Available Online at ESci Journals
International Journal of Entomological Research
ISSN: 2310-3906 (Online), 2310-5119 (Print)
http://www.escijournals.net/IJER

ENTOMOLOGICAL DATABASES: CHALLENGES AND OPPORTUNITIES IN DATA MANAGEMENT AND RETRIEVAL

[ab]Luke Melita*, [b]Gopinath Ganapathy, [c]Kaliyaperumal Karunamoorthi, [d]Sebsibe Hailemariam
[a] *Department of Computing, Jimma University, Jimma, Ethiopia.*
[b] *School of Computer Science and Engineering, Bharathidasan University, India.*
[c] *Department of Environmental Health, Faculty of Public Health and Tropical Medicine, Jazan University, KSA.*
[D] *Department of Computer Science, Addis Ababa University, Addis Ababa, India.*

A B S T R A C T

Documenting and preserving the entomological data is an important aspect in the fight against insect-transmitted diseases in terms of public health and agriculture concern and it is also helpful in formulating the life-saving novel chemotherapeutic as well as immunotherapeutic agents. The objective of this scrutiny is to assess the existing entomological databases and to identify the major challenges and future perspectives in terms of effective information retrieval. To pursue with the study, a detailed search on Google and American Online Search engines has been carried out with the key words "Entomology", "Entomology Databases", "Entomology Collections", "Insects", "Insect Collection" and "Insect Database", in differing orders, in order to extract websites or databases, containing information on Entomology and entomological databases. In this scrutiny among the output of first 253 web-links, 67 appropriate websites/databases were selected for further analysis. From the observed data, several important inferences are made such that effective counter-measures can be taken-in-part. A data input standard for framing an ideal database management system and some methods for retrieving entomological data effectively through different kinds of interfaces based on ontologies, common framework or meta-search have been suggested. The findings recommend for renewed interest and support in establishing entomological databases in the resource-constrained settings with efficient platforms for distributed retrieval of information.

Keywords: Entomological databases, Biological Databases, Integration, Information Retrieval.

INTRODUCTION

Biological databases meant for storing and organizing biological data, collect information from life sciences, scientific experiments, computational analyses and contents from published literatures (Wren and Bateman, 2008; Attwood *et al.,* 2013). Varieties of information like genomics, proteomics, metabolomics, microarray gene expression, phylogenetics etc are also stored and maintained as biological databases in terms of genome databases, protein structure databases, protein sequence databases, taxonomic databases etc (Altman, 2013). These computerized web databases, facilitate effective data management and analysis, help in knowledge transmission, make coherent information to be available to researchers, planners and other users, and feed the circle of information exchange between the studies and public audience (Ningthoujam *et al.,* 2012).

* Corresponding Author:
Email: karunamoorthi@gmail.com

Significance of Biological Databases: Biological databases play a vital role in bioinformatics, since it helps varieties of researchers to access and analyze data from different parts of the world. The knowledge obtained helps them to address with disease-oriented and environment-related issues and to make vital policy decisions. Biological databases remain as a tool to identify insect species as well as in comparing their relationships with related species. This biological knowledge distributed among different general and specialized databases, requires to be retrieved with efficient queries.

Importance of Entomological Databases: Insects (Insecta) are the most diverse of all animal groups. Until today, over one million insect species were described/identified, which represents more than half of all known living organisms (Chapman, 2009; Wilson, 2006). It has been estimated that nearly six-to-ten million species may exist worldwide (Chapman, 2009; Erwin, 1982; 1997), and potentially they represent over ninety-percent of the differing metazoan life forms on Earth (Erwin, 1982). However, less than one percent insect species are serious pests that affect mankind, livestock and crops. Though these pests or vectors are very small in number, they are able to cause serious negative socio-economic, public health and clinical impacts by means of low yield through transmission of several destructive diseases to various crops**,** humans and animals (Tripepi *et al.,* 2013). In these perspectives, understanding about entomology is quite imperative to maintain good health and to enhance the food supplies. Insect-transmitted diseases impose an enormous burden on the world population in terms of loss of life (millions of deaths per year) (Jacobs-Lorena, 2006). Since humankind often suffer due to various vector-borne diseases, particularly in the resource-limited settings, it calls for a serious documentation/databasing of insect pests/vectors, particularly the taxonomic details of insect species, those that often impact on the wealth as well as the health of humanity. Taxonomy that involves the identification, classification and naming of organisms, needs to be documented or recorded systematically based on the phylogenetic relationships of the arthropods and related groups (Edwards and Cavalli-Sforza, 1964). At the moment, fairly a large number of websites and databases exist to document, store and maintain the entomological data. However a few of them only are well-organized with reference links and impotent articles, but then the majority of them contains only images, with insufficient details. This is an attempt to explore with the available entomological databases, their data organization, structure and management. In these contexts, this review becomes more significant and pertinent. It is an attempt to identify the potential barriers of the existing EDs in terms of retrieval of information, and to identify the emerging opportunities to design ideal EDs in the near future.

MATERIAL AND METHODS

Data Extraction Procedure: In order to collect the appropriate research materials for the present scrutiny, a detailed search on Google and American Online Search engines has been carried out for the time period January 2013 – November 2013. A Boolean search strategy was adopted and the key words entered for search are “Entomology”, “Entomology Databases”, “Entomology Collections”, “Insects”, “Insect Collection” and “Insect Database” in differing orders, in order to extract websites or databases, containing information on Entomology and entomological databases.

Data Extraction: Concerning with the data extraction, the most appropriate web-links and databases were selected for the present investigation. The selected websites/databases were carefully reviewed and critically analyzed in terms of their content type, specialization, origin and facilitating software tools. The websites of entomology-based educational institutions, organizations, societies and databases were analyzed. Subsequently institutions without entomological collections or databases were excluded in the later search. The databases or files which are specific to only one species were also excluded in the later phase, in due requirement for identifying more databases containing wide-spread entomological data. From a total of 253 web-links, excluding the repeated websites, 75 websites were further analyzed and finally the most appropriate 67 websites with or without databases were selected for further investigation. These selected pages were analyzed for educational and research details, established year of entomological institution, whether the entomological websites contain all or specific information, whether the information provided is structural or textual or file formatted, whether it contains images, and special mode of searching, the availability and accessibility of research articles, and their extensibility, integration, annotation and versioning possibilities etc. The identified databases are analyzed for the information categories and quality, and for their mode of information retrieval.

Data Processing and Analysis: Each collected data was cleaned, checked for completeness, coded and analyzed with an IBM compatible micro computer using the statistical package for the Social Science (SPSS) (Window version 16.0, Chicago, IL USA) for computing statistics and frequency distributions. Relevant tables and figures are used to display with the results. Range and mean were analyzed and appropriate tables, graphs and percentage were displayed.

