生物多样性 2014, **22** (3): 264–276 Doi: 10.3724/SP.J.1003.2014.14012
Biodiversity Science http: //www.biodiversity-science.net

Scratchpads 2.0: 互联网时代的生物多样性虚拟研究环境

王利松[1*] Vincent S Smith[2] 张红瑞[1] 张宪春[1]

1 (中国科学院植物研究所系统与进化植物学国家重点实验室, 北京 100093)
2 (Biodiversity Informatics Group, Natural History Museum, Cromwell Road, London, SW7 5BD, UK)

摘要: Scratchpads 2.0系统是支持在线环境下生物多样性基础数据的创建、管理和高效利用的虚拟研究平台。本文对该系统的研发背景和现状、系统使用的关键特征(包括个人数据和在线资源的动态整合机制、多语言内容的创建和管理、系统使用授权、动态数据追踪、团队协作, 以及数据论文的发表机制), 以及系统开发和管理者关心的主要技术问题(包括系统安装和高效维护管理、分布式系统架构、模块化开发和管理机制、相关的技术标准)进行了介绍。并针对与Scratchpads 2.0相关的生物多样性信息工具研发和应用的问题进行了讨论。Scratchpads准确的角色定位、学科业务需求的深度挖掘和优越的技术实现, 决定了它是网络时代分类学研究的重要基础设施之一, 将为世界在线植物志的实现提供重要的技术基础。

关键词: 生物多样性信息学, 数据发表, 在线植物志, 虚拟研究环境

Scratchpads 2.0: a virtual research environment for biodiversity sciences in the Internet era

Lisong Wang[1*], Vincent S Smith[2], Hongrui Zhang[1], Xianchun Zhang[1]

1 *State Key Laboratory of Systematic and Evolution Botany, Institute of Botany, Chinese Academy of Sciences, Beijing* 100093, *China*
2 *Biodiversity Informatics Group, Natural History Museum, Cromwell Road, London, SW7 5BD, UK*

Abstract: We describe key features of the Scratchpads 2.0 Virtual Research Environment (VRE), which supports the creation, management and reuse of biodiversity data. This paper provides an introduction to recent developments and status of the Scratchpads 2.0 system, including its technical architecture. Key features include mechanisms to integrate individual research data and online resources, creation and management of multilingual content, license and authorization of system and data content, dynamic tracing of data editing history, research team cooperation, and methods of data paper publication. Important technical features include effective maintenance and installation of the system, ability to build distributed architecture, modularized function and development, and implementation of related information standards. These are put into a context with related biodiversity informatics tools. Scratchpads was designed with accurate role orientation, a deep understanding of taxonomic research requirements, and excellent technical solutions. All of these attributes contribute to Scratchpads' importance to e-infrastrature in the Internet era for taxonomy, thereby providing us with a promising tool to complete ambitious projects like World Online Flora.

Key words: biodiversity informatics, data-publishing, online flora, virtual research environment

过去近20年来, 以物种2000(http://www.sp2000.org/)、全球生物多样性网络(Global Biodiversity Information Facility, GBIF) (http://www.gbif.org/)、生命条形码联盟(Consortium for the Barcode of Life, CBOL)(http://www.barcoding.si.edu/)、JSTOR Plant Science(http://about.jstor.org/global-plants)和网

收稿日期: 2014-01-13; 接受日期: 2014-04-30
基金项目: 国家自然科学基金(30800057)和科技部基础性专项(2013FY112600)
* 通讯作者 Author for correspondence. E-mail: lswang@ibcas.ac.cn

络生命大百科(Encyclopedia of Life, EOL)(http://www.eol.org)等为代表的生物多样性信息网络的建立, 使原本深藏于标本馆和图书馆的大量原始数据和文献资料可以方便地获取, 并在这些规模数据基础上开展较大时空尺度的研究工作(Yesson, 2007; Whitlock, 2011), 发现新的科学知识和规律(Bebber *et al.*, 2010; Fontaine *et al.*, 2012)。

然而, 科研网络信息化并不仅仅是将科研人员已经发表的成果和数据进行加工, 上传到数据库通过网络共享, 它还包括科研网络虚拟研究环境(Virtual Research Environment, VRE)的构建, 也即将基本的科研工作流程, 包括数据采集、管理、分析和发表等环节, 尽可能地实现网络化(Duin & van den Besselaar, 2011)。基于研究领域的VRE环境建设也是国际科研信息网络建设的热点问题之一(Chavan & Ingwersen, 2009; Smith, 2009; Wieczorek *et al.*, 2012; Costello *et al.*, 2013b)。VRE的建立对具有“小研究大科学模式”(small studies and big sciences)特征的学科, 如研究自然历史的生物多样性、分类学和系统学来说尤为重要。VRE将极大地改变传统研究的工作模式, 加速生物多样性的记录和发现过程(Smith & Penev, 2011; Johnson, 2012; Maddison *et al.*, 2012; Scotland & Wood, 2012), 并最终使生物多样性等基础研究所产生的“小数据”发挥大作用(Moore, 2011; Winker & Withrow, 2013)。VRE的核心问题是它能够在目前的科学技术条件下, 有效解决大科学命题下学科数据资源的科学管理和分析问题。

