taxonomy informatics using sequence data has been submitted by members of our group earlier this year. This information will all ultimately be accessible from a single website. All such information will ultimately be made freely available online for all approx. 180,000 species of Lepidoptera within five years, given appropriate funding. An additional goal is to add online images of the underside and upperside of type specimens for each name, which would be obtained in the first instance by raiding taxonomists' existing collections of type photographs and using bulk-scanning equipment to digitize the images. Adding information about types and DNA sequences is not a component of this GBIF proposal, but it is a long-term aim of the Lepidoptera Taxome Project and *LepIndex* to have these fields fully documented where information exists. Finally, many types lie unnoticed and unlabeled within major collections, and, particularly for Lepidoptera, within the NHM collection itself; these would need to be hunted down and digitized, but cataloguing the existing literature information, as proposed here, will provide the essential scaffold for the more complete database information envisaged in the Taxome Project.

xi) Risk assessment. The main risk to this project would be the incapacitation or unavailability of its main scientist, Gerardo Lamas; this would be a major blow to butterfly taxonomy in general, as well as to this GBIF project. However, since we hold electronic copies of the Neotropical checklists generated by Dr. Lamas and colleagues, we would be able to cover much of the work, even in this eventuality. Alternative butterfly taxonomists would need to be hired to make the important taxonomic decisions, but several younger taxonomists are available who could substitute for these tasks.

While the loss of Dr. Lamas would undoubtedly be a problem, this project has as a major aim to document electronically and add to the information Lamas alone has at his fingertips, as rapidly as possible. An even worse risk for world butterfly nomenclature would be that Dr. Lamas' nomenclatural researches became limited to traditional, small-circulation printed and copyrighted documents.

xii) Documentation of lessons learned. We very much regard butterflies as suitable model organisms for taxonomy informatics, among many other disciplines (Ehrlich & Hanski 2004), and we undertake to document any problems we encounter and short-cuts we discover as we carry out this innovative project. All checklist parsing programs or macros will be made available to others who may use them, though because each checklist has distinct coding standards, such programs usually require considerable alteration. Nonetheless, programming will aim for generality wherever possible and the more general we can make such programs the more useful they will be both for our own checklists and for others working on different organisms. Finally, we adhere to a completely open-information policy, and will broadcast any techniques and findings from this work in replies to email requests or as online or printed scientific and popular articles.

c) Who will benefit from Global Butterfly Names?

i) What are the benefits to user communities? It might be argued that *LepIndex* is of interest mainly to taxonomists, rather than to a broader, biodiversity-oriented user community. However, it is essential that conservation decisions and biodiversity management goals depend on the best available science. In particular, taxonomic inflation is such a major problem (Isaac et al. 2004) that it is simply not feasible to support management decisions based on out-of-date or static lists. *Global Butterfly Names* will validate and add to *LepIndex* the most current information, as well as completing classification details to allow *LepIndex* to act as the main global source for Lepidoptera scientific names and classification, allowing the most informed possibilities for name standardization to all user communities. We are committed to making LepIndex an ECAT Data Node for GBIF, and links are underway with Species 2000; in the long term, the underlying data will be made freely available to the entire global biodiversity community and especially to any other taxonomic data servers that adhere to an open-access ideal. We have already received an approach from Peter Hall, Director of Biodiversity Information Services at the United Nations Environment Programme-World Conservation Monitoring Centre in Cambridge about the possibility of incorporating our data into a world list of butterflies for conservation purposes. In the future, *LepIndex* will be linked via GBIF or other sites to biological data such as photographs, distribution data, life history and ecological interactions, &c.

ii) What are the benefits to global biodiversity science? Currently, taxonomic information on butterflies is buried in myriad obscure print publications. We aim to make complete, global, up-to-date information on taxonomy of butterflies freely available to the widest possible user community. In addition, alternative taxonomic arrangements can be easily implemented by organizations that disapprove of the standard *LepIndex/GLOBIS* classification; their data nodes will be able freely to obtain information from our online nameserver, while rejecting our preferred classification. Our open provision of data encourages, but does not enforce a standardized, consensus classification preferred by acknowledged world experts.

