

台湾生物多样性资料整合之经验与策略

邵广昭* 赖昆祺 林永昌 柯智仁

李瀚 洪铃雅 陈岳智 陈丽西

(台湾中研院生物多样性研究中心, 台北市南港区 11529)

摘要: 台湾生物多样性数据库之整合从2001年开始, 是因为数位典藏计划、生物多样性推动方案, 及台湾加入全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF), 均在这一年启动。2002年“中研院”开始建置台湾物种名录数据库(TaiBNET), GBIF之台湾入口网TaiBIF则是在2004年时建置, 用来整合台湾生物多样性的资料并与国际接轨。所采用之方法及格式均依循GBIF所发展的交换标准, 一来可以整合台湾的数据, 二来可及时与国际交换数据。虽然TaiBNET及TaiBIF已突破智慧财产权(知识产权)的障碍, 可搜集整合数位典藏各子计划逐年累积的资料, 但跨部门间及非数位典藏计划所产生的数据, 仍因各单位及个人的本位主义而难以整合分享, 特别是生态分布原始数据。因此2008年在“中研院”成立了跨主管部门的委员会, 制订可行之资料搜集、整合与公开的政策, 并要求各主管部门在委办合约中纳入。无人否认数据库整合的重要, 但在现行对研究人员考评制度下, 研究人员大多不愿投入数据库建置的学术服务工作, 所获得的人力与经费亦日益短缺而难以永续经营。亟需相关部门的重视与支持。TaiBIF在过去6年来的推动成果虽未臻理想, 但所获的经验和心得仍有可供外界参考与借镜之处。**关键词:** 生物多样性信息学, GBIF, TaiBNET, TaiBIF, 数据库

Experience and strategy of biodiversity data integration in Taiwan

Kwangtsao Shao*, Kunchi Lai, Yungchang Lin, Chihjen Ko, Han Lee, Lingya Hung, Yuehchih Chen, Leesea Chen

Biodiversity Research Center, Academia Sinica, Taipei 11529

Abstract: The integration of Taiwan's biodiversity databases started in 2001, the same year that both the Digital Archives Program (later renamed Taiwan e-Learning and Digital Archives Program; TELDAP) and Biodiversity Action Plan were launched and Taiwan joined the Global Biodiversity Information Facility (GBIF) as an Associate Participant. In 2002, Academia Sinica began the creation of the "Catalog of Life in Taiwan" database (TaiBNET). Taiwan's node of GBIF, TaiBIF, established in 2004, integrates Taiwan's biodiversity data and shares it with the global community. Both TaiBNET and TaiBIF have broken through the barrier of Intellectual Property Rights and can collect and integrate data accumulated by TELDAP's various sub-projects. However, raw data, especially those on ecological distribution, generated by different agencies or non-TELDAP projects are still dispersed due to parochialism. A cross-agency committee was thus established in Academia Sinica in 2008 to formulate policies on data collection and integration, as well as mechanisms to increase public availability of data. Any commissioned project will hereafter include these policy requirements in the contract. The results of TaiBIF's efforts over the past six years, though not perfect, can provide some information and insights for others to reference or replicate.

Key words: biodiversity informatics, GBIF, TaiBNET, TaiBIF, database

1 2001年是数据整合的元年

在2001年之前, 台湾生物多样性的数据库可以

说都是各行其是, 分散在各个机构或学术单位, 并没有作任何真正的横向整合, 最多只是在各网站里有相关链接, 可以连到其他网站或数据库的首页。

如各主管部门下有一些不同的署处或试验研究单位, 每一个都有各自的网站和数据库, 如渔业部门、林业部门、水产试验所、林业试验所、特有生物保育研究中心等。台湾与生物多样性相关的民间组织也甚多, 其中建置有网站的至少有30个以上, 如蝴蝶保育、珊瑚礁、鱼类、鲸豚学会、荒野保护协会等。此外, 台湾所推动的大型或整合型研究计划, 或是重要的任务工作也有各自的网站, 如林业主管部门的植群图(vegetation mapping)计划、自然资源生态数据库; 防检部门的入侵种网站; 经建部门及内政主管部门之国土地理信息系统; 或科学发展部门资助的生态长期研究网等等。但这些网站的内容多半只是计划简介、研究报告、文献、政策及法令的倡导等文件数据, 以及与民众的沟通、相关知识或新闻等, 几乎都没有提供各项研究计划的后设资料(元数据, metadata)或原始数据(primary data or raw data), 顶多只是消化过的图表数据。而且不少网站之内容贫乏, 缺少更新及维运, 甚至并未完全公开, 或须经授权或付费后才能点阅。

直到2001年起, 台湾生物多样性的数据才开始有了整合。特别是数位典藏科技计划的展开, 除将人文社会科学领域的文史、生活、档案、器物、艺术、语言、新闻、图像、地图、建筑、影音等文化遗产的数据建文件保存外, 也把自然遗产的生物与自然, 如生物标本及本土物种信息列入典藏的范围。台湾《生物多样性推动方案》也正好是在2001年才核定通过, 其中的D1201案, 即责成主办单位科学发展部门及九个协办单位, 开始搜集及整合生物多样性的资料并与国际接轨, 其内容包括专家名录、物种名录、典藏标本、地理分布、时空分布、入侵种、物种解说、文献及生物资源等。同年全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF)正式成立时台湾即加入为副会员(Associate Participants), 也因此可以配合其后设资料及交换平台之技术与标准, 如Darwin Core及DiGIR(Distributed Generic Information Retrieval) (<http://digir.net/>)来推动台湾生物多样性数据之整合, 并与其接轨。