小研究大科学模式的含义是: 一方面, 研究自然历史的学科常常都有非常宏大的科学命题, 比如完整地记录世界的生物多样性, 破解物种起源之谜和构建生命之树。这些命题经常是学科的终极目标。从这个意义上说它们是“大科学”。另一方面, 在研究实践中, 由于研究对象的广博和复杂性, 研究者常只能对有限区域的有限对象(如具体的科属种)有较为详尽和深入的研究, 研究成果也常以新物种的描述、科属类群的修订等大量“小文章”的形式出现。这些研究成果所积累的数据和信息对宏大的科学命题来说都是具体的案例和数据元素。从这个意义上说它们的科学实践模式是“小研究”。也有国外研究者将分类学的这种实践过程描述为“团队项目”(team sport)(Knapp, 2008)。

由于传统出版模式的制约, “大部头”植物志在使用中有诸多缺陷: 如体积庞大不易携带、语言版本不同、信息更新不及时和原始数据丢失等(Kress, 2004; Thomas *et al.*, 2011)。利用现代的网络信息技术将分类学相关研究活动移到互联网, 是生物多样性研究群体当下最热门的话题之一(Godfray *et al.*, 2007; Mayo, 2008; Clark *et al.*, 2009; Smith *et al.*, 2009; Costello *et al.*, 2013a; Marhold *et al.*, 2013)。那么, 对于具体的研究人员来说, 如何使用网络化的工具来参与类似在线植物志的研究工作, 从而实现宏伟的大科学目标, 这是本文介绍Scratchpads 2.0(以下简称S2)平台的主要目的之一。

1　什么是Scratchpads 2.0?

S2(http://scratchpads.eu/)是一个生物多样性在线虚拟研究环境。由本文作者之一的Vincents Smith博士牵头的英国自然历史博物馆(Natural Historical Museum, London, NHM)生物多样性信息学研究小组来主导系统的研发任务。S2的定位是: 在开放科学(Open Sciences)(Vision, 2010)理念下, 实现包括数据采集、管理、分析到发表的无缝集成的研究工作流, 为研究人员打造现代化的e-taxonomy网络研究平台(Smith, 2009; Smith *et al.*, 2009, 2011; Brake *et al.*, 2011; Smith & Penev, 2011)。

S2是一个开源软件(Free and Open Source Software, FOSS), 任何个人和机构都可以在自己的研究和项目中公开使用, 并通过它向同行和社会分享自己的研究进展和成果, 以及建立自己的研究网络。利用S2建立的站点内容可以是某个特定的分类群(比如科和属)、特定区域的生物区系以及任何与研究自然历史相关的内容(表1)。S2最初由EDIT(European Distributed Institute of Taxonomy, http://www.etaxonomy.eu/)项目基于开源内容管理系统(Content Management Systems, CMS) Drupal 5发布。从2008年底到2011年, 系统实现了从Drupal 5到Drupal 7核心模块的升级更新, 并开发了大量针对生物多样性基础研究的模块。S2得到欧盟第七框架(EU/FP7)的ViBRANT (http://vbrant.eu/)项目和英国环境调查委员会(Natural Environment Research Council, NERC)的eMonocot(http://e-monocot.org/)项目以及NHM的共同资助。S2包括非常庞大的研发计划(Smith *et al.*, 2009), 它目前针对分类学研发了90

表1 Scratchpads 2.0的应用案例
Table 1 Case study of using Scratchpads 2.0

内容分类 Category	站点名称 Name of sites	使用的S2关键功能 Key Scratchpads features used
农业和园艺 Agriculture and horticulture	世界茄科在线 Solanaceae Source (http://solanaceae.myspecies.info/)	植物分类 Botanical taxonomy 野外调查页面 Field work pages BRAHMS 数据导入 BRAHMS data import 物种多媒体 Taxonomic media 协同管理 Collaborator management
动物多样性 Animal biodiversity	非洲鱼类学门户 African Ichthyology Portal (http://africhthy.org/)	多语言支持(英语/法语) Multi-lingual support (English/French) 社区论坛 Community forums 动物分类 Zoological taxonomy 学术文献 Scientific literature 物种多媒体 Taxonomic media
公众科学 Citizen science	硅藻在线 Diatoms Online (http://diatoms.myspecies.info/)	达尔文核心标准 Darwin-core specimen records 动物分类 Zoological taxonomy 多用户博客 Multi-user blogs
保护 Conservation	植物红皮书索引 IUCN Sampled Red List Index for Plants (http://threatenedplants.myspecies.info/)	保护评估 Conservation assessments IUCN数据整合 IUCN data integration 多元植物分类 Multiple botanical classifications 科学文献 Scientific literature
入侵种 Invasive species	认识蚂蚁 Antkey (http://antkey.org/)	多语言支持(英语/中文/印度尼西亚语) Multi-lingual support (English/Chinese/ Indonesian) 解剖学词汇 Anatomical glossary 物种分布图 Species occurence maps
植物多样性 Plant Biodiversity	加拿大和阿拉斯加极地植物区系 Arctic Flora of Canada and Alaska (http://arcticplants.myspecies.info/)	植物分类 Botanical taxonomy 自定义页面 Custom pages 自定义数据内容 Embedded custom content Google地图 Google Maps
专业组织 Society	国际蝽类昆虫学会 The International Heteropterists' Society (http://ihs.myspecies.info/)	个人群组 Private organisational groups 群组交流工具 Group communication tools 科学文献 Scientific literature 动物分类 Zoological taxonomy 物种多媒体 Taxonomic media
系统学 Systematics	蚊子的分类和编目 Mosquito Taxonomic Inventory (http://mosquito-taxonomic-inventory.info/)	解剖学词汇 Anatomical glossary 动物分类 Zoological taxonomy 物种多媒体 Taxonomic media 图库 Galleries 科学文献 Scientific literature

