EXPERIENCE AND STRATEGY OF BIODIVERSITY DATA INTEGRATION IN TAIWAN

K T Shao*, K C Lai, Y C Lin, L S Chen, H Y Li, C H Hsu, H Lee, H W Hsu, and G S Mai

Biodiversity Research Center, Academia Sinica, 11529 Taipei, Taiwan
**Email:*zoskt@gate.sinica.edu.tw

ABSTRACT

The integration of Taiwan's biodiversity databases started in 2001, the same year that Taiwan joined GBIF as an associate participant. Taiwan, hence, embarked on a decade of integrating biodiversity data. Under the support of NSC and COA, the database and websites of TaiBIF, TaiBNET (TaiCOL), TaiBOL, and TaiEOL have been established separately and collaborate with the GBIF, COL, BOL, and EOL respectively. A cross-agency committee was thus established in Academia Sinica in 2008 to formulate policies on data collection and integration as well as the mechanism to make data available to the public. Any commissioned project will hereafter be asked to include these policy requirements in its contract. So far, TaiBIF has gained recognition in Taiwan and abroad for its efforts over the past several years. It can provide its experience and insights for others to reference or replicate.

Keywords: Biodiversity informatics, GBIF, TaiBNET, TaiBIF, Database

1   2001: THE YEAR DATA INTEGRATION BEGAN

Before 2001, the biodiversity databases in Taiwan were scattered in various government agencies, private organizations, and academic institutions. There was no real horizontal integration; these databases, at most, provided links to other sites or the home pages of relevant databases on their websites. These agencies and institutions may have departments or research units under them, each in turn may have its own websites and databases. For example, under the Council of Agriculture (COA), there are the Fisheries Agency, Forestry Bureau, and Taiwan Endemic Species Research Institute; under the Construction and Planning Agency, there are many national parks. As for the biodiversity-related private organizations, more than 30 of them have established databases and websites. The large-scale or integrated research projects promoted by the government also have their own websites, such as the Forestry Bureau's National Survey and Mapping of Floral Diversity Project, the Bureau of Animal and Plant Health Inspection and Quarantine's invasive species project, the Council for Economic Planning and Development's National Geographic Information Systems, and the National Science Council's Long-Term Ecological Research Network. However, these sites usually cover project introductions, research reports, literature, news articles, and policy and regulation guidance but lack metadata, raw data, or primary data from research projects. Moreover, the reports are often in paper form and are kept at the funding agencies, making it difficult to achieve the goal of sharing and integrating research data across agencies.

Taiwan began to integrate its biodiversity data in 2001. The new National Digital Archives Program aimed to archive not only data in the field of humanities and social sciences but also data in biological and natural sciences, such as specimens and species information. The Executive Yuan approved the Biodiversity Promotion Plan in the same year. One of the projects under the Promotion Plan is for the National Science Council (NSC), leading nine co-organizers, to collect and integrate biodiversity data and exchange them with global organizations. The data collected cover expert lists, species checklists, specimen information, geographical distribution, spatial and temporal distribution, invasive species, species description, literature, and biological resources. Also in 2001, the Global Biodiversity Information Facility (GBIF) was formally established, and Taiwan joined it as an associate participant. As a result, Taiwan can apply the technologies and standards of GBIF's metadata and exchange platform to promote the integration of its biodiversity information and the exchange with GBIF.

In 2002, NSC began to provide funding for the Biodiversity Research Center, Academia Sinica to create the website of the Taiwan Species Checklist "TaiBNET" (http://taibnet.sinica.edu.tw). More than a hundred taxonomists are involved in this work. In 2004, GBIF's Taiwan Portal, TaiBIF (Taiwan Biodiversity Information

Facility), was built. Its tasks are to consolidate Taiwan's biodiversity data, develop software tools, and hold educational training workshops in related metadata standards and technologies. In 2007, the second phase (2007-2012) of the National Digital Archives Program was transformed to Taiwan e-Learning and Digital Archives Program (TELDAP), and an International Collaboration and Promotion Division was added. One of the tasks of this division was to promote international cooperation in biodiversity information. TaiBIF was then incorporated into TELDAP, and its tasks were re-defined (Shao et al., 2008a). TaiBIF is to collect the ecological distribution data and data in *Biota Taiwanica* (in English), which is commissioned by the NSC. It is also in charge of the integration of cross-agency and institution biodiversity data. Further, TELDAP is to integrate data from its Institutional Projects, Request-for-Proposals Project, and Biosphere and Nature Thematic Group. It concentrates on the collection of species descriptions (in Chinese), specimens, and literature information. Its task of collaboration with international organizations is assisted by TaiBIF.