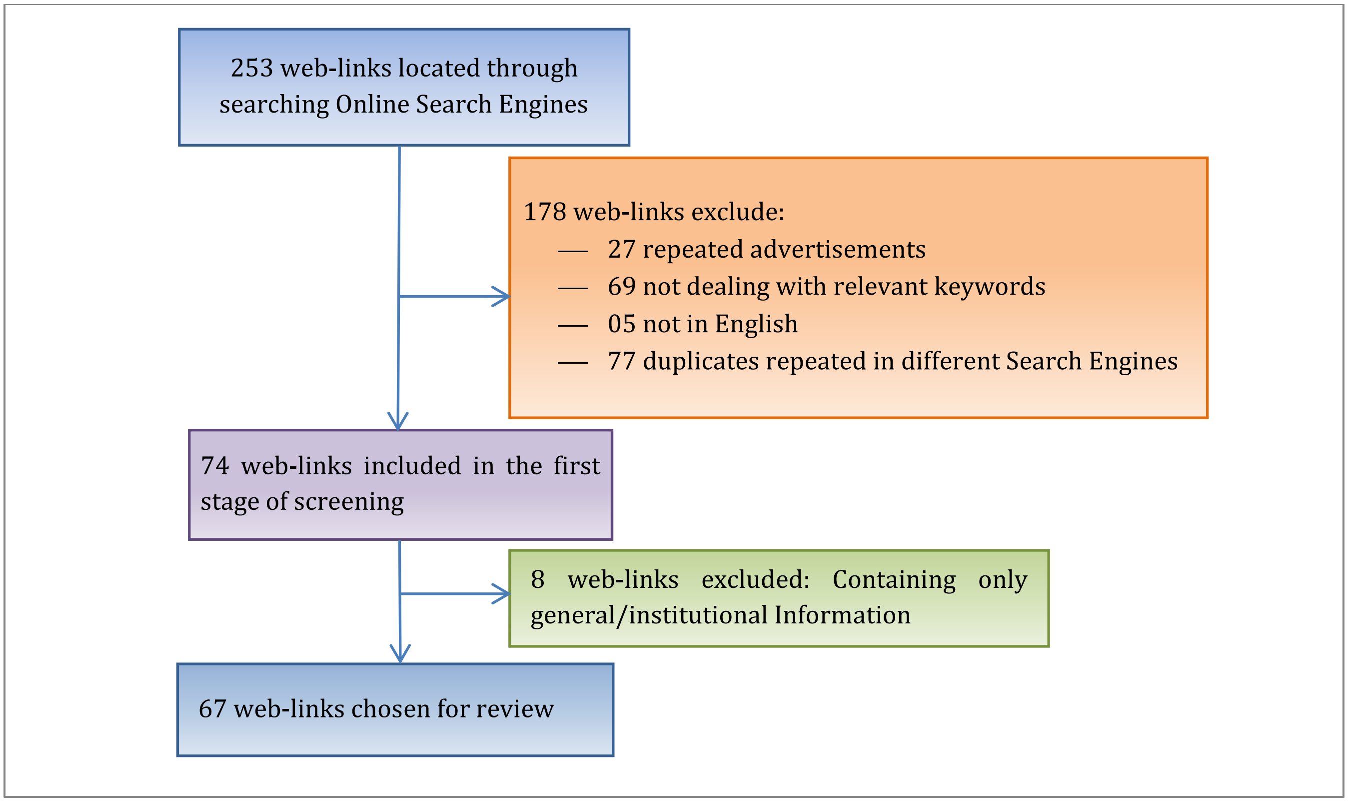

Figure 1. Inclusion and Exclusion Criteria for the Selection of Entomological Websites/Databases for the present scrutiny.

RESULTS AND DISCUSSION

The data collected from various websites are listed in the following tables in Table 1 and Table 2. Details about the entomological databases in terms of distribution, content, accessibility, quality of data and structured databases are shown in Fig. 1, 2 and 3. By using the proper Boolean search strategy, the present study has identified nearly thirty three properly organized databases from the first 253 web-links (Table 2). The process of Insect identification, classification and data storage has been described in the fig. 4. Table 3 indicates some of the important parameters that are to be considered for the construction of idealistic entomological databases in the near future.

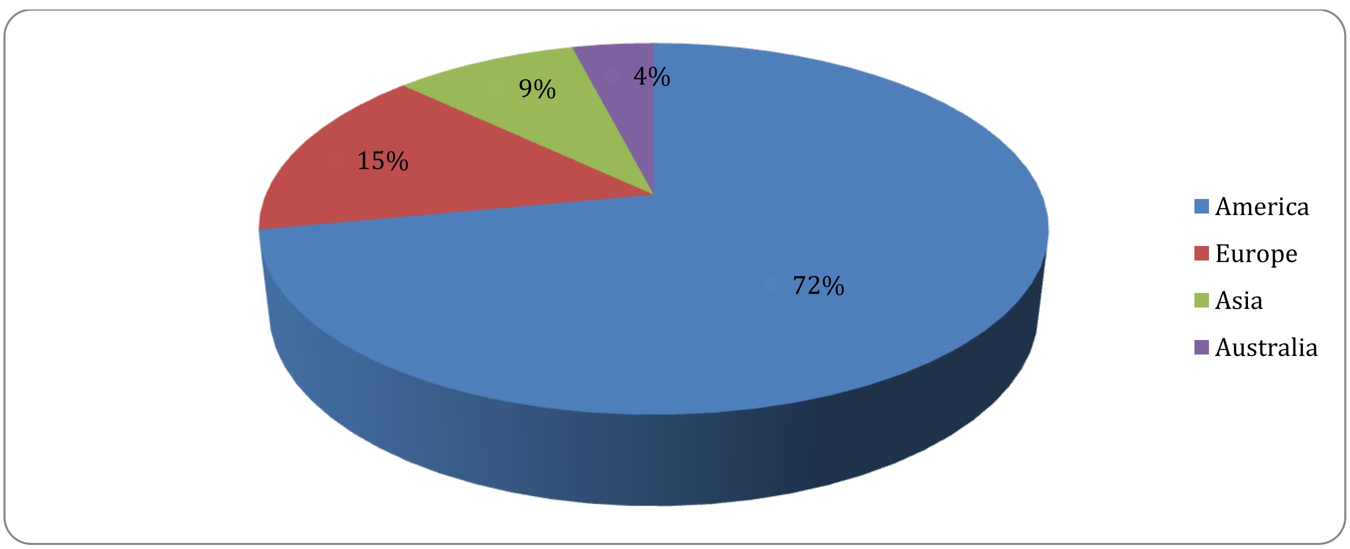

Figure 2. Distribution of selected entomological websites continent-wise.

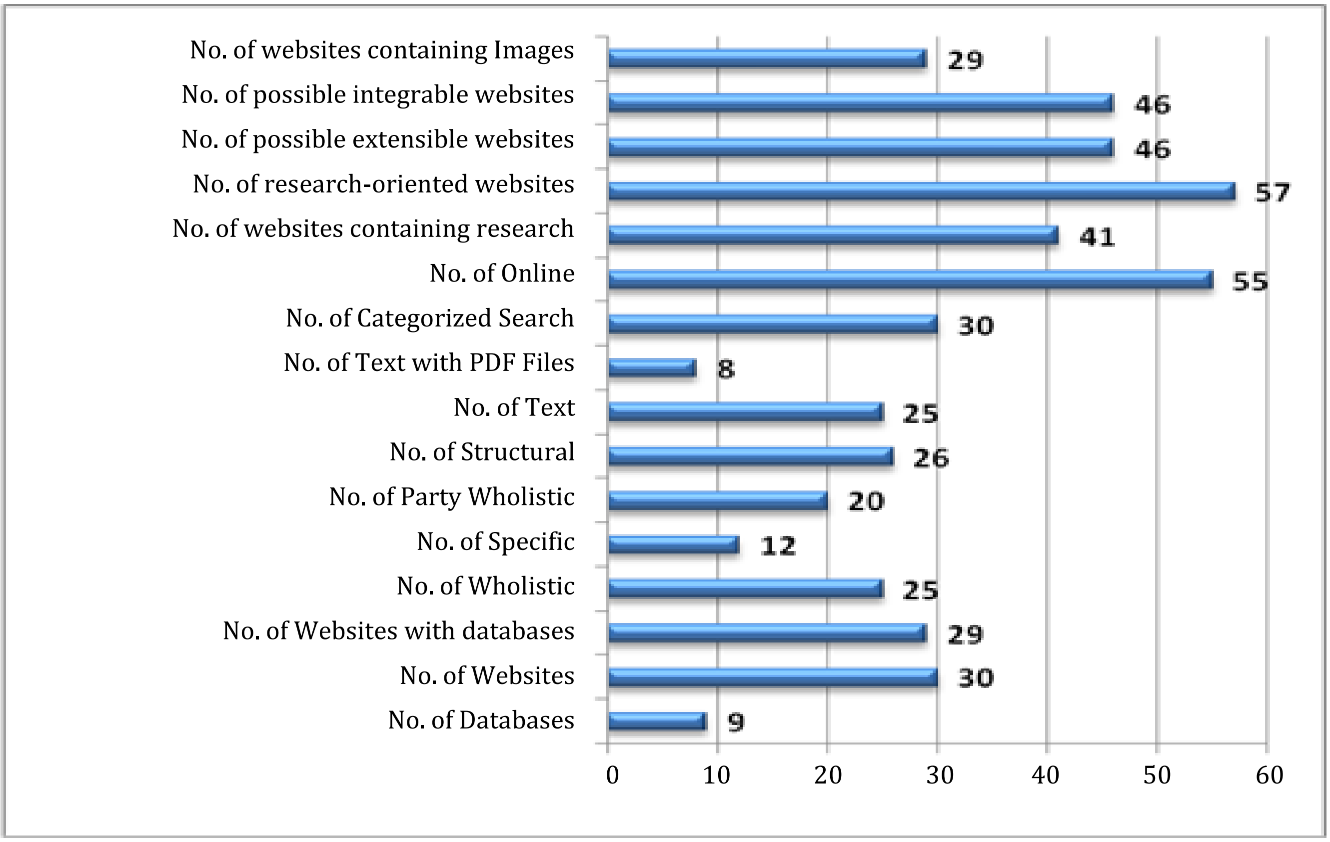

Figure 3. Outline of content, accessibility and quality of data among entomological websites.