此外, 科学发展部门从2002年起, 开始资助“中研院”整合台湾百余位分类学者建置了台湾物种名录之网页TaiBNET(<http://taibnet.sinica.edu.tw>)。而GBIF之台湾入口网TaiBIF(Taiwan Biodiversity In-

formation Facility) (<http://taibif.org.tw>)则是在2004年科学发展部门所资助的一年经费及农业主管部门半年的经费下建置。到2006年, 科学发展部门又再度补助“中研院”进行TaiBIF计划, 从事网页维护及其内容增修订与整合的工作, 包括软件工具之开发与教育训练等。科学发展部门则每年补助学者出席与GBIF相关国际会议的经费。于2007年因数位典藏科技计划进入第二期计划, 同时于2008年与数字学习计划合并, 成为数位学习与数位典藏计划(Taiwan e-Learning and Digital Archives Program, TELDAP), 共五年期计划(2008–2012), 并加入国际合作的项目, 其中也包括生物多样性的数据与国际合作, 所以我们就将TaiBIF之工作纳入TELDAP计划, 并重新做了整合与分工(邵广昭等, 2008a)。TaiBIF计划负责收集科学发展部门所委托生物志(英文版)与生态分布数据, 并与跨主管部门之数据作整合; 而TELDAP则是整合来自数典计划所委托之机构及公开征选计划所搜集之资料, 或生物与自然主题小组项下所有生物多样性的数据, 主要是标本、物种解说(中文)及文献等的资料, 并与国际接轨。

2 数位典藏计划整合成果

由于科学发展部门所主办的数位典藏科技计划的机构计划, 每年均有稳定及足够的经费, 可聘请专任的约聘助理, 协助研究人员将多年来累积的丰富藏品与数据数字化, 因此大幅增进了数字典藏的质与量。数位典藏计划的目的除了信息技术研发外, 更要将数据数字化, 予以典藏及整合, 并朝创意增值与产业结合方向作应用。所幸这项重要的基础建设工作, 在内容方面除了文化外, 亦包括生物的部分。生物部分的整合, 在过去一期计划中曾分成动物、植物两个主题小组来作横向或跨数据库的整合, 但在数典二期中, 动、植物已和地质主题小组合并为“生物与自然资源主题”一个组。

第一期五年的数典计划是由联合目录来搜集与整合所有的数典数据, 其网址为<http://catalog.ndap.org.tw>。第二期的计划则是由“成果入口网”来征集(图1)。它所收集之资料来源, 除台大、“中研院”、科博馆三个机构外, 还有二十几个公开征选的计划所收集的资料, 每一个计划各自均有自己的网页, 如“中研院”之鱼类、贝类、植物数据库, 科博馆的网页等。迄今已提供给联合目录整合的约有22万

笔数据, 影像数据大约24万笔。但仍有一些单位迄今迟迟未能将其数据提供到联合目录及GBIF。其原因除了智财权的问题外, 还包括其内部数据无法成功地整合以及不少标本仍有鉴定困难等问题。GBIF目前已累积全球各国节点共约2亿笔的物种原始分布数据可供下载及使用(Chapman, 2005)。

在智财权方面, 原本数典计划系采用“创用”(Creative Common, CC)授权的办法。即在符合姓名标示(Attribution)、非商业性(Noncommercial)、禁止改作(No Derivative Works)、相同方式分享(Share Alike)等不同组合的规定下, 能够上网公开, 供民众浏览使用。但数典成果中, 文化部分的内容, 不少仍有智慧财产权未厘清的问题, 有些资料的性质并非CC所能规范, 如原住民之传统智慧财产权已另有法律保护。因此数典计划在2008年起乃不得不重新进行逐笔数据的智慧财产权之盘点工作, 并请

数典计划的主持人重新签署授权同意书。如此才解决了联合目录的数据无法提供给TaiBIF, 再经由TaiBIF提供到GBIF的问题。

此外, 在联合目录的网页上虽然可以用类群或字符串(种名)来查物种信息, 但其格式设计主要是为了藏品的项目(items or collections), 如标本等, 而非其他更进一步的物种或生物信息, 且其呈现方式并不符合生物领域的习惯。这主要是因为联合目录是要整合人文与生物两种属性不同的数据, 若采用有限的Dublin Core的字段, 很多字段无法作适当的对应, 譬如只有标本信息(含图片)而没有较多的物种解说及GIS的分布资料, 也没有办法和国际上所用的Darwin Core作接轨及整合。故我们乃建议各计划之后所设数据应加入Darwin Core之必填栏位作标记, 并改由TaiBIF来收集数据, 再转给联合目录作计划绩效考核项目之一。