As to the issue of intellectual property rights (IPR), TELDAP originally employed the "Creative Commons (CC)" licenses approach. Under a combination of Attribution, Non-Commercial, No Derivative Works, and Share-Alike requirements, data are to be put on the Internet to be browsed by the public. However, for a great deal of cultural content, the IPR issues were not clarified. In addition, the nature of some information makes it unsuitable for CC. For example, the IPR of Taiwanese aborigines is covered by separate laws. Therefore, TELDAP in 2008 had to conduct an inventory of IPR for every item and require project directors to re-sign license agreements, solving the problem of TaiBIF not being able to receive data from the Union Catalog and then submitting them to GBIF.

2 INTEGRATE BIODIVERSITY DATA IN TAIWAN AND EXCHANGE THEM INTERNATIONALLY

Currently TaiBIF/TELDAP integrates cross-agency, cross-institution, cross-project, NGO, and private biodiversity databases through the use of species name and GPS (latitude and longitude). The most critical element is to build a species checklist as the scientific name is the keyword or link to the integration of biodiversity data. Via the scientific name of a species, its specimen information, DNA barcode (Barcode of Life, BOL), phylogeny (Tree of Life, TOL), ecological distributional data (in EML format), and other information (Encyclopedia of Life, EOL) can all be accessed.

There are three different ways to exchange and share Taiwan biodiversity data internationally (Shao, 2006; Shao et al., 2007b). The first path is to link to GBIF directly through TaiBIF. The second path is to go through regional networks that are the GBIF's Associate Participants themselves. The third path is for the local organism databases to send Taiwanese native or endemic species information to global species databases (GSDs), such as FishBase and AntBase. They in turn can link to GBIF through large-scale international cooperation projects, such as COL, BOL, EOL, BHL, OBIS, BioNET-International, etc. (Fig. 1).
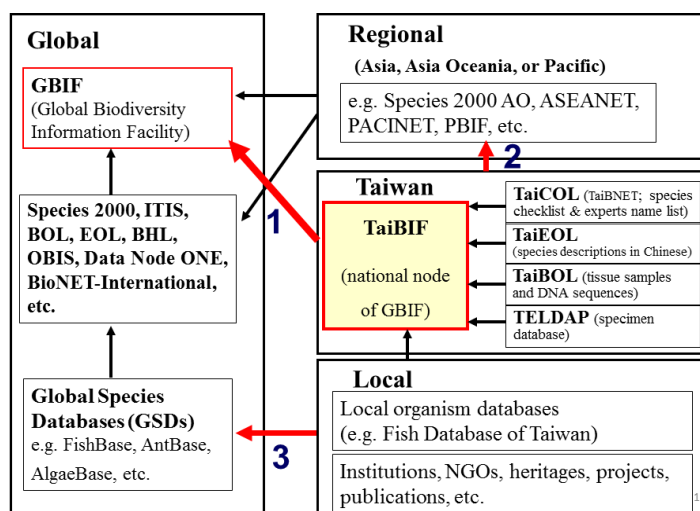

Figure 1. Integrate biodiversity databases in Taiwan and link them globally

In order to more effectively and broadly employ the integration framework and advancement of global biodiversity information, Taiwan will (1) cooperate with the Catalogue of Life (COL) of Species 2000 to continue maintaining and operating Taiwan's COL, i.e., the database and website of TaiCOL (renamed from TaiBNET); (2) cooperate with CBOL and iBOL to continue maintaining and operating Taiwan's BOL and change the name of the present "Cryobanking Program for Wildlife Genetic Material in Taiwan"(http://cryobank.sinica.edu.tw) to TaiBOL; (3) cooperate with EOL and build Taiwan's EOL, TaiEOL.