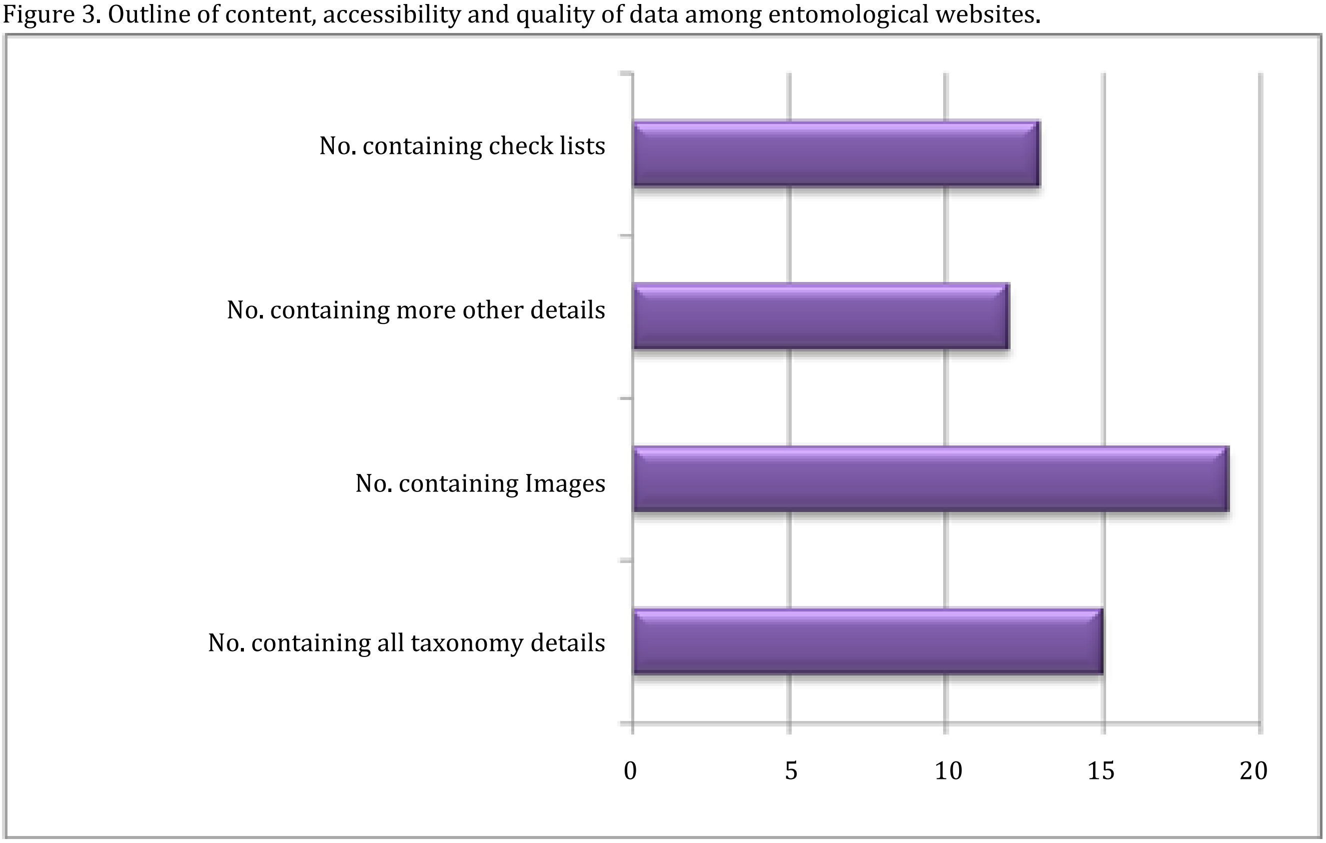

Figure 4. Summary of Database content among structured databases.

The present study found that among the total of 67 website studied with, 48 were from America, 10 from Europe and rest of them from the Asia, Australia and none of them from Africa (Figure 1). About 30 websites contained either general information about the institution or basic information on entomology or description on the projects or research carried over by the faculty or the insect identification service (Table 1). It has been estimated that only 25 websites had wholistic approach on the spread of data and they covered all parts of entomology, whereas 12 were species-specific and 20 contained some or more of the species, but not all. Only 6 institutions had both specific databases and databases for the rest of all species or some species. Figure 1 clearly shows that the entomological databases are well established in the developed counties than in the developing economies. It could be possibly explained that the poor infrastructure is due to the resource constraints in terms of lack of skilled personnel, expertise, scientific technology and more particularly the financial constraints (Woldetensae, 2007). These have to be addressed effectively in order to address with the vector-borne diseases burden and to enhance with the agricultural productivity by identifying the major deadly pest species' and to adopt appropriate pest control strategies. It has been identified that approximately only 30 web-links had a categorized search facility in order to search on the taxonomic details or other relevant details. Although 55 websites have the online access facility, 6 of the most important entomological websites were restricted with access. This restricts access by the individuals other than the members of the institution and remains as a major constraint for researchers, individuals and students from other parts of world to obtain necessary details. This issue can be addressed by providing concession or some percentage of subsidies, as provided by Health Inter Network Access to Research Initiative (HINARI) for researchers from the developing countries. In addition, the findings clearly show that only 41 websites have the facility to access various research articles; however, some of them offered access to the articles by purchase or upon a special request. It has been estimated that out of all websites analyzed, up to 85% of websites were research-oriented. It is interesting to note that nearly 69% of websites had the capability of extending their features to adapt integration of databases (Figure 2). Table 2 indicates that only less than half (43%) of the websites, contained information in the form of database, which shows that stringent measures are to be taken, to document insect information worldwide. Moreover, this calls for the idea of integrating databases or some other means, to provide users with a unified view of these data (Lenzerini, 2002). This process becomes significant in a variety of situations, which include both commercial (when two similar companies need to merge their databases) and scientific (combining research results from different bioinformatics repositories, for example) domains. Data integration is highly necessitated with increasing frequency as the volume and the need to share existing data explodes (Lane, 2006). Though the data acquired are organized, only 37% of them contain information about a lot of species and the rest contain very less information. Apart from the literature databases, among the data organized for insect species, only 39% are of structured datatype, whereas the rest contain massive text descriptions or articles in PDF file format (Table 2). Among the structured data, nearly 45% have the efficient way of searching and retrieving such data, unless otherwise only browsing through links has been made possible (Table 2). In certain cases, it has been observed that data are compiled as text files (12%), with enormous parameters. This suggests that the databases constructed and to-be-constructed must follow a well-defined data model and a proper mode of search and this is discussed in detail in the Data Standards section.

Systematic Databases: The European Database Directive defines a database as "a collection of works, data or other independent materials arranged in a systematic or methodical way and capable of being accessed by electronic or other means" (Hunsucker, 1996). Indeed nearly 87% (33/38) of entomological databases were observed to be systematically designed with proper data organization and management. Among these databases, 45% of them had each and every taxonomic detail of all the species and nearly 58% contained details with images (Figure 3). Among the 33 databases, nearly 21% of the databases were from several museum online databases, which shows that they are not independent or sole entomological databases and they are only a part of all kinds of exhibits (Figure 3).

Table 1. List of Entomological Websites and their details [explain content and abbreviations in one sentence].