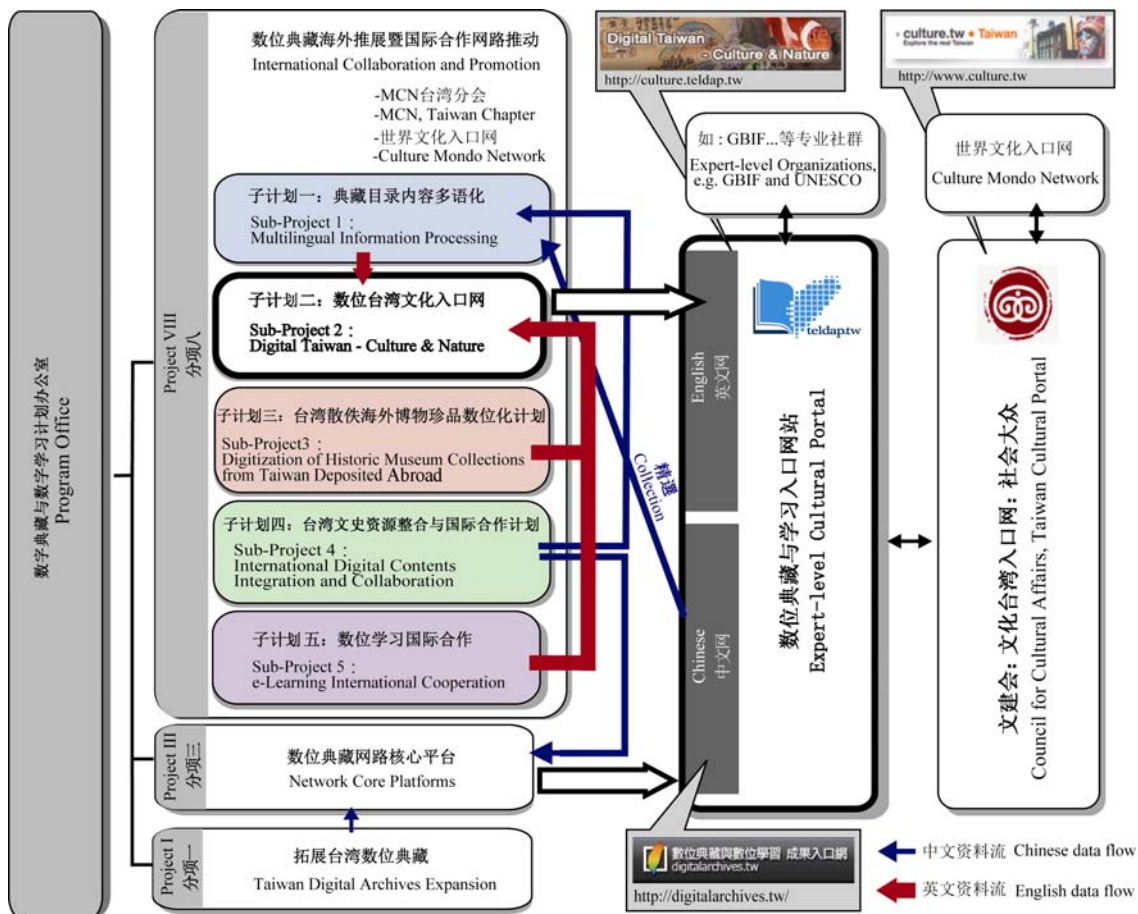


图1 数字典藏与数位学习国际合作架构图

Fig. 1 International collaboration chart of Taiwan e-Learning and Digital Archives Program

3 如何整合台湾生物多样性数据并与国际接轨

目前TaiBIF/TELDAP整合生物多样性数据库之方式, 主要是利用种名及GPS(经纬度)来整合跨主管部门、跨计划及民间, 或各学术单位及个人所建构之资料, 直接链接到各网站中该物种之信息上, 而不是先链接到各网站之首页, 再从首页向下搜索。故先建置一个权威的本土物种名录最为重要, 因物种名是生物多样性数据整合的关键词或linker, 可以经由学名查询到物种之标本数据、生命条形码(Barcode of Life, BOL)、生命之树(Tree of Life, TOL)、生态数据(EML格式)或最新的物种信息, 如网络生命大百科(Encyclopedia of Life, EOL)等。物种名录(Catalog of Life, COL)是由Species 2000 (Sp2K)及ITIS互相合作再提供给GBIF, 其目标在完成全球180万个物种的电子名录, 在2007年5月COL即已经达到了100万种。在BOL方面, “中研院”也是BOL联盟之成员, 2007年9月在“中研院”举办了第二届国际生命条形码会议。其中生命条形码数据库之整合也是一项重要的工作, 目前全球有6万多种78万多笔物种条形码已上线。

此外, 台湾方面除了进行物种出现记录整合外, 还具有植物的种原保存中心, 如植物园、改良场、种子库。在动物方面, 除动物园、水产试验所及畜产试验所之繁殖场或种原中心外(Shao *et al.*, 2008), 林业主管部门自2006年起也在推动一项野生动物冷冻遗传物质的典藏计划, “中研院”负责统筹及数据库建置, 目前已收集到液氮保存的约3,000种9,000多件标本。此冷冻遗传物质数据库可查到目前由各单位已收集到的标本信息(<http://cryobank.sinica.edu.tw>)。包括“中研院”(鱼)、特有生物研究保育中心及师范大学(鸟、哺乳类)、台北市立动物园及师范大学(两爬)、自然科学博物馆(无脊椎及昆虫)及畜产试验所(备份库)。

网络生命大百科已于2009年2月26日正式公布其首批3万种之网页, 每种一页, 共3万页的数据, 预计在2017年完成全球180万种之电子百科全书。首批完成的主要是鱼类, 因为全球鱼库(FishBase) (<http://www.fishbase.org>)已是全球最成功、最完整之物种数据库, 做起来最容易。台湾鱼类数据库 (<http://fishdb.sinica.edu.tw>)已与FishBase合作多年,

是FishBase的重要合作伙伴(Shao *et al.*, 2007a), 我们也打算与EOL合作, 推广其物种上传接口Lifedesk, 并提供台湾生命大百科中台湾特有种的网页。譬如图2中的台湾樱花钩吻鲑(*Oncorhynchus masou formosanus*)之样张左边是中文, 右边是英文, 内有生态照、原始描述、文献、相关报告, 甚至有此种鱼的钞票及邮票之图案等。

至于与国际接轨及分享数据, 共有三种不同途径(邵广昭, 2006; Shao *et al.*, 2007b)。除了TaiBIF直接连到GBIF的第一条路径外, 也可以通过与局域网络合作把数据送到国际上, 这些局域网络其实都是GBIF的副会员, 或是经由不同生物类和各类群之全球数据库(GSD)如FishBase或AntBase合作, 经由这些全球物种数据库, 再把台湾本土或特有种之资料送到国际合作计划, 如COL, BOL, EOL, BHL等不同目的的大型整合性数据库, 再串连到GBIF(图3)。