All the databases mentioned above will be integrated into TaiBIF (http://taibif.org.tw), which is linked to global databases such as GBIF. The old URLs will remain accessible. The new URLs and their website contents are listed below:
(1) TaiCOL (http://col.taibif.tw) = TaiBNET: 54,000 native species, 1,351 alien species, along with conservation species and fossil species.
(2) TaiBOL (http://bol.taibif.tw): 10,457 tissue samples from 2,981 species, 1,255 DNA sequences from 844 animal species.
(3) TaiEOL (http://eol.taibif.tw): established in 2011 and will have information in Chinese on 16,000 species by the end of 2013.
(4) TaiBIF (http://taibif.tw): In addition to the information on the three websites mentioned above, there is *Biota Taiwanica* with English information on 11,000 species and videos, literatures, specimens, etc. Other than database integration work, TaiBIF provides biodiversity communities in Taiwan with information technology assistance to accelerate the integration and sharing of data.

3   STARTING WITH A SPECIES CHECKLIST—TaiCOL (TaiBNET)

The scientific name of a species is the most important keyword in linking biological information; hence, the most basic task of integrating biodiversity data is to first establish an accurate, authoritative, and complete species checklist. TaiBNET was established in 2002 to integrate and update the checklist of valid species in Taiwan. So far a total of 54,417 native species in eight kingdoms (Virus, Bacteria, Archaea, Protozoa, Chromista, Fungi, Plantae, Animalia), 59 phyla, 143 classes, 662 orders, 3,128 families, and 17,706 genera have been compiled, including more than a thousand cultivars, alien species, and fossil species. TaiBNET offers users handy search functions such as search by partial scientific name, Chinese name, common name, keyword, or string. Users can also browse by the taxonomy tree. Each of the tree's classification levels displays the numbers of recorded organisms in its sub-levels. When a species is chosen, a description page with its classification, synonyms, and literature is shown. For some organisms, links to other databases such as the Fish Database of Taiwan, FishBase, Discovery Life, and EOL are given. A mechanism is provided to let users submit related information and images to the site. A "2008 Workshop: Research and Status of Taiwan Species Diversity" was held at the National Museum of Natural Science in Taichung on August 15-16, 2008. Half a year later, the books "2008 Taiwan Species Diversity I. Research and Status" and "2008 Taiwan Species Diversity II. Species Checklist" were published along with a DVD. The book and DVD of "Taiwan Species Checklist 2010" were published in 2010 and can be freely downloaded as well (Shao et al., 2008b, 2008c, 2010).

The database TaiBNET is currently maintained by a full-time assistant. Authorized taxonomists and their doctoral students can update data online. Afterward, the data are reviewed by top ranking scholars in each taxon. Another way to check the validity of data is to compare the checklist with existing databases such as Sp2000 or the electronic version of *Biota Taiwanica*. When discrepancies are found, experts in that taxon are invited to double check. Additionally, a fungus can be matched against CABI's Index of Fungi and marine organisms can be matched against WoRMS.

One advantage of international cooperation is to obtain a wealth of other information. For example, Taiwan has 3,086 valid fish species as of 2011; from FishBase, 14,000 synonyms of its native fishes were obtained. The collection of synonyms is important because scientific names and classification systems are constantly being revised. For fish, 1/10 of fish names are changed approximately every 10 years (Pauly & Froese, 2000; Shao et al., 2007a). Even taxonomists will not be able to track and remember the changes and can only rely on automatic database comparison to detect them. When a user enters an invalid name of a species, the system automatically finds the valid name and links to the right webpages. Otherwise, much of the valuable specimen information in herbaria and museums would be lost if users only know synonyms of a species name.

4 ESTABLISHING THE TAIWAN ENCYCLOPEDIA OF LIFE WITH COMMUNITY ENGAGEMENT

Advocated by biologist E. O. Wilson, the Encyclopedia of Life project began in 2007. Its goal is, through the joint efforts of scientists around the world, to gather and share scientific knowledge about the 1.9 million known organisms in a single online resource. The TaiEOL team started to communicate and exchange information technology with the EOL team in 2009. In 2011, the team began to work on the Taiwan Encyclopedia of Life (TaiEOL, in Chinese) and hopes to complete in three years the online information content of 20,000 species, including 8,000+ endemic species.

4.1 Status and analysis of species data in traditional Chinese characters

TaiEOL from the start has utilized its own Biodiversity Bibliography System to record species with Chinese descriptions from biological publications in Taiwan. There are now more than 400 books written in traditional Chinese and a cumulative total of 45,000 scientific names in the system. Using Insecta as an example, there are 21,000 recorded insect species in Taiwan; yet only 4,700 species have Chinese descriptions. Furthermore, only 1,600 out of 6,000 endemic insects have information written in Chinese. These statistics help with the decision on which taxon to invite researchers to fill in with information. They also provide users with the bibliographies of a species when they retrieve information on that species.