Sr.	Continent	Country	Concern	UG	PG	Research	Year	DB/Web-site	Wholistic/ Specific/ Partly Wholistic	Structural/ File / Textual	Less/ Categor-ized Search/ Browsing/ Detail	Access - Online/ Offline/ Restricted	Annotation/ Versioning Possibility	Contains Research Articles?	Research/ Basic Knowledge	Extensibility /Integrity Possibility	Images/ Videos
1	America	US	University of Kentucky	Yes	Yes	Yes	1891	Website	Specific, Partly Wholistic	Text, PDF file	less	Online	No	No	Basic	No	NA
2	America	US	Iowa State Univerity	Yes	Yes	Yes	1880	DB/Website	Wholistic	Text	less	Online	No	No	Basic	No	NA
3	America	US	Entomological Society of America	No	No	Yes	1889	DB/Website	Wholistic	Structural	Categorized Search	Online - Restricted	Yes	Yes	Research	Yes	NA
4	America	US	University of Alaska Museum of North	No	No	Yes	2000	DB/Website	Partly Wholistic	Structural	Categorized Search	Online	No	No	Research	Yes	NA
5	America	US	Auburn University	Yes	Yes	Yes	NA	Website	Partly Wholistic	Text	less	Online	No	Yes	Basic	No	NA

6	America	US	Smithsonian National Museum of Natural History	No	No	Yes	1910	DB/Website	Wholistic	Structural, PDF	Categorized Search	Online - Restricted	Yes	Purchase	Research	Yes	Images
7	America	US	University of California/ Bohart Museum of Entomology	Yes	Yes	Yes	1946	Website	Partly Wholistic	DB under Construction	NA	NA	NA	NA	NA	NA	Images
8	America	US	Texas A & M University	Yes	Yes	Yes	1977	Website	Wholistic	Text, PDF	Articles	Offline	NA	Yes	Research	Yes	Images
9	America	US	Cornell University	Yes	Yes	Yes	1870	Website	Specific (Beetle)	Text, PDF file	Less Articles	Offline	No	A few - yes	Research	No	NA
10	America	US	Penn State University	Yes	Yes	Yes	NA	Website	Wholistic	Text	Less Articles	Online	No	A few - yes	Research	No	NA
11	America	US	Purdue University	Yes	Yes	Yes	1884	Website	Partly Wholistic	Text, PDF file	Articles	Online	No	Yes	Research	Yes	NA

12	America	US	University of California, Riverside	Yes	Yes	Yes	NA	Website	Partly Wholistic	Text	Less Articles	Online	No	No	Research	No	NA
13	America	US	Michigan State University	Yes	Yes	Yes	NA	Website	Specific	Text	NA	NA	No	No	NA	No	NA
14	America	US	Rutgers University	Yes	Yes	Yes	1888	Website	Partly Wholistic	Text	less	Online	No	No	Basic	No	NA
15	America	US	University of Wisconsin - Madison	Yes	Yes	Yes	1909	DB/Website	Partly W	Structural	Browse (link)	Online	No	No	Research	No	NA
17	America	US	California Academy of Research	No	No	Yes	1862	DB/Website	Wholistic	Structural	Categorized Search, General Collection DB	Online	Yes	Yes	Research	No	Images
18	America	US	University of Arkansas	No	Yes	Yes	1991	Website	Specific, Partly Wholistic	Images	Only Insect Identification Service	Online	No	Yes	Research	Yes	Images

19	America	US	University of Minnesota	Yes	Yes	Yes		DB/Website	Wholistic	Structural - Restricted	Categorized Search - Restricted	Online - Restricted	Yes	Yes	Research	Yes	NA
20	America	US	University of Arizona	No	Yes	Yes	1940	DB/Website	Wholistic, Specific - Bee, Butterfly	Structural - Restricted	Categorized Search - Restricted	Online - Restricted	Yes	No	Research	Yes	Images
21	America	US	University of California/ UCR Entomology Research	No	No	Yes	NA	DB	Wholistic	Structural	Categorized Search	Online - Restricted	Yes	Purchase	Research	Yes	Images
22	America	US	Harvard University/ Museum of Comparative Zoology (MCZ Type DB)	No	No	Yes	1636	DB	Wholistic	Structural	Categorized Search	Online	Yes	On request	Research	Yes	NA
23	America	US	Illinois State Museum	No	No	Yes	1877	DB/Website	Partly Wholistic	Text	Listing Only	Online	No	Yes	Research	Yes	NA

24	America	US	Hawai'I Biological Survey/ Bishop Musuem Entomological Types Database	No	No	Yes	1889	DB/Website	Wholistic	Structural	Categorized Search	Online	Yes	NA	Research	Yes	NA
25	America	US	Entophiles - Insect Pictures and Bio	No	No	Yes	NA	Website	Partly Wholistic	Text	less	Online	No	Yes	Research	Yes	Images
27	America	US	University of Colorado, Boulder/ Museum of Natural History	No	Yes	Yes	1902	DB/Website	Wholistic and Specific	Structural (Hierachical)	Categorized Search	Online	Yes	Yes	Research	Yes	Images
28	America	US	Hulett Environmental Services, Florida - Bug Database	No	No		NA	Website	Specific	Text	less	Online	No	No	Basic knowledge	No	Images
29	America	US	University of Nebraska State Museum	No	No	Yes	1886	DB/Website	Specific - Scarabeoidea, Partly Wholistic	Text, PDF file	Family Search	Online	Yes	Yes	Research	Yes	Images

30	America	US	Bio-diversity Information System, Kansas State University	No	No	Yes	1879	DB	Wholistic	Structural	Categorized Search	Online	Yes	List of Plants	Research	Yes	NA
31	America	US	University of Delaware	No	No	Yes	1973	DB/Website	Partly Wholistic	Text	less	Online	No	Yes	Research	Yes	NA
32	America	US	University of Florida, South West Florida Research and Education Centre	No	No		NA	DB/Website	Partly Wholistic	Text, PDF file	Only Articles	Online	No	Yes	Research	NA	NA
33	America	US	Mandala Database, University of Illinois	No	No		NA	DB	Specific - fly	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	Images
34	America	US	Armed Forces Pest Management Board	No	No	Yes	1962	DB/Website	Partly Wholistic	Text	less	Online	No	Yes	Research	No	NA

35	America	US	Oklahoma State University	Yes	Yes	Yes	NA	Website	Partly Wholistic	Text	less	Online	No	Yes	Research	No	NA
36	America	US	West Fly Database	No	No	Yes	NA	DB/Website	Partly Wholistic	Text	less	Online	No	Yes	Research	No	NA
37	America	US	Santa Barbara Museum of Natural History/ Online California Beetle Database	No	No	Yes	NA	DB	Specific - Beetle	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	Images
39	America	US	Antbase, American Museum of Natural History , Ohio State University	Yes	Yes	Yes	1995	DB/Website	Specific - Ant	Structural	Categorized Search, Browsing	Online	Yes	Yes	Research	Yes	NA

40	America	US	Sonoran Anthropod Studies Institute	No	No	Yes	1986	Website	Wholistic	Text, PDF File	less	Online	Yes	Yes	Research	Yes	NA
41	America	US	Blattodea Species File	No	No	No	2004	Website	Partly Wholistic	Text	less, categorized search	Online	Yes	Yes	Research	Yes	Images
42	America	US	US Geological Survey	No	No	Yes	NA	Website	Specific - Lepidoptera	Text	less, categorized search	Online	Yes	Yes	Research	Yes	NA
43	America	US	University of California/ Natural History of Orange Country, Irvine	No	No	Yes	NA	Website	Specific - butterfly	Text	less, tabulated	Online	No	Yes	Research	Yes	Images
44	America	US	University of Michigan, Museum of Zoology	No	Yes	Yes	1929	Website	Wholistic	Text	less	Online	No	Yes	Research	Yes	NA