4 从物种名录(TaiBNET)的建置开始

生物多样性数据整合最基础的工作, 就是要先建立一个正确、权威、完整的物种名录。因此从2002年起我们就先建立TaiBNET网站, 来整合及更新台湾有效种的物种名录。迄今我们已收集了7界(细菌、古菌、原生生物、原藻、真菌、植物、动物)及滤过性病毒, 共56门, 2,945科, 51,572种的本土物种, 不包括栽培种、外来种、化石种名录在内。使用者可由种名、关键词或字符串来查询, 也可以用分类系统树来查询, 即由界、门、纲、目到科的拉丁学名、中文名或俗名来查该科下所有的种, 分类树形图亦提供每一分类阶层目前在台湾已记录到的属和种数。使用者如点选所要查的科, 即会列出该科所有的种。如点选种则可进一步连到各个数据库的该物种数据, 如台湾鱼类数据库、FishBase, 甚或Discovery Life及EOL之网页。2008年8月15日在自然科学博物馆举行的“2008台湾物种多样性研究现况研讨会”的成果, 即是在半年后正式汇编出版, 包括上册的研究现况及下册的物种名录, 二本书及一片光盘, 目前均已上网可供下载(邵广昭等, 2008b, c, d)。

台湾物种名录数据库的维护目前有专任助理负责, 同时亦依不同生物类群开放授权给各类群的分类学者或其博士生, 以按笔计酬的方式请他们自行上网作增补修订。最后再请各类群最顶尖的分类

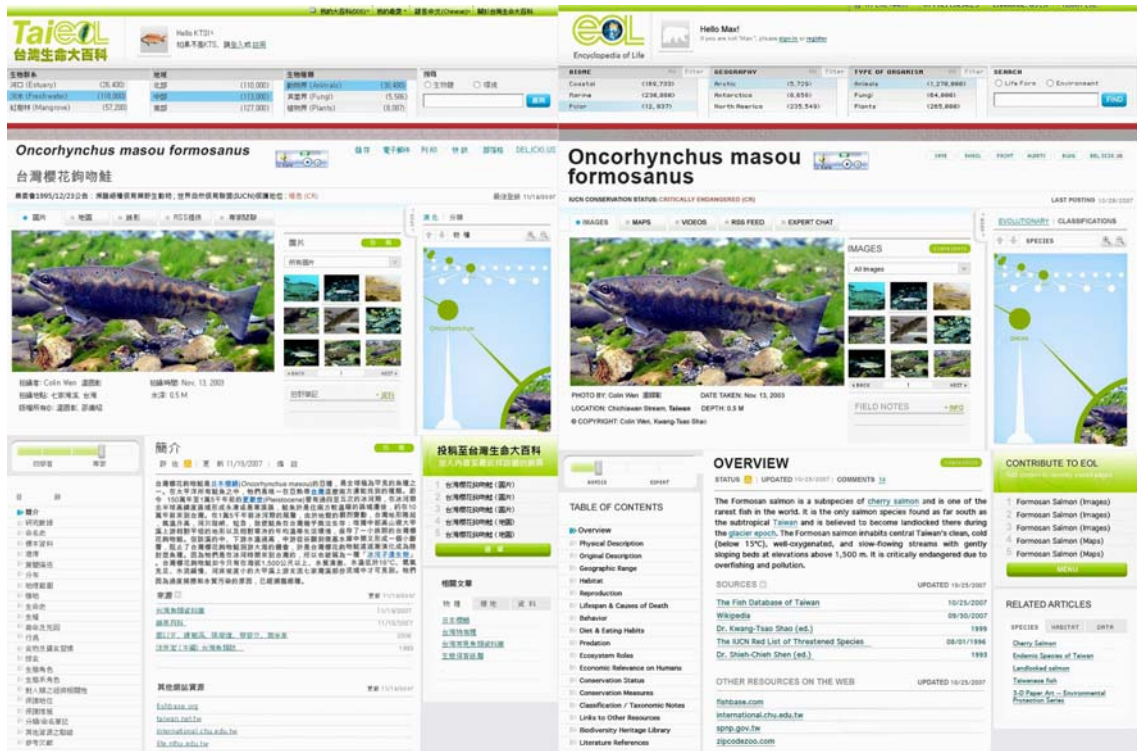


图2 中英文版台湾生命大百科示范页
Fig. 2 Chinese and English versions of TaiEOL

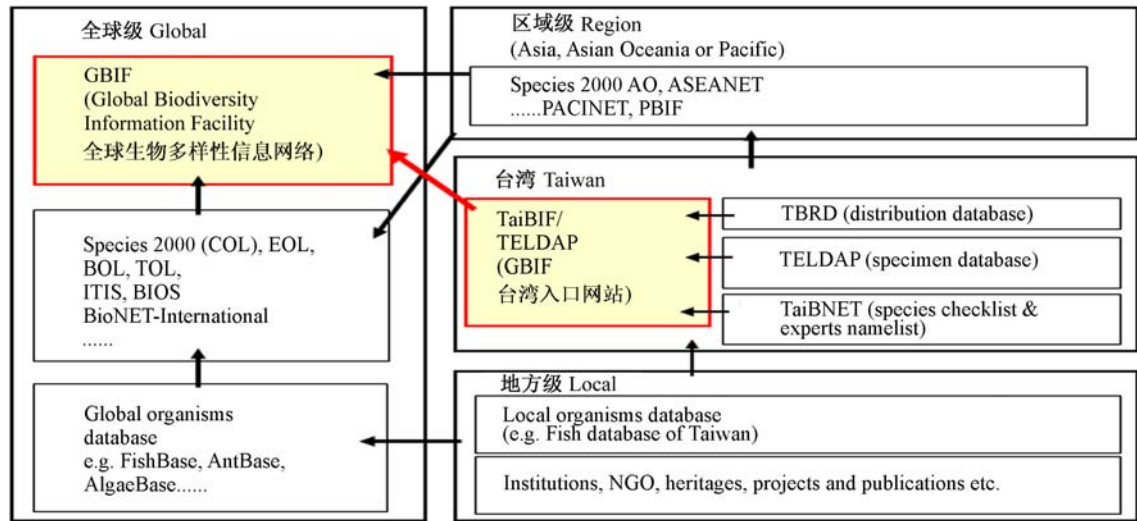


图3 整合台湾生物多样性信息网并与国际接轨
Fig. 3 Integration of biodiversity databases in Taiwan and link to global databases