4.2 Open source software platform of TaiEOL portal

The information collaboration and user-participation concept of Web 2.0 has become an international trend in sharing and integrating biodiversity data. LifeDesks, the modules developed by the EOL project, and Scratchpads, provided by the European Distributed Institute of Taxonomy (EDIT), are two of the most popular participatory platforms. Based on these two types of platforms and using open source software, TaiEOL began to develop its Chinese version with species information as its core.

The TaiEOL portal will build on the existing species checklist of Taiwan. For the 5.4 million species, it will, in staggered phases, invite taxonomists and citizen scientists to join the effort and contribute to its content. Taiwan's endemic species especially need descriptions, images, and popular science material so that the public can easily learn about them and become knowledgeable in Taiwan's biodiversity. Moreover, TaiEOL can provide information on endemic species to EOL to be shared by the global community. Through the portal's biological content management system and user interaction mechanisms, the biodiversity information of Taiwan can be directly and effectively shared and exchanged.

5 INTEGRATING DATABASES AND LINKING GLOBALLY THROUGH TaiBIF

TaiBIF, the Taiwan node of GBIF, uses the metadata format and tools recommended by GBIF to integrate Taiwan's biodiversity information and share it with the global community. The homepage of TaiBIF is shown in Figure 2. TaiBIF consolidates various databases in Taiwan, including species checklists, *Biota Taiwanica*, species occurrence data, Taiwan Encyclopedia of Life, etc. Using biodiversity informatics and tools, TaiBIF enables the general public, academic scholars, and government agencies to gain access to the data on its platform. In addition, TaiBIF introduces international data standards and protocols so that it can meet the two objectives of "deepening Taiwan people's biodiversity knowledge" and "helping to make the world's biodiversity picture more complete."

Figure 2. Homepage of TaiBIF

5.1   Information currently available online at TaiBIF portal (Shao et al., 2009)

5.1.1   Species checklist

The Catalogue of Life in Taiwan project (TaiBNET) offers species names (scientific names and Chinese names) and classification hierarchy and keeps track of taxonomic work. Being the backbone of the project, these data provide means to access all the biodiversity information.

5.1.2   Primary species occurrence data (specimen and observational data)

There are specimen data from TELDAP and observational data from various collaborating domestic databases. A total of 1.59 million digital specimens and observational records are now available. The data come from institutions such as Academia Sinica, National Museum of Natural Science, National Taiwan University, National Taiwan Museum, Taiwan Forestry Research Institute, Fisheries Research Institute, Agricultural Research Institute, Taiwan Endemic Species Research Institute, National Museum of Marine Biology and Aquarium, National Taiwan Ocean University, National Tsing Hua University, National Sun Yat-sen University, etc.

5.1.3   *Biota Taiwanica*

NSC's Division of Life Sciences, in order to fulfill the requirement of the Biodiversity Promotion Plan, assists domestic taxonomists in composing contents for *Biota Taiwanica*. So far, information on more than 10,000 species has been uploaded online and released to the public.

5.1.4   Literature

Relevant research papers on biodiversity are provided by the Science and Technology Policy Research and Information Center of the non-profit National Applied Research Laboratories. Currently, there are research projects (8,079 articles), research reports (3,987 articles), periodical papers (10,611 articles), conference papers (4,767 articles), Master's theses and Doctoral dissertations (4,171 articles), English journal papers (2,270 articles), and publications (499 articles).

5.1.5   Eco-Photo

Using Cooliris software, TaiBIF presents the ecological photos and videos it collects. It will include the images gathered by TaiEOL in the future.

5.1.6   Species description

The text and images of native Taiwan species as well as popular science material from TaiEOL will be incorporated into TaiBIF in the future.