45	America	US	American Museum of Natural history	No	No	Yes	1869	DB/Website	Wholistic	Structural	Categorized Search, Browsing	Online	Yes	Yes	Research	Yes	Images
46	America	Canada	Entomological Society of Canada, Biological Survey of Canada	No	No	Yes	1977	Website	Partly Wholistic	Text	Articles	Online	No	Yes	Research	Yes	Images
47	America	Canada	University of Alberta, E.H. Strickland Entomological Museum	No	No	Yes	1922	DB/Website	Wholistic	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	Images
48	America	Mexico	University of New Mexico, The Anthropod Museum	No	No	Yes	1980	Website	less	Text	less	NA	NA	NA	NA	NA	Images

50	Europe	United Kingdom	Natural History Museum/ The British Museum	No	Yes	Yes	1753	DB/Website	Wholistic	Structural	Categorized Search - less categories	Online	Yes	Purchase	Research	Yes	Images
51	Europe	United Kingdom	Royal Entomological Society, UK	No	No	Yes	1833	DB/Website	Unknown	NA	Categorized Search	Online - Restricted	Yes	NA	Research	Yes	NA
52	Europe	United Kingdom	Hosts Database, Natural History Museum	No	No	Yes	NA	DB	Specific - caterpillar	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	Images
53	Europe	Latvia, Baltic States	Entomological Society of Latvia	No	No	Yes	1951	Website	Partly Wholistic	Text	less	Online	No	Yes	Research	No	NA

54	Europe	Germany	Dermestidae of the World, by Andreas Herrmann	No	No	Yes	NA	Website	Specific - Dermestidae	Text	good explanation	Online	Yes	Yes	Research	Yes	Images
55	Europe	Germany	SYSTAX (DORSA), Museum of Natural History, Germany	No	No	Yes	1810	DB/Website	Wholistic	Structural	Search - Browsing	Online	Yes	Research Details	Research	Yes	NA
56	Europe	Germany	The Bavarian State Collection of Zoology	No	No	No	1759	Website	Wholistic	Text	Browsing	Online	Yes	Research Details	Research	Yes	NA
57	Europe	Germany	SeSam DB, Senckenberg, World of Biodiversity	No	No	No	1886	DB/Website	Wholistic	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	NA

58	Europe	Netherlands	Netherlands Entomological Society	No	No	No	NA	Website	Wholistic	Text	Browsing	Online	Yes	Yes	Research	Yes	Images
59	Asia		Hein Bijlmakers (personal website)	No	No	No	1995	Website	Basic	Text	less	Online	No	No	Basic knowledge	No	NA
61	Asia	China	Overseas Chinese Entomologists Association	No	No	Yes	1998	Website	NA	NA	NA	NA	NA	NA	NA	NA	NA
62	Asia	Japan	Asian Insect Information Center (Beetle Collection) - COLSasaji DB	No	No	Yes	NA	DB	Specific - Beetle	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	NA

63	Asia	Thailand	Thai Bugs	No	No	Yes	NA	Website	Partly Wholistic	Text	less	Online	No	Yes	Research	Yes	Images, Videos
64	Asia	Japan	Japanese Ant Image Database	No	No	Yes	NA	Website	Specific - Ant	Structural	good	Online	No	Yes	Research	Yes	NA
65	Australia	Australia	Insect Reference Collection Database	No	No	Yes	2000	DB/Website	Wholistic	Structural	Categorized Search	Online	Yes	No	Research	Yes	Images
66	Australia	Australia	Australian National Insect Collection Database	No	No	Yes	NA	DB	Wholistic	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	Images
67	Australia	Newzealand	BUGZ - Bibliography of New Zealand Terrestrial Invertebrates Online	No	No	Yes	1775	DB/Website	Wholistic - only Research Articles	Structural	Categorized Search	Online	Yes	Yes	Research	Yes	NA

Table 2. Details of Structured Databases with category search.

S. No.	Concern	Common name/ Scientific Name	Species/ sub species	Genus	Family/ sub family	Order/ Sub order	Specimen No.	Location	Author	Type	Other Details	Images	Year Collected	List of Species/ Check list	No.of Species
1	Entomological Society of America	Yes	Yes	Yes	Yes	Yes	No	No	Yes	NA	Common name/ Scientific name, Notes	No	NA	NA	NA
2	Smithsonian National Museum	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Class, Collection, Type Status, Physical description type	Yes	Yes	Yes	3 million
3	California Academy of Research	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Category, Sex, type Status, Life stage,	Yes	Yes	NA	NA

4	University of Arizona (Butterfly)	Yes	Yes	Yes	Yes	Yes	No	No	No	NA	NA	Yes	NA	Yes	Plenty - Number NA
5	UCR Entomology Research	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	Plant Association	Yes	NA	NA	NA
6	Museum of Comparative Zoology (MCZ Type DB)	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Common name	No	NA	NA	NA
7	Bishop Museum Entomological Types DB	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Mode of origin, distribution, endangered species status, publication volume, ecological data	No	Yes	Yes	16,000
8	Oxford University Museum	No	No	No	Yes	Yes	No	Yes	Yes	NA	Collection, Description	No	Yes	Yes	25,000 - no.of types

9	Essig Museum of Entomology Collections	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	Advanced Search, associated names, Type status, gender, Summary	Yes	Yes	Yes	NA
10	Natural History Museum	Yes	No	No	No	No	Yes	Yes	No	Yes	Stage, Type Status, preservation, sex	Yes	NA	NA	1 million species, 279,225 types
11	Museum of Natural History, Colorado	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Type	Yes	No	Yes	NA
12	Entomology Database, Konchur	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Flying season, Flower/plant association, parasite, references, synonymy, distribution, keywords (taxonomy, morphology, ecology, biology, etc.) , notes, type depository, host records, etc	Yes	Yes	Yes	NA

13	Entomological Society of Canada	Yes	Yes	Yes	No	Yes	No	No	No	NA	NA	No	NA	NA	NA
15	Royal Entomological Society	Yes	No	No	No	Yes	No	No	No	NA	Class, Sub class	No	NA	Yes	NA
16	Biodiversity information System	Yes	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA	No	NA	Yes	828,000
17	University of Delaware	Yes	No	No	No	Yes	No	No	No	NA	Details	No	NA	NA	NA
19	Mandala DB	Yes	Yes	No	No	No	Yes	Yes	Yes	NA	More other details	Yes	NA	NA	NA

20	Insect Reference Collection DB	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Status	Yes	NA	NA	NA
21	West Fly	Yes	Yes	Yes	Yes	Yes	No	No	No	NA	Details	No	NA	NA	NA
22	Hosts	Yes	Yes	Yes	Yes	No	Yes	No	No	NA	NA	Yes	NA	NA	22,000 Lepideptera species
23	Australian National Insect Collection DB	Yes	Yes	No	No	No	Yes	Yes	Yes	NA	Other details	Yes	NA	NA	NA
24	Online California Beetle DB	Yes	Yes	Yes	Yes	No	No	Yes	No	NA	Other details	Yes	NA	NA	50,000 records of beetle

25	Drexel University	Yes	Yes	Yes	Yes	No	No	Yes	Yes	NA	NA	No	NA	NA	NA
26	Asian Insect Information centre - COLSasaji DB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	NA	No	Yes	NA	NA
27	American museum of Natural History (Ant base)	Yes	Yes	Yes	Yes	Yes	No	No	Yes	NA	NA	No	Yes	NA	NA
28	Blattodea Species file, Orthoptera Species File, Illinois Natural History Survey	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	NA	Data source, Expert, Publication, Scrutiny, Specimen (count, sex, depository, etc.), Specimen depository, Statistics	Yes	NA	Yes	NA
29	Dermestidae of the World	Yes	Yes	Yes	Yes	No	No	No	Yes	NA	NA	Yes	NA	NA	NA

30	University of California, Irvine	Yes	Yes	Yes	No	No	No	No	No	NA	Plant Association, Flight period	Yes	NA	Yes	NA
31	University of Alberta	Yes	Yes	Yes	No	No	Yes	Yes	Yes	NA	Other details	Yes	NA	NA	NA
32	SeSam Database	Yes	Yes	Yes	No	No	Yes	Yes	Yes	NA	NA	No	NA	Yes	NA
33	American museum of Natural History	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	Other details	Yes	NA	NA	NA

Table 3. Suggested Parameter Set.