学者来审订。另项检核数据的方法则是将TaiBNET的名录和Sp2K每年所提供之全球物种名录光盘作比对, 如发现不符合的种名, 可能是Sp2K尚未收录的种类, 也可能是笔误或同种异名等, 均会再请各分类群的专家协助核对。此外, 真菌尚可与全球

CABI之真菌名录比对; 海洋生物名录可与“全球海洋生物名录登录系统(WoRMS)” ([http:// www. marinespecies.org/](http://www.marinespecies.org/))作相互交叉比对。此外, 与国际合作的另一好处是可以取得对方丰富的数据, 如台湾本土鱼种之同种异名录14,000种, 即可由FishBase

中取得, 而台湾的有效鱼种迄2009年止, 共为3,061种。

同物异名之收录十分重要, 因为分类系统及物种学名一直不断地被修订。以鱼类为例, 约每隔10年, 1/10的鱼名就会被更动(Froese & Pauly, 2000)。即使是分类专家亦无法追踪与记忆, 惟有靠数据库来作自动比对察觉, 而把使用者输入的无效名自动指向有效名去作链接。否则许多标本馆中的标本如用无效名, 就永远搜寻不到了。

5 以TaiBIF与国际接轨

除物种名录外, 其他内容与国际合作及接轨主要是利用TaiBIF入口网, 包括物种解说、生态分布资料及文献、新闻等。目前除科学发展部门和农业主管部门已有资助经费在作名录的更新、生物志及生态分布数据之收录外, 其他数典及跨主管部门数据之整合则系由TELDAP来资助。图4为TaiBIF的首页。

以下是TaiBIF网站的内容及操作方法(邵广昭等, 2009)

5.1 目前在TaiBIF信息入口网上可查询的数据

(1) 物种名录: 承继台湾物种名录计划(TaiBNET)成果, 包含物种名称(学名、中文名)及分类阶层。目前记录有51,572个分类群及51,394笔同物异名资料。此数据除了提供查询, 扮演着本计划的骨干角色, 还用链接所有生物多样性数据及其在分类上的沿革关系。

(2) 物种出现记录(标本与观测数据): 来自台湾各主要数据库, 如数字典藏与数字学习科技计划(TELDAP)之标本典藏, 已包含台湾十多个单位: “中研院”、自然科学博物馆、台湾大学、台湾博物馆、林业试验所、水产试验所、农业试验所、海洋生物博物馆、海洋大学、“清华大学”及“中山大学”等参与, 透过TaiBIF可查询到台湾数字化标本及观测点, 总资料笔数为558,112笔。

(3) 生物志: 科学发展部门为落实推动方案之要求, 协助分类学者进行生物志内容撰写。目前包含第一期资料共有5,275种。

(4) 文献资料: 由财团法人实验研究院科技政策研究与信息中心提供与生物多样性相关文献资料, 目前所搜集的资料包含研究计划(8,079笔)、研究报告(3,987笔)、期刊论文(10,611笔)、会议论文(4,767笔)、硕博士论文(4,171笔)、英文期刊论文

(2,270笔)、出版品(499笔)等资料。

(5) 生态照: 结合cooliris之方式呈现台湾生物多样性信息入口网所搜集之生态照片与影片。

5.2 目前TaiBIF提供4种主要的查询方式

(1) 关键词、学名查询: 系统提供简易与进阶两种方式进行查询。输入时能通过AJAX链接TaiBNET中物种名录符合的学名, 除了可避免错字外, 及提供建议搜寻关键词(suggestion keyword), 亦可得知想要搜寻的名字是否在台湾的纪录中。且因物种的学名常因分类架构改变而产生同物异名的状况, 在TaiBIF的查询系统中亦会结合台湾物种名录(TaiBNET)中分类系统, 提供同物异名的查询机制以扩大学名的查询范围。

(2) 物种浏览: 以生物分类阶层进行浏览, 此物种分类阶层与台湾物种名录系统相同, 灰色字体表示尚未有数据开放链接, 明亮字体显示系统已建立整合数据。每一阶层的整合信息后方会显示(标本)链接, 点选后, 会显示系统中具有该阶层所有的标本馆藏数据及其分布的状况。

(3) 机构浏览: 以数据提供机构进行浏览, 点选任一个“数据提供机构”的链接, 立即能查询到该机构所提供的全部数据。

(4) 地图浏览: 以地理空间进行查询, 利用系统平台上地理空间框选的功能以获得该地区物种分布信息。

6 信息技术追随国际的主流

2004年开始时TaiBIF是使用DiGIR作为与GBIF秘书处间交换数据之存取协议, 由于若干数据域位对应上的问题及限制, CRIA(Centro de Referência em Informação Ambiental)开发了另一套TAPIR(TDWG Access Protocol for Information Retrieval)(http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/tdwg_tapir_specification_2009-09-08.htm)以改善此情况, 并获得GBIF正式采用, 同时也在2009年10月正式成为TDWG官方的标准之一。尽管该协议并非全然令人满意, 本计划着眼于标准推动的困难及其字段对应的灵活性, 在后续的合作中, 亦采用此协议并加以推广, 作为协助台湾机构成为GBIF数据提供者的协议工具。