Without GBIF funding, the chance for validating and making the full *LepIndex* records for butterflies available will be lost, and nomenclature information for this important and charismatic group will remain unavailable to the majority of the world's scientists, as at present. Successful implementation of this GBIF-funded project, on the other hand, will show our administrators the value of our overall approach to Lepidoptera, and now Coleoptera scientific names at the NHM. We see GBIF seed funding as essential for leveraging larger projects to be fully funded from other sources in the future.

d) Capacity building. The Natural History Museum, UCL, and collaborating institutions are committed to capacity-building, especially where it involves international links or developing countries. We provide training and qualifications to the highest international standards of scientific taxonomy, to masters, PhD, and postdoctoral levels. Current and past examples for Lepidoptera systematics include: yearly MSc courses in Taxonomy and Biodiversity, MRes courses in Biosystematics (many international students, held in NHM and Imperial College London), Angel Viloria (Venezuelan, PhD, NHM and University of London (UL) 1998), Shen-Horn Yen (Taiwan, PhD, NHM and UL in 2004), Margarita Beltrán (Colombian, PhD, UCL, expected 2004), Carlos Peña (Peruvian, PhD, Univ. Stockholm, expected 2008), Dr. Igor Emelianov (Belarus, BBSRC postdoctoral fellow, 1993-2001), Dr. Jesús Mavárez (Venezuelan, EU Marie Curie postdoctoral fellow, UCL, 2002-2004).

References

Ackery,PR; Smith,CR; Vane-Wright,RI (1995): Carcasson's African Butterflies: an Annotated Catalogue of the Papilionoidea and Hesperioidea of the Afrotropical Region. CSIRO & British Museum, Natural History Publications, East Melbourne & London.

Bascombe,MJ; Johnston,G; Bascombe,FS (1999): The butterflies of Hong Kong., London.

Beccaloni,GW; Scoble,MJ; Robinson,GS; Pitkin,B (2004): The Global Lepidoptera Names Index (*LepIndex*). The Natural History Museum, London. (*http://ww.nhm.ac.uk/entomology/lepindex/*)

Boggs,CL; Watt,WB; Ehrlich,PH (Eds.) (2003): Butterflies: Ecology and Evolution Taking Flight. Univ. Chicago Press, Chicago.

Braby,MF (2000): Butterflies of Australia: Their Identification, Biology and Distribution. CSIRO, Collingwood. 1008 pages.

Bridges,CA (1979-94): Catalog of the Names of the Butterflies. 2nd ed. C.A. Bridges, Urbana, Ill. Held in electronic form at the NHM, London.

Chou,I (1998): Classification and Identification of Chinese Butterflies. Henan Scientific and Technological Publishing House.

Corbet,AS; Pendlebury,HM; Eliot,JN (1992): The Butterflies of the Malay Peninsula. 4th ed., Kuala Lumpur.

de Jong,R; Treadaway,CG (1993): The Hesperiidae (Lepidoptera) of the Philippines. Zoologische Verhandelingen 288, 116-121.

Ehrlich,PR; Hanski,I (2004): On the Wings of Checkerspots. A Model System for Population Biology. Oxford University Press, Oxford.

Häuser,C; de Jong,R; Lamas,G; Robbins,RK; Smith,CR; Vane-Wright,RI (2001): Papilionidae - Revised GloBIS/GART Species Checklist. 2nd ed. Staatliches Museum für Naturkunde, Stuttgart. (*http://www.insects-online.de/frames/papilio.htm*)

Hemming,F (1967): Generic names of butterflies of the world and their type-species (Lepidoptera: Rhopalocera). Bull. Brit. Mus. (Nat. Hist.), Entomol. 1967 (Suppl. 9), 1-509. [Now superseded by Pitkin & Jenkins (2002)].

Isaac,NJB; Mallet,J; Mace,GM (2004): Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology and Evolution (in press). (*http://abacus.gene.ucl.ac.uk/jim/pap/isaac04.pdf*)

Karsholt, O; Razowski, J (1996): The Lepidoptera of Europe: a Distributional Checklist. Apollo Books, Steenstrup.

Lamas,G; Robbins,RG; Field,WD (1995): Bibliography of Butterflies. An Annotated Bibliography of the Neotropical Butterflies and Skippers (Lepidoptera: Papilionoidea and Hesperioidea). (Series Ed: Heppner,JB. Atlas of Neotropical Lepidoptera, 124.) Association for Tropical Lepidoptera, Gainesville, Florida.