S. No.	Collection Attributes
1.	Global unique identifier
2.	Biological Systematic Name
3.	Taxonomy (Order/Suborder, Family/Subfamily, Genus, Species/Subspecies)
4.	Biological Synonymous Name/ Common Name/ Associated Names
5.	Location
6.	Author
7.	Date/Year Collected
8.	Endangered species status
9.	Mode of origin
10.	Morphology/ Physical description type – with several attributes like legs, body parts, antennae, pair of wings, color, shapes and sizes, structure, pattern, mode of locomotion etc
11.	Distribution
12.	Habitat, food habits and mode of feeding
13.	Climatic Conditions
14.	Sex
15.	Type
16.	Type Status
17.	Life Stages
18.	Special Attributes
19.	Flying season
20.	Ecology
21.	Host records
22.	Image
23.	Beneficial Insects
24.	Flower/Plant Association
25.	Harmful Insects
26.	Agricultural/ Medical/ Veterinary importance
27.	Parasite
28.	Disease Description: Disease transmission, transmission period, transmission condition, Affected organisms, disease severity
29.	Citation Reference Code (if any)
30.	Publications
31.	Summary/ Description

Therefore, constructing well-defined and specialized entomological databases is extremely important, because among the all living organisms nearly 90% of them belong to insects. Although majority of insects are beneficial to human kind, some of them are potent vectors like mosquitoes, tsetse flies, Black flies, Sand flies, bedbugs, fleas and louse causing enormous mortality and morbidity on the world population. Therefore, it is essential to adopt effective vector-control interventions by accessing the entomological databases for particular vicinity and other concerns. Furthermore, it shall be helpful as it is easier as well as economical too for employing successful pest control programme, in terms of agricultural concerns.

Specific Databases: Species-specific databases are available for some species; mainly those that are often used in research (For example, consider Online California Beetle Database). Nearly 18% of entomological databases are observed to be species-specific, which shows that not all species are covered as such. Observing the data, it is obvious that, though species-specific databases are structured, they do not contain all information as the parameters listed in Table 3. So it is suggested that these databases should not only pertain to data integration standards, but also should be maintained with current research-relevant updates (Rhee and Crosby, 2005).

Database Management System (DBMS) Design: Database design is the process of producing a detailed data model of a database and it stands for the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the database management system (DBMS) (Gehani, 2006). Here we shall discuss about the data (insects) that is to be designed and structured and the data standards applied and the mode of data retrieval in detail. In order to structure insect data as a standard, we must know what characteristics and attributes are to be stored, to define the insect identified. In this aspect the following section remains significant and pertains.

Data Standard: Data standards are documented agreements on representations, formats, and definitions of common data. The use of common data standards among databases will foster consistently defined and formatted data elements and sets of data values, and provide public access to more meaningful data. The purpose of the standard is to ensure uniformity and comparability in the identification of biological organisms in the collection, analysis, and exchange of environmental data. Maintaining these standards shall also help to have improved data quality, increased data compatibility, improved consistency and efficiency of data collection, reduced data redundancy and improved data access. The entomological databases observed in the study, have different ways of representing the details and it is obvious that data management differs for databases and some lack data standards, which challenges data integration. Since the key components of a data standard are data element names, definitions, and formatting rules (EPA, 2007), we tend to suggest the entomological databases to be designed based on a common standard or at least covering the parameters and specifications listed in the Table3.

From table 2, we find that nearly 46% of databases have similar taxonomic structure for databases, whereas the rest have different parameters, in describing the data. Nearly 36% of databases vary in the way of representing data. This shows the diversity in the representation of data, which is a check point for integrating databases or to have a unified view of data. Since the complexity, heterogeneity, and the size of biological data also raise difficult issues in the area of data models, in order to create flexible and complex access to biological databases, new data models that are sensitive to the novel characteristics of biological data and queries are required (Singh, 2003).

Apart from these, for the purpose of data integration or unified data view, it is mandatory to have similar approach on data modeling too. Though relational data models and hierarchical data models are preferred for some of the databases, object databases do better with complex data types, specifically for the data that are not repetitive and have difficulty in being described using tables and any arbitrarily data type can be stored in the object database, eliminating the need for files (Image files and PDF files) altogether. When it comes to integration at the enterprise level, object databases provide significant advantages. The main advantage of object design as an approach to software development is the ability to scale-up to very large applications. The object database at the enterprise level contains thousands of classes and millions of objects. “Integration is achieved by the interconnections between objects, crossing subject domains and leading to interdisciplinary associations” (Beck, 2001). Analyzing the structure of several entomological databases present in the study and by referring to several research articles and policies by standards bodies (like United States’ EPA data standards), here we suggest a possible parameter or attribute set, for structuring a database in Table 3, in order to provide a better data standard for all entomological databases and hence to help in integrating these databases.

Data Retrieval: Data retrieval stands for extracting required information from the database by querying. There are two types of querying that are usually allowed. One is that the user can specify the query for the data to be retrieved. The second type of querying is an automatic query builder that reduces the burden of the user (Kappara *et al*., 2011). Mostly web-based biological databases offer some type of web-based forms which act as query builders, to allow users to author DB queries. These query forms are quite restricted in the complexity of DB queries that they can formulate. Writing precise queries against biological DBs is usually left to a programmer skillful enough in complex DB query languages like SQL. However, it is mandatory to provide facilities to extract information by direct querying for sophisticated users and to construct a web interface for building precise queries for biological DBs that can construct much more precise queries and that is user friendly enough to be used by biologists. The interface needs to support queries containing multiple conditions, and connecting multiple object types, without using the join concept, which is unintuitive to biologists (Latendresse and Karp, 2010). Among the entomological databases observed, no database is observed to have an efficient way of adaptive querying in order to provide the biologists with ease of usage, like providing tools to modify or optimize queries based on previous search etc. Therefore providing intelligent interfaces for structured entomological databases shall help to obtain fast, effective and optimized results.