TaiBIF过去采用DiGIR与GBIF相互分享数据, 但因DiGIR当时使用的Darwin Core为1.2版, 与新的



图4 台湾生物多样性信息入口网的首页

Fig. 4 Homepage of Taiwan Biodiversity Information Facility

Darwin Core(目前为1.4版)没有完整的对应,而对于中文数据域位亦无法自由对应,权衡之下中文必须与拉丁文或英文放在同一个字段,或选择不导出数据。另外,虽然可以整合查询目前已经架设好 DiGIR provider 的资源,但这样的页面对使用者而言未臻友善。GBIF 早期的数据查询接口也使用 DiGIR,后于2007年时,改以自行开发的数据入口网为主, TaiBIF 也利用类似的概念设计与开发台湾生物多样性信息入口网,除了前端的检索应用程序以外,后端的数据链路协议则使用 TAPIR。

TAPIR 是一个经由 HTTP 传送的、REST style 的分布式数据传输协议。TAPIR 合并与加强了 Bio-CASE 与 DiGIR 的优点,让使用者可以用更简洁的方法与数据提供者沟通。

TAPIR 提供三种不同程序语言的支持,分别为 PyWrapper3 (Python)、TAPIRLink(PHP), 以及 Tapir-DotNet (.NET)。除此之外,对于数据流通的标准也提供选择,包含不同版本的 Darwin Core, 以及提供自订 XML 以扩充数据定义的能力,在数据的分享上提供了更多的弹性。本计划团队也藉此发展中文字段对应的 XML 定义,满足利用 TAPIR 分享中文数据的需求。

TaiBIF 计划藉由与各单位信息人员的合作,将 TaiBIF 用以与 GBIF 接轨的信息技术(目前主要为 TAPIR)建置在各单位数据库中,并同时与该单位建立使用社群联系,共同解决技术问题,分享应用心得及成果。2009 年已协助生物资源保存及研究中心及特有生物保育中心架设。目前本计划正积极与台湾各主要生物多样性资源管理单位商谈合作细节,未来在 TaiBIF 及 GBIF 的平台上应可见到更多物种出现原始资料(图5)。

7 生物多样性数据整合之困难与解决方法

观测分布的原始数据搜集及整合最难,标本、文献、名录等资料较易,此乃因经过分类及生态学家鉴定及记录了物种的分布后,一旦放到网站上公开,则可以用于分析及写报告,不像标本,尚需取得原采集或管理者之同意才能借到。因此,过去调查者在发表报告前,均不愿提供及公开其数据,顶多只是提供经分析过的图表而已,更多的情况是数据遗失、毁损或被永远被埋没,十分可惜(Shao *et al.*, 2007c)。因此最近我们也在建议当局各委办单位在委托计划签约时,需明文规定计划结束后,计划主持人需要缴交原始数据文件。譬如渔业主管部门规