Lamas,G; Callaghan,C; Casagrande,MM; Mielke, O; Pyrcz,T; Robbins,R; Viloria,A (2004a): Hesperioidea and Papilionoidea. (Series Ed: Heppner,JB. Atlas of Neotropical Lepidoptera, Checklist: Part 3.) Association for Tropical Lepidoptera, Gainesville, Florida (in press).

Lamas,G; Häuser,C; de Jong,R; Nielsen,E; Robbins,RK; Vane-Wright,RI (2004b): Global standards for higher classification and family group names of the butterflies (Lepidoptera). (submitted).

Lees,DC; Kremen,C; Raharitsimba,H (2003): Classification, diversity and endemism of the butterflies (Papilionoidea and Hesperioidea): a revised species checklist. In: The Natural History of Madagascar. (Eds: Goodman, SM; Benstead,JP) University of Chicago Press, Chicago, Illinois, 762-793.

Lees,DC; Minet,J (2003): Lepidoptera: systematics and diversity. In: The Natural History of Madagascar. (Eds: Goodman,SM; Benstead,JP) University of Chicago Press, Chicago, Illinois, 748-761.

Mallet,J; Willmott,K (2003): Taxonomy: renaissance or Tower of Babel? Trends Ecol. Evol. 18, 57-59.

Maruyama,K (1991): Butterflies of Borneo. Vol. 2(2). Hesperiidae. Tobishima Corporation. 11+69 pages.

Nielsen,ES; Edwards,ED; Rangsi,TV (1996): Checklist of the Lepidoptera of Australia. (Monographs on Australian Lepidoptera, 4.) CSIRO, New Melbourne.

Opler,PA; Stanford,RE; Pavulaan,H (Eds.) (1995): Butterflies of North America. 30 Dec 2002 ed. Northern Prairie Wildlife Research Center Home Page, Jamestown, ND. (*http://www.npwrc.usgs.gov/resource/distr/lepid/bflyusa.htm*)

Opler, PA; Warren, AD (2002): Butterflies of North America. 2. Scientific Names List for Butterfly Species of North America, north of Mexico. Contributions of the C.P. Gillette Museum of Arthropod Diversity, Colorado State University, Fort Collins. 79 pages.

Otsuka,K (1988): Butterflies of Borneo. Vol. 1., Tokyo. 162 pages.

Parsons,MJ (1998): Butterflies of Papua New Guinea, Princeton University Press, Princeton NJ. 928 pages.

Pelham, J (in preparation): Checklist of Butterflies of the Nearctic.

Pitkin,B; Jenkins,P (2004): Butterflies and Moths of the World. Generic Names and their Type-Species. The Natural History Museum, London. (*http://www.nhm.ac.uk/entomology/butmoth/*)

Ruggiero,M; Bisby,F; Wilson,K; Shimura,J (2002): Towards a "Catalogue of Life". Final Report of the GBIF Scientific and Technical Group (STAG) Workshop on An Electronic Catalogue of Names of Known Organisms (ECAT). Sydney, Australia, 15-16 March 2002.

Seki,Y; Takanami,Y; Otsuka,K (1991): Butterflies of Borneo. Vol. 2(1). Lycaenidae. Tobishima Corporation. 10+103 pages.

Tennent,J (1996): Butterflies of Morocco, Algeria, and Tunisia., Wallingford, Oxon. 243 pages.

Tennent,J (2002): Butterflies of the Solomon Islands: Systematics and Biogeography., Dereham, Norfolk.

Treadaway,CG (1995): Checklist of the butterflies of the Philippine Islands. Nachr. entomol. Ver. Apollo N.F. 16 (Suppl. 14), 7-118.

Vane-Wright,RI (2003): Indifferent philosophy versus Almighty Authority: on consistency, consensus and unitary taxonomy. Syst. Biodiv. 1, 3-11.

Vane-Wright,RI; de Jong,R (2004): The butterflies of Sulawesi: annotated checklist for a critical island fauna. Zoologisches Verhandelingen 343, 1-267.

Vane-Wright,RI; Humphries,CJ; Williams,PH (1991): What to protect? – systematics and the agony of choice. Biol. Conserv. 55, 235-254.

Williams,MC (2003): Afrotropical Butterflies and Skippers. A Digital Encyclopaedia. Privately published CD-ROM, Rietondale, South Africa.