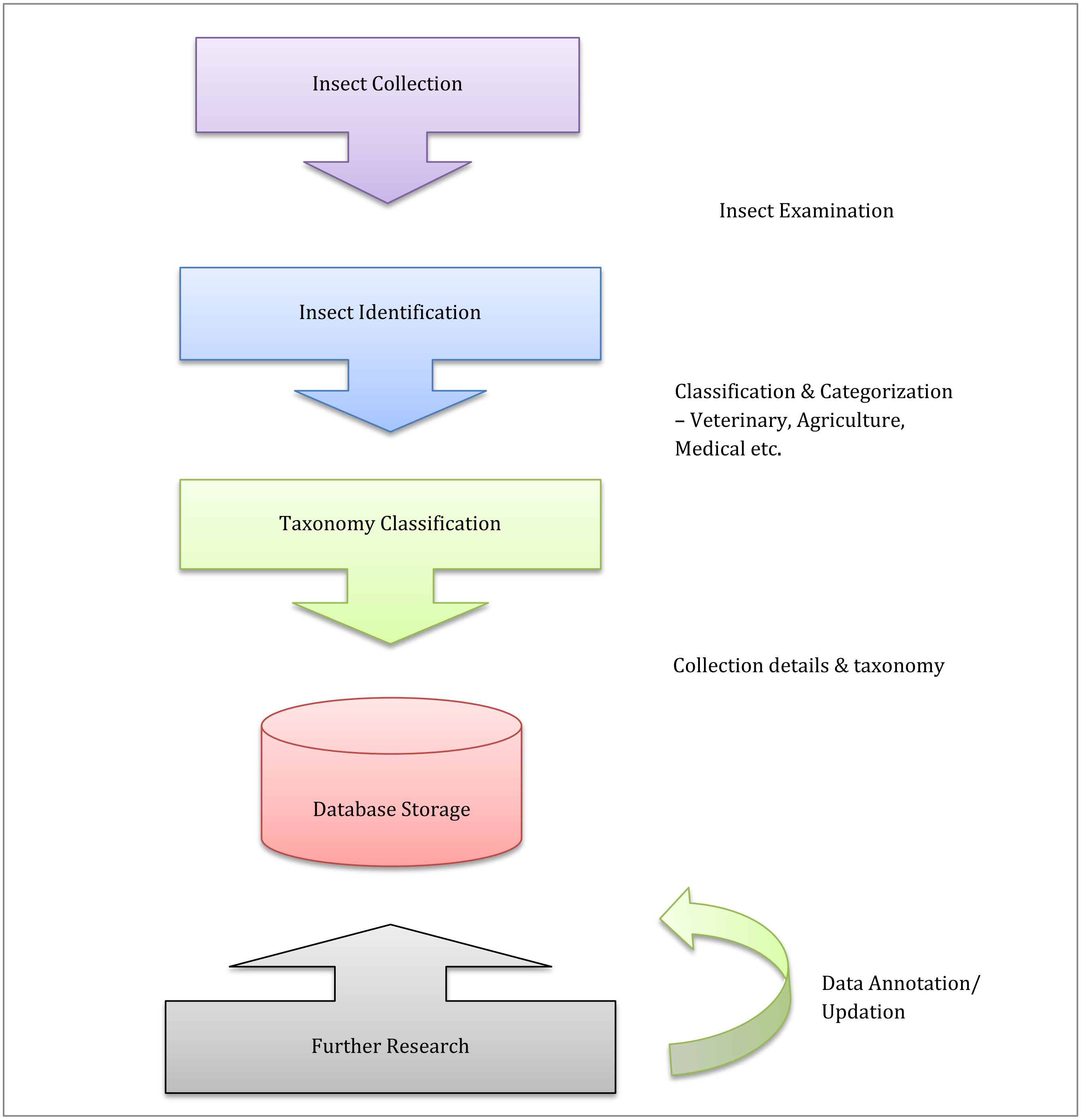

Figure 5. Process of Entomological Data Collection, Analysis and Storage.

Challenges and Opportunities in Entomological Databases: Several challenges with the entomological databases are relative to any other biological databases, in terms of collection of insects, design of database, retrieval of information, complex querying, extension of databases, integration of databases, in providing annotations and versioning facilities, and in authentication. Based on the present study results, this is evident. Similar to other biological databases, sustainability issue is observed with many websites that become inaccessible or relocated to other websites (Galperin, 2006; Wren and Bateman, 2008). It has been commented that such issue is common in many project-funded databases, specified for project data during the funding period (Rhee and Crosby, 2005). Other limitations like lack of frequent update (Wren and Bateman, 2008) and non-disclosure of year of creation are also noticed. Maintenance of digital data in long term basis in the databases has also been associated with many challenges including evolution of hardware and software, and the risk of systems becoming obsolete. Since maintenance in long-term basis is costly, and possible only in major institutions and government agencies, aggregation of information stored in different databases has been the expected solution or option (Canhos et al., 2004; WHO, 1993; Hussey et al., 2006). It has also been observed among the databases that each database has its own way of uniquely identifying the data stored, by providing accession numbers or serial numbers. However it is always significant to have global digital identifiers in order to uniquely identify each data. Certain check points must be viewed in the sole light of entomological prospects, in addition to the afore-said issues. Any entomological database should have at least the parameters suggested in Table 3. Unless otherwise the databases will not contain all the basic data required for further research and presently the available entomological databases do not have all such data. Moreover, not only data must be stored, it must also be available and accessible for usage. The information centers or databases are geographically diverse in different locations, challenging to reach the remote areas due to inaccessibility and unavailability due to restriction with authentication checks and resource constraints respectively. In addition, the existing entomological databases discourage discovery-minded users from asking data-intensive, user-specific and complex biological questions. Fewer provisions have been provided to compare and comprehend, and to make massive retrieval from several relevant databases. For instance, it is quite necessary to connect and compare with the insect details, disease transmission strategies and the nature of the parasites involved with, by referring to data from entomological databases, disease/parasitological databases and other relevant biological databases respectively. In these such cases, it is quite essential to integrate data among entomological databases. Expecting for integration of databases for collective retrieval, cannot be the only suggestion. Data integration is a challenging task, since the sources contain data from a single lab or project, or from definite repositories for very specific types of information (Example, Beetle database) and due to sheer volume of data with different data formats and data access methods. However, this could be overcome by providing a common data model for transforming data from all sources (Lacroix and Critchlow, 2003), just as suggested in Table 3. It is also quite ever possible to share and query data among databases, by making international collaborations between organizational databases, similar to that of the International Collaboration of DDBJ, ENA, GenBank databases for nucleotide sequences (Cochrane et al., 2012). Despite the challenges with respect to massive retrieval and data integration, it is still possible to retrieve collective data and to make further analysis, by introducing a meta-search system that connects all entomological databases and relevant biological databases automatically, even if they have slight differences in their data formats and storage methods. The following figure 6 shall demonstrate a framework for a meta-search interface for entomological databases. The meta-search interface, receiving input from the user shall validate and parse the input string or query. The parsed query shall be identified with the pattern and checked if such retrieval has been made recently. If a recent retrieval is identified, the same result set is brought out. If else, the parsed query is formatted, processed and re-defined accordingly, and sent to relevant databases, data sources and search engines respectively to get the comprehensive result. The result details are also logged. For implementing an effective interface as demonstrated above, retrieval from ontology specific databases shall be of additional helpful, since the semantics are clearly identified and the retrievals can be more meaningful and relevant.

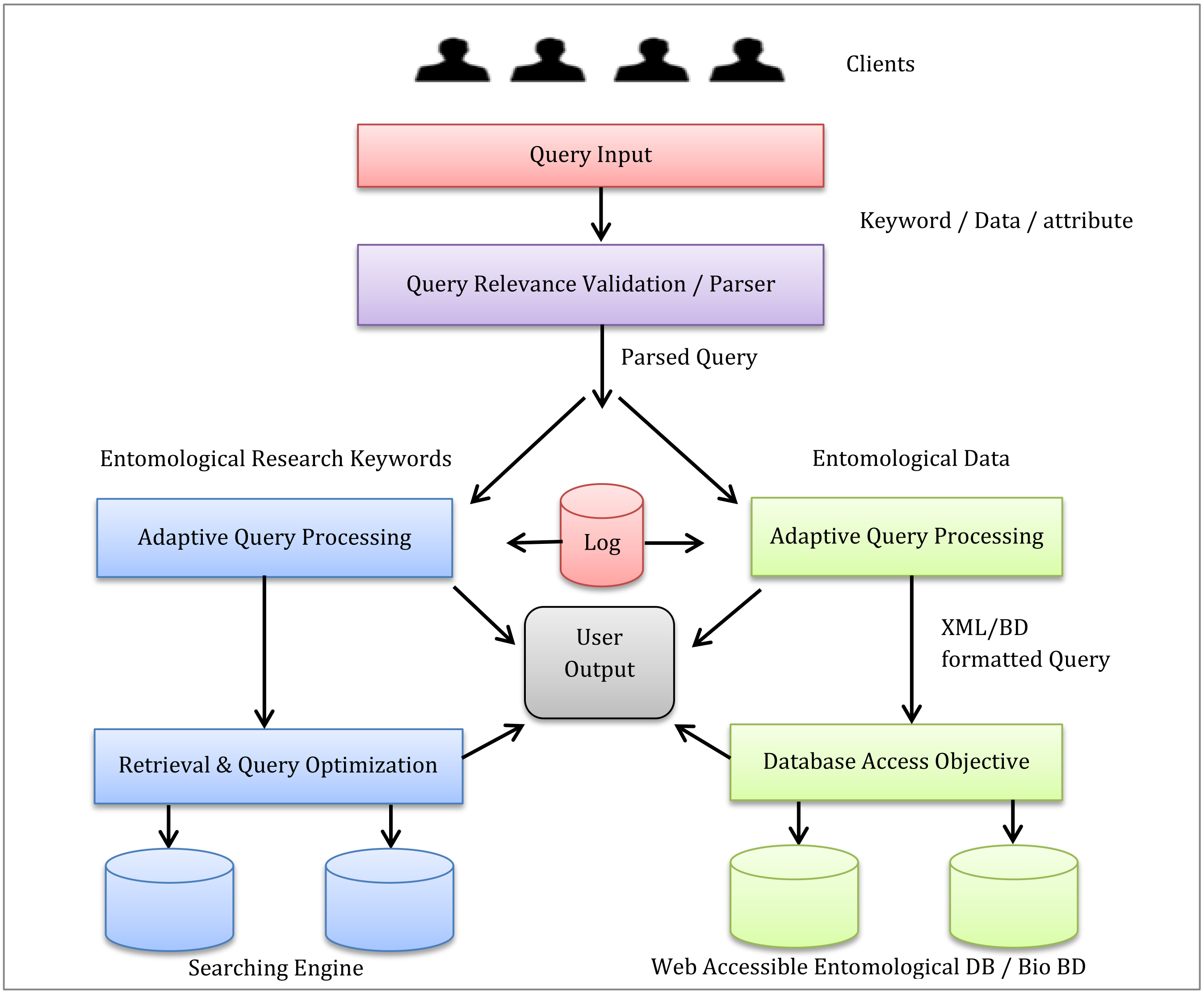

Figure 6. A Meta-search Design for retrieving from Entomological Databases.