5.2   Web service and tool development

After years of research and development effort, the TaiBIF team has made available many biodiversity information tools to accelerate the integration and sharing of biodiversity data. For example, there are tools to check scientific names and geological coordinates:

  i.   The scientific names are matched against TaiBNET and Sp2000, and any errors in spelling are highlighted and reported back to the users for reference. Additional information on the taxon level, author, and publication year of the species is provided in the correct sequence so that users can easily standardize the format and increase the accuracy of data.

    ii.   TaiBIF converts geological coordinates to the format commonly used in Taiwan so that users can update old ecological distribution data and conduct further analysis and research work. TaiBIF also verifies the coordinates. Its tools will check spatial data for errors; e.g. based on the description of a location and latitude and longitude, the tools can determine if the coordinates fall outside the range. TaiBIF will pass the information back to the users (data providers) to make necessary changes. In doing so, TaiBIF increases the quality of the species occurrence data and accomplishes the goal of integrating and accessing biodiversity data through geographic information.

6   INFORMATION TECHNOLOGY FOLLOWING INTERNATIONAL TREND

In terms of information technology, GBIF uses DiGIR, BioCASE, and TAPIR with Darwin Core as its metadata standard to integrate global species occurrence data (Hill et al., 2009; Chapman, 2005) and has accumulated 312 million occurrence data to date. TaiBIF in 2004 started to use DiGIR (Distributed Generic Information Retrieval) as the protocol to exchange data with the GBIF Secretariat. Due to the limitations of certain fields, CRIA (Centro de Referência em Informação Ambiental) developed TAPIR (TDWG Access Protocol for Information Retrieval) to try to solve the problem. TAPIR is a REST style distributed data exchange protocol transmitted by HTTP. It combines and reinforces the strengths of BioCASE and DiGIR and allows users to communicate with data providers in a simple and direct way. It also provides more choices in data distribution standards, including different versions of Darwin Core and the capacity to define data by custom XML, making it more flexible to share data. TAPIR was officially adopted by GBIF and has been one of the official TDWG standards since October 2009.

TAPIR was chosen by the TaiBIF team to be its data exchange protocol, integrating data (in English) into the original information architecture. However, the majority of Taiwan's species occurrence data also contains Chinese characters. TaiBIF, therefore, made use of TAPIR's customization feature and created a Chinese-language XML extension. Consequently, not only the English data can be shared with GBIF but also the Chinese data can be shared within the original Darwin Core framework. Furthermore, a retrieval platform was established at TaiBIF (Fig. 3). This framework constantly assists with system deployment at the biodiversity-related institutions in Taiwan. Various institutions and museums have now accumulated over 1.59 million species occurrence data.
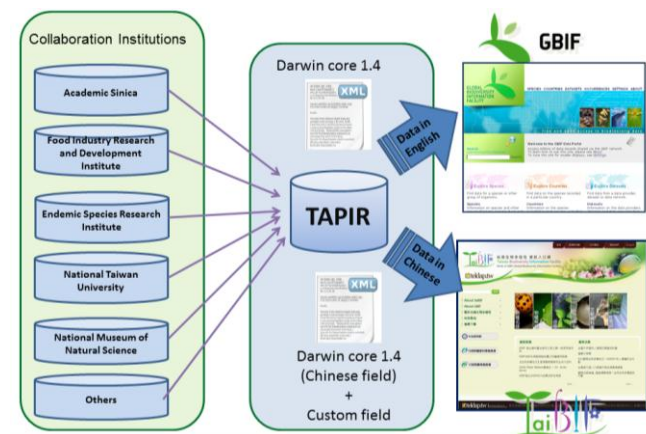

Figure 3. Flow of data to TaiBIF (Chinese) and GBIF (English) using TAPIR

7   DIFFICULTIES AND SOLUTIONS OF BIODIVERSITY DATA INTEGRATION

Compared to specimen, literature, and species checklists, the collection and integration of the observational raw data are much more difficult. This is due to the fact that once a species is identified and its distribution is recorded by taxonomists or ecologists, the information can be analyzed and written into papers if the information is posted online. However, specimens can only be borrowed with the agreement of their original collectors or managers. As a result, most of the researchers may be willing to provide analyzed charts or tables but hesitate to disclose and share data before they have a chance to publish their papers. It is fairly common, and a regrettable

happening, for research data to get lost, destroyed, or buried somewhere (Shao et al., 2007c).