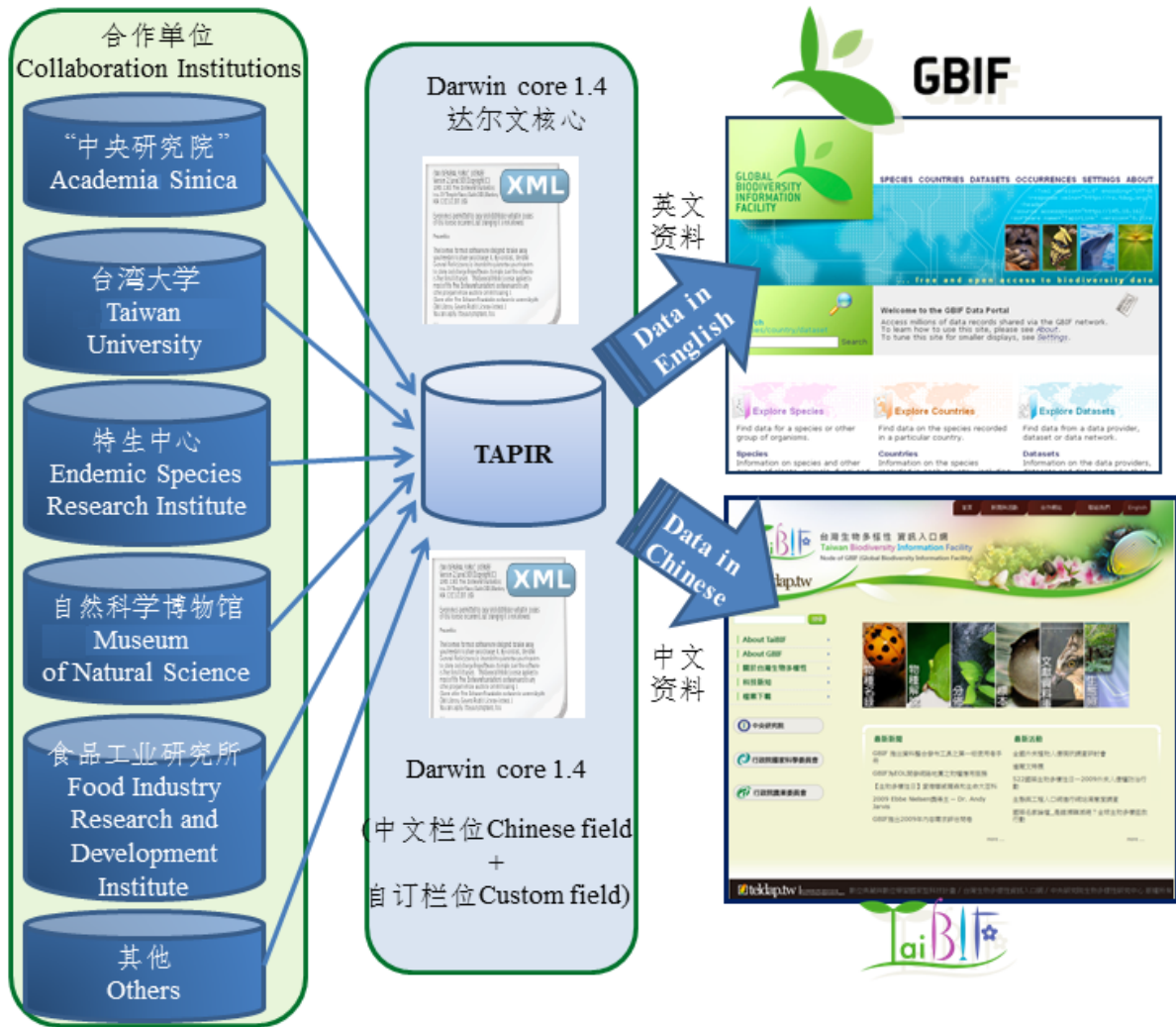


图5 利用TAPIR将各单位数据汇入台湾生物多样性资讯机构(TaiBIF) (中文)及全球生物多样性资讯机构(GBIF)(英文)流程图
 Fig. 5 Flow of data to Taiwan Biodiversity Information Facility (TaiBIF) (Chinese) and Global Biodiversity Information Facility (GBIF) (English)

定各计划都要每年分期中及期末两次依“通用数据库格式”(XML)缴交原始分布数据, 包括详细说明计划的目的、材料方法等。采用XML的好处是未来不论用哪一种元数据(metadata)均可以转换过去。

台湾研究发展考核主管部门于2006年虽已行文各相关主管部门, 要求推动信息公开, 但却无法落实执行及管理考核。复因智财权问题复杂, 大多数研究人员均不愿自动送缴原始数据文件, 委办单位也缺乏信息的典藏和管理单位, 以及有明确执行力的信息政策, 故各主管部门内或跨单位的数据整合在过去20年来始终没有明显的进展。有鉴于此, 在科学发展部门的委托下, 2008年“中研院”成立了

“GBIF台湾委员会”, 由各主管部门之代表及实际负责生物多样性数据库的学者专家们所组成, 开始建立数据交换标准, 研商制订可行之数据政策, 再请各主管部门在委办合约中纳入缴交原始资料及在一定期限内公开的条款。

7.1 生态分布资料建文件缴交原则

经过多次委员及专家会议, 于2009年7月20日的会议中已达成“GBIF—台湾委员会推动公务预算生态分布数据之搜集与建文件原则”的结论, 供永续会及生物多样性推动方案参考:

- (1) 数据缴交种类及范围。第一阶段仅要求缴交生态分布数据(物种出现纪录)之原始数据数字文

件, 后续将纳入监测结果等相关资料。

(2) 搜集并要求缴交调查资料。请主管部门于委办调查研究计划之合约书等相关资料内纳入应缴交原始资料之条文。

(3) 数据缴交格式、方法及管考: (i) 数据缴交应包含调查方法、调查项目、定义及调查数据, 其格式可采用既有之国际惯用数据标准格式。如 Darwin Core, Ecological Metadata Language, EML) (Jones *et al.*, 2006), ISO19115等; (ii) 各主管部门依其技术能力指定数据搜集(存储)单位; (iii) 各主管部门应研拟管考办法, 如资料缴交后始得拨款核定经费。

(4) 资料公开之期限(智财权问题)。由各主管部门自行决定数据公开之期限。

7.2 生态分布数据缴交之格式

目前已达成之共识要求所有各机构所有委办计划应缴交之生态分布数据, 应包含 1–12 项之 metadata 及第 13 项之原始数据(raw data):

(1) 名称(title);

(2) 数据拥有人(owner)信息, 至少包含姓名、组织、职称与联络信息;

(3) 研究合作个人或机关(associated parties)信息, 至少包含姓名、组织、职称与联络信息, 若无此项免填;

(4) 研究内容摘要(abstract);

(5) 关键词(keywords);

(6) 数据提供使用规则订定(usage rights);

(7) 数据使用联络人(contacts)信息, 至少包含姓名、组织、职称与联络信息;

(8) 数据内容描述, 包含材料方法(methods)、研究时间(temporal coverage)、地点(geographic coverage)、对象之分类(taxonomic coverage),

(9) 变量名称(name);

(10) 变量内容说明(label);

(11) 变量定义(definition);

(12) 变量量测定义, 包含量测种类(category)的细项定义, 如无顺序文字型态(nominal)或顺序尺度(ordinal)选项, 必须包含值的描述与定义, 若为等距(interval)或等比尺度(ratio)选项, 则必须包含单位(unit)、精度(precision)与数值型态(number type), 时间(date-time)类型则须有格式(format)与精度(precision)。

(13) 依 9–12 项所述的原始数据(raw data)。

更详细的内容、格式、规范及范例, 目前科学发展部门已委托林业试验所及“中研院”在制订, 在 2010 年中开始公布实施, 并推广到其他主管部门。

7.3 未来的挑战及对策

为能确实达成目标, 我们也已在更高的决策层次, 以由上而下的方式来推动。

但数据库整合工作在许多机构仍面临下列共同困难与挑战, 难以解决, 惟有靠各单位主管的认同及贯彻执行的决心, 增加所需的人力与物力, 及给予工作上的肯定与鼓励。

(1) 无专责信息业务之人员及编制, 承办之研究人员无法分心来进行管理及维护, 只好委外发包厂商来开发系统, 而不采用 open source, 以致于在未来无法永续维护与经营管理数据库。委托计划结束, 数据库的生命亦随之终结。

(2) 因属经常性之经营管理, 经费常遭到逐年压缩。

(3) 承办人员或研究人员之绩效未获肯定, 研究绩效之评比仅重视论文之发表(SCI点数), 以致于研究人员多不愿投入此工作或主动提供资料, 也造成工作传承的困难。

除上述困难外, 目前 TaiBIF 推动数据征集及整合的方法及策略还包括:

(1) 分工合作及分布式管理: TaiBIF 以分类为主(名录、标本及其分布、文献及解说等)并负责与国际接轨, 与 NGIS(National Geographical Information System)之生态分布观测数据及其他数据库作区隔。并建议采取分布式管理, 使各机构或研究人员能有自己的业绩及维护各自的网站。

(2) 多与其他网站合作、链接: 如此, 被点阅率才会增加。

(3) 经费及成果分享、建立公信力: 公平分配经费, 愿公开自己的资料, 成果归功于大家及数据提供者, 整合者需建立自己的公信力。

(4) 挑容易的先做: 先有成果才能继续申请经费, 不要好高骛远。

(5) 满足使用者之需求: 而非只考虑信息技术及数据整合的方便, 以使用者及数据提供者的方便为优先。

(6) 管考的落实: 才能有效地征集到数据。

(7) 多培育人才: 争取年轻学子出国开会及主

办研讨会之学习交流机会。

(8) 道德劝说数据整合的好处:

- 可协助作资料备份(异地备份)
- 资料检核(质量管理与改进)
- 回馈社会, 学术服务, 纳税人的权利
- 数据库被点阅到的机会愈多
- 协助将当局所投资之庞大科学研究及调查经费落实, 将原始资料建文件保存, 不致于在计划结束后, 无法取得原始数据, 或只是取得文件式报告
- 数据可再利用更新的统计软体来分析及作模拟, 如使用EML, Ecological Niche Model等
- 资源保育及永续利用, 经营管理政策之重要工具, 如生物多样性量化指标。

8 结论

整合工作的推动涉及流程的建立及其中各环节困难问题的克服, 问题触及之面向包含资料政策、智财权、数据质量观念、数据生产标准作业程序、数据标准及软硬件开发技术等等。本文仅就“中研院”生物多样性研究中心在过去近十年来所推动的经验与心得予以整理, 供大家参考。

参考文献

- Chapman AD (2005) *Uses of Primary Species-Occurrence Data*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Froese R, Pauly D (2000) *FishBase 2000: Concepts, Design and Data Sources*. ICLARM, Los Baños, Laguna, Philippines. 344 p.
- Jones MB, Schildhauer MP, Reichman OJ (2006) The new bioinformatics: integrating ecological data from the gene to the biosphere. In: *Annual Review of Ecology, Evolution and Systematics*, **37**, 519–544.
- Shao KT (邵广昭) (2006) Biodiversity informatics and the integration of Taiwan and global information. In: *Teaching Materials on Biodiversity—Introduction to Biodiversity* (生物多样性教材——生物多样性概论) (ed. Chou CH (周昌弘)), pp. 25–40. Pioneer Project Promotion Office for Biodiversity Personnel Training, Ministry of Education, Taipei. (in Chinese)
- Shao KT, Lin YC, Lin HH (2007a) Linking the Taiwan fish database to the global database. *Data Science Journal*, **6**, 164–171.
- Shao KT, Peng CI, Yen E, Lai KC, Wang MC, Lin YC, Lee H, Yang JL, Chen XY (2007b) Integration of biodiversity database in Taiwan and linkage to global databases. *Data Science Journal*, **6**, 2–10.
- Shao KT, Lai KC, Lin YC, Lee H (2007c) Progress and impediment of the integration of biodiversity database. In: *Handbook of Open and Free: New Enterprise in the Information Age — An International Workshop*, pp.17–18. Institute of Information Science, Academia Sinica, Taipei.
- Shao KT, Huang SC, Chen S, Lin YC, Lai KC, Ko BCJ, Chen LS, Yang AJ (2008) Establishing a Taiwan Biodiversity Information Network and Its Integration with Germplasm Databanks. APEC-ATCWG Workshop (2008) Capacity Building for Risk Management Systems on Genetic Resources, pp. 71–78. Hsin-Hua, Taiwan.
- Shao KT (邵广昭), Lai KC (赖昆祺), Lin YC (林永昌), Ko CJ (柯智仁), Chen LS (陈丽西), Lee H (李瀚), Lin HH (林欣桦) (2008a) Integration of TELDAP's biodiversity databases and linkage to global databases. In: *Proceedings of Insect and Mite Resource Management and Application Symposium* (昆虫与螨蛛资源之管理与应用研讨会专刊) (eds Huang KW (黄坤伟), Chan ML (詹美铃)), pp. 283–289. Museum of Natural Science, Taichung. (in Chinese with English abstract)
- Shao KT (邵广昭), Lee H (李瀚), Lin YC (林永昌) (2008b) Establishing Taiwan Species Checklist—TaiBNET. In: *2008 Taiwan Species Diversity. I. Research and Status* (2008台湾物种多样性I. 研究现况), 283–289. Forestry Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica, Taipei. (in Chinese with English Abstract)
- Shao KT (邵广昭), Peng CI (彭镜毅), Wu WJ (吴文哲) (2008c) *2008 Taiwan Species Diversity. I. Research and Status* (2008台湾物种多样性I. 研究现况). Forestry Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica, Taipei. (in Chinese with English abstract)
- Shao KT (邵广昭), Peng CI (彭镜毅), Wu WJ (吴文哲) (2008d) *2008 Taiwan Species Diversity. II. Species Checklist* (2008台湾物种多样性II. 物种名录). Forestry Bureau, Council of Agriculture and Biodiversity Research Center, Academia Sinica, Taipei. (in Chinese)
- Shao KT (邵广昭), Lai KC (赖昆祺), Hung LY (洪铃雅), Ko CJ (柯智仁), Gao JH (高君和), Chen LS (陈丽西), Chen YC (陈岳智), Yang YL (杨杰伦), Lin YC (林永昌) (2009) *TaiBIF Data Portal: A Practical Hands-on Tutorial* (TaiBIF—数据入口网操作手册). Biodiversity Research Center, Academia Sinica, Taipei. (in Chinese)

(责任编辑: 时意专)