CONCLUSION

Though the issues described with the entomological databases, are similar to any other database, the necessity for entomological databases stands unique as it associates with all living beings existence. Documenting entomological details will help to obtain knowledge and hence to facilitate the fight against diseases and crop loss, assist in the development of drugs and preventive measures, and in discovering the basic relationships between species. However, it is quite essential to take measures, to repair with the following issues:

- Funding agencies to support and encourage in establishing research-oriented databases even in the resource-constrained regions, to enable further research and analysis and documentation of entomological data and in constructing a common platform to connect several databases.
- Integrate existing databases or to provide intelligent interfaces with unified view of data from all databases.
- Identify and acknowledge common framework or data model for representing data.
- Make policy decisions in collaborating existing databases and to share data beyond boundaries.
- Make efforts to effectively integrate entomological databases with other types of databases, to help in identifying disease patterns, drug design etc.

REFERENCES

Altman, R. B. 2004. Building successful biological databases. Briefings in Bioinformatics, 5(1):4-5. DOI:10.1093/bib/5.1.4.

Attwood, T. K., A. Gisel, A. N. E., Eriksson and E. Bongcam-Rudloff. 2011. Concepts, Historical Milestones and the Central Place of Bioinformatics in Modern Biology: A European Perspective. Bioinformatics - Trends and Methodologies. InTech, Croatia. Available at: http://www.intechopen.com/articles/show/title/concepts-historical-milestones-and-the-central-place-of-bioinformatics-in-modern-biology-a-european-, Retrieved 8 Nov 2013.

Beck, H. 2001. Agricultural enterprise information management using object databases, Java, and CORBA. Computers and Electronics in Agriculture, 32:119-147.

Biosystematics Division. 2000. Collecting and Preserving Insects and Arachnids - A Manual for Entomology and Arachnology. SAFRINET, the Southern-African (SADC) LOOP of Bio-net International.

Canhos, V. P., S. Souza, R. Giovanni, and D. A. L. Canhos. 2004. Global biodiversity informatics: setting the scene for a “new world” of ecological modeling. Biodiversity Informatics, 1:1-13.

Chapman, A. D. 2009. Numbers of living species in Australia and the World Report for the Australian Biological Resources Study. Canberra, Australia. pp. 60. Available at:
http://www.deh.gov.au/biodiversity/abrs/publications/other/species-numbers/index.html.

Cochrane, G., I. Karsch-Mizrachi, and Y. Nakamura. 2012. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research, 40(D1):D33-D37.

Edwards, A. W. F. and L. L. Cavalli-Sforza 1964. Reconstruction of evolutionary trees. In Heywood, Vernon Hilton; McNeill. Journal of Phenetic and Phylogenetic Classification, 67-76. OCLC 733025912.

Environmental Protection Agency's (EPA) 2007. Data standards program - Data Standards Briefing Paper. Available at: http://www.epa.gov/ttnchie1/conference/ei12/panel/kohn.pdf. Retrieved 12th December 2013.

Erwin, T. L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. Coleopterists Bulletin, 36(1):74-75.

Erwin, T. L. 1997. Biodiversity at its utmost: Tropical Forest Beetles. In: Reaka-Kudla, M.L., D.E. Wilson and E.O. Wilson (Eds.). Biodiversity II. Joseph Henry Press, Washington, D.C, pp. 27-40.

Galperin, M. 2006. The molecular biology database collection: 2006 update. Nucleic Acids Research, 34:D3-D5.

Gehani, N. 2006. The Database Book: Principles and practice using MySQL. 1st Ed., Summit, New Jersy.: Silicon Press, USA.

Hunsucker, G. M. 1996. European Database Directive: Regional Stepping Stone to an International Model. Fordham Intellectual Property, Media and Entertainment Law Journal. pp.697, 704.

Hussey, C., S. Wilkinson, and J. Tweddle. 2006. Delivering a name-server for biodiversity information. Data Science Journal, 5:18-28.

Jacobs-Lorena, M. 2006. Genetic approaches for malaria control. In: Knols, B. G. J., and C. Louis (Eds). Bridging laboratory and field research for genetic control of disease vectors. Springer Part 2:53-65.DOI: 10.1007/1-4020-3799-6_5,2006.

Kappara, V. N. P., R. Ichise, and O. P. Vyas., 2011. LiDDM: A Data Mining System for Linked Data. LDOW2011. March 29, 2011. Hyderabad, India.

Lacroix, Z., and T. Critchlow. 2003. Bioinformatics: Managing Scientific Data. The Morgan Kaufmann Series in Multimedia Information and Systems, Morgan Kaufmann Publishers, Massachusetts (San Francisco, California), USA.

Lane, F., 2007. IDC (International Data Corporation): World Created 161 Billion Gigs of Data in 2006. Available at: http://www.toptechnews.com/story.xhtml?story_id=01300000E3D0. Retrieved 12 December 2013.

Latendresse, M. and P. D. Karp. 2010. An Advanced Web Query Interface for Biological Databases. Database - The Journal of Biological Databases and Curation, baq006. doi: 10.1093/database/baq006.

Lenzerini, M., 2002. Data Integration: A Theoretical Perspective. Dipartimento di Informatica e Sistemistica, Universita di Roma “La Sapienza” Via Salaria 113, I-00198 Roma, Italy. Available at:
http://www.cs.ubc.ca/~rap/teaching/534a/readings/Lenzerini-pods02.pdf. Retrieved 12 December 2013.

Margaret, A. J. 1997. Carl Linnaeus: Father of Classification. United States: Enslow Publishers.

Ningthoujam, S. S., A. D. Talukdar, K. S. Potsangbam and M.D. Choudhury 2012. Challenges in developing medicinal plant databases for sharing ethnopharmacological knowledge. Journal of Ethnopharmacology, 141(1):9-32.

Novotny, V., Y. Basset, S. E. Miller, G. D. Weiblen, B. Bremer, L. Cizek, and P. Drozd 2002. Low host specificity of herbivorous insects in a tropical forest. Nature, 416(6883):841-844.

Rhee, S. Y., B. Crosby 2005. Biological Databases for Plant Research. Plant Physiology, 138:1-3.

Singh, A. K. 2003. Querying and Mining Biological Databases. Journal Integrative Biology, 7(1):7-8. doi:10.1089/153623103322006472.

Tripepi, R. R., M. Bauer, S. M. Bell, and W. B. Jones 2013. The Idaho Master Gardener Program Handbook. Educational Publications Warehouse, Master Gardener Idaho University of Idaho Extension. Available at: www.extension.uidaho.edu/mg/resources/handbook/mghbook.pdf

WHO, 1993. Research Guidelines for Evaluating the Safety and Efficacy of Herbal Medicines. World Health Organization, Regional Office for the Western Pacific, Manila, Phillipines.

Wilson, E. O. 2006. Threats to Global Biodiversity. Available at: http://www.globalchange.umich.edu/globalchange2/current/lectures/biodiversity/biodiversity.html. Retrieved 8 Nov 2013.

Woldetensae, Y 2007. Optimizing the African Brain Drain: Strategies for Mobilizing the Intellectual Diaspora towards Brain-Gain. Association of African Universities, Libya, October 21-25.

Wren, J. D. and A. Bateman 2008. Databases, data tombs and dust in the wind. Bioinformatics, 24(19):2127-2128. DOI:10.1093/bioinformatics/btn464.