The reluctance to submit raw data is only one of the reasons why, in the past 20 years, there has not been much progress in the information integration process among or across government agencies in Taiwan. Other reasons are: the complicated issues of IPR, the lack of information collection and management unit in the agencies, the absence of clear, executable policies for information, etc. Hence, entrusted by the National Science Council, Academia Sinica in 2008 set up the National Committee for GBIF (GBIF-ROC), with committee members being either representatives from biodiversity-related agencies or researchers who are in charge of databases. The Committee is establishing data exchange standards, studying and developing viable information policies, and requesting all agencies to include a term in their contracts demanding all their commissioned projects to submit the raw data when the projects end. After a certain period of time, agreed upon beforehand, the data will be open to the public.

7.1   Principles for submitting ecological distribution data

After several discussions, GBIF-ROC reached a conclusion on July 20, 2009, entitled "The principles of government-funded ecological distribution data collection and archiving, promoted by GBIF-ROC Committee" to be considered by the Sustainable Development Research Committee and the Biodiversity Promotion Plan.

1. Type and scope of the submitted data: In the first stage, only the digital raw data of the ecological distribution (species occurrence data) need to be submitted. Other related data will be added later.
2. Collection and submission of survey and monitor data: All government agencies are requested to include a term in their contracts demanding all their commissioned projects to submit raw data when the projects end.
3. Format, methodology, and evaluation of data submitted: (i) Submitted information should include survey or monitor methodology, definitions, and data. The format should use international customary standards, such as Darwin Core, Ecological Metadata Language (EML) (Jones et al., 2006), and ISO19115. (ii) Each agency should designate its data collection (storage) unit based on its technology capability. (iii) Each agency should define an evaluation system; e.g., further funding is provided only after data are submitted.
4. When the data are made public (IPR issue): Each agency should determine when to make the data accessible by the public.

7.2   Format of ecological distribution data submitted

Consensus has been reached about the format of the information that the commissioned projects are required to submit. There are 12 items of metadata and one item of raw data.

1. Title of the project;
2. Data owner information, at least including name, organization, position title, and contact information;
3. Associated parties information when applicable, at least including name, organization, position title, and contact information;
4. Research summary (abstract);
5. Keywords;
6. Usage rights;
7. Contact window information, at least including name, organization, position title, and contact information;
8. Content description, including materials and methods, temporal coverage, geographic coverage, and taxonomic coverage;
9. Variable names;
10. Variable labels;
11. Variable definitions;
12. Variable measurement definitions, including detailed definition of measurement categories. If it is nominal or ordinal, it should include description and definition of the values. If it is interval or ratio, it should include unit, precision, and number type. If it is date-time, it should include format and precision.

13. Raw data of items 9-12 mentioned above.

7.3   Future challenges and strategies

In order to achieve its goals of collecting and integrating biodiversity data, TaiBIF has officially proposed to higher-level governmental decision makers, such as the National Science and Technology Conference and the National Council for Sustainable Development, the adoption of a top-down approach. However, the database integration work in many agencies still faces some common problems and challenges that are listed below. The difficulties can only be overcome with the support and determination of the managers in each agency. The managers need to commit more manpower and material resources to the work and also bestow recognition and encouragement on their employees.

1. Without professional IT personnel to serve on the staff, the researchers are forced to outsource system development as they usually don't have the time to manage or maintain the databases on their own. When open source software is not used, it is difficult to maintain a database sustainably. The database typically ceases to exist when a project ends.
2. Database management in general is regular work. Consequently, its funding is often cut back year after year.
3. The accomplishment of the researchers is not recognized. The evaluation of a researcher's performance is based exclusively on his/her published papers (SCI scores). Hence, most researchers are reluctant to spend their time on databases or share their data, making it difficult to conduct further work on the data. Recently GBIF has started to promote the publication of "data papers" in scientific or SCI journals, a good incentive strategy for scientists to publish their data and give the public open access to them.

Other methods and strategies promoted by TaiBIF are as follows:

1. Cooperating with other websites to increase page views through mutual links.
2. Sharing budget and accomplishment. Data integrators need to establish credibility by distributing funds equitably, making their own data public, and attributing achievement to team members and data providers.
3. Meeting the needs of users. The convenience of users and data providers should take precedence over that of information technology or data integration.
4. In addition to the number of SCI (Science Citation Index) papers and their impact factors, digitized material, such as the number of records (including specimens collected, ecological distribution data, and DNA sequences) archived and uploaded to the Internet, the so-called "Repository Impact Factor," should be taken into account when assessing research performance. Data compilation and submission in the form of "data papers" to academic journals should also be encouraged. This mechanism facilitates the publishing of biodiversity data resources (Chavan et al., 2011).
5. Implementing performance evaluation effectively in order to receive data effortlessly.
6. Nurturing talents. Provide opportunities for young scholars to organize conferences or attend international meetings so that they can learn and exchange ideas.
7. Advocating the benefits of data integration:
   ✓ Offer offsite backup (data stored in different places)
   ✓ Help with data validation to improve data quality
   ✓ Contribute to the society; academic services; the taxpayers' rights
   ✓ Increase page views of the data
   ✓ Assist government in making sure it receives concrete results from the huge investment of funds it allocates to scientific researches and surveys. Raw data are archived and preserved so that when a project ends, these data, rather than summarized reports, are available to the public.
   ✓ Be able to reanalyze and re-simulate models with the data, using newer statistical software such as cluster analysis, ecological model, etc.
   ✓ Can be used as an important tool, e.g., a quantitative indicator of biodiversity for resource conservation, sustainable use, and management policies.

8   CONCLUSION

In order to meet the tasks stipulated by the Convention on Biological Diversity, it is elemental to integrate biodiversity data and make it accessible to the public. Only through sufficient information sharing can the multi-dimensional objectives of biodiversity conservation, research, education, and resource sustainability be achieved. The promotion of data integration involves establishing work flow and overcoming difficulties occurring in the process. The difficulties cover many aspects, such as intellectual property rights, data policy, data standards, data quality concept, standard operating procedure of collecting data, and software/hardware development technologies. In this paper, we present the results of the ten-year experience of the Biodiversity Research Center, Academia Sinica in promoting biodiversity data integration. Future work will consider the addition of environmental and hydrological conditions to analyze in order to increase sustainable biodiversity development and policymaking.

9   REFERENCES

Chapman, A.D. (2005) *Uses of Primary Species-Occurrence Data*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. GBIF, pp 1–117.

Hill, A.W., Guralnick, R., Flemons, P., Beaman, R., Wieczorek, J., Ranipeta A., et al. (2009) Location, location, location: utilizing pipelines and services to more effectively georeference the world's biodiversity data. *BMC Bioinformatics 10*, S3.

Jones, M.B., Schildhauer, M.P., & Reichman, O.J. (2006) The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Review of Ecology, *Evolution and Systematics* 37, pp 519–544.

Pauly, D. & Froese, R. (2000) *FishBase 2000: concepts, design and data sources*. ICLARM, Los Banos, Laguna, Philippines, pp 1–344.

Shao, K.T, Lin, Y.C., & Lin, H.H. (2007a) Linking the Taiwan fish database to the global database. *Data Science Journal* 6, pp 164–171.

Shao, K.T., Peng, C.I., Yen, E, Lai, K.C., Wang, M.C., Lin, Y.C., Lee, H., Yang, A., & Chen, X.Y. (2007b) Integration of biodiversity database in Taiwan and linkage to global databases. *Data Science Journal 6*, pp 2–10.

Shao, K.T., Lai, K.C., Lin, Y.C., & Lee, H. (2007c) Progress and impediment of the integration of biodiversity database. *Handbook of Open & Free: New Enterprise in the Information Age — An International Workshop*, pp 17–18. Institute of Information Science, Academia Sinica, Taipei.

Shao, K.T., Huang, S.C., Chen, S., Lin, Y.C., Lai, K.C., Ko, C.J., Chen, L.S., & Yang, J.L. (2008a) Establishing a Taiwan biodiversity information network and its integration with germplasm databanks. TARI, COA. *APEC-ATCWG Workshop on Capacity Building for Risk Management Systems on Genetic Resources*, pp 71–78. Hsin-Hua, Taiwan.

Shao, K.T., Peng, C.I., & Wu, W.J. (2008b) *2008 Taiwan Species Diversity I. Research and Status*, pp 1–373. Taipei: Forest Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica. (in Chinese)

Shao, K.T., Peng, C.I. & Wu, W.J. (2008c) *2008 Taiwan Species Diversity II. Species Checklist*, pp 1–796. Taipei: Forest Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica. (in Chinese)

Shao, K.T., Peng, C.I. & Wu, W.J. (2010) *Taiwan Species Checklist 2010*, pp 1–840. Taipei: Forest Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica. (in Chinese)

Chavan, V. & Penev, L. (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12 (Suppl 15):S2.

(Article history: Available online 17 February 2013)