多个功能模块, 建立了完善的研发工作流程、支持服务系统以及在线培训和推广计划。在不到两年的时间内, 获得了生物多样性研究群体的广泛认可。目前, 全世界已有605个基于S2的子站点, 7,193位用户使用S2来管理自己的研究项目和日常工作。这些站点信息量包括了动植物和微生物的141万物种(Smith *et al.*, 2011)。一些典型的应用案例见表1。

2 重要的使用功能和特点

S2通过大量新技术的应用建立了易用且功能丰富的可视化管理界面(附图1–3)。作为用户, 研究者在研究过程中已积累了大量研究资料, 这些资料可能以多种形式保存, 比如小型办公数据库(Access)、电子表格(Excel, csv)、文档(pdf, doc, txt)、各种专业软件(如PAUP和DELTA), 以及各种文献管理工具, (如EndNote和Reference Manager)。研究者除了已积累的数据, 可能还需要利用大量的在线数据资源。如何将这些分散的数据资源通过S2进行集中管理? 如何在网络环境下有效保护自己的知识产权和对数据相关信息的授权使用? 如何与同行和团队的其他成员协同工作? 如何管理数据的动态修改历史? 以及如何将这些数据通过数据论文(data paper)的形式向期刊投稿? 下面将从这些角度来介绍S2的核心功能。当然, 这里并不是详尽的

S2使用手册, 我们只是希望从用户的角度来概述S2的部分重要和独特的功能。如果希望详细了解如何使用S2, 可以参考其在线帮助系统(http://help.scratchpads.eu/w/Main_Page)。

2.1 数据整合

根据数据内容(content type)及其表现形式(presentation type), S2将生物多样性基础数据分为物种名录(taxonomy)、文献(biblio)、类群描述(taxon description)、地名(location)、标本(specimen/observation)、图像(media gallery)、分子序列(molecular sequence)、形态性状(characters)和系统发育(phylogeny)9个大类, 分别对应相关的数据模块(图1)。因为物种学名在生物多样性信息系统建设中起着最重要的标识作用, 是整合和链接其他信息的关键纽带(Patterson *et al.*, 2010; Platnick, 2013), 因此, S2系统组织所有信息的框架都是生物分类系统(界/门/纲/目/科/属/种), 每一条学名在S2中定义为分类术语(taxonomic term)。建立自己的S2系统的第一个步骤就是导入所研究类群的学名。有了分类术语基础, 后续进入系统的标本、物种描述、文献和分子序列等信息就可以和学名建立关联(图1), 从而使系统中所有信息形成相互交织的链接关系, 为信息的高效检索和利用建立基础。目前, S2可以通过内置的电子表格模板(Excel)对名录、类群描述、地名和标本4类数据进行批量导入, 文献可以通过EndNote、Google Scholar、Reference Manager的通用文本格式导入。S2与许多常用的数据工具, 比如植物学研究和标本馆管理系统(Botanical Research and Herbarium Management System, BRAHMS)和DELTA建立了无缝的数据交换机制(图1)。

S2设计了自动化的方式利用在线数据。比如物种名录, 使用者可以选择从物种2000、NCBI (National Center for Biotechnology Information)、整合分类信息系统(Integrated Taxonomic Information System, ITIS)或者是EOL等数据源导入所需要的物种工作名单(附图1), 只需要输入要导入的类群学名即可。如用户要导入整个松科植物的名录, 只需要在界面输入松科的学名“Pinaceae”, 剩下的工作由系统自动完成。导入完成后就可以在可视化的编辑界面上对数据进一步整理(附图2)。

除了存储在数据库中的数据, S2设计了自动化的方式整合来自NCBI的分子序列、GBIF的标本、生物多样性遗产图书馆(Biodiversity Heritage Library, BHL)的文献、EOL的物种信息以及其他网络多媒体和图片信息(图1)。S2对外源数据的利用方式类似撰写论文中的“文献引证”, 即以Mashup的形式集成相关的信息在类群页面(Taxon Page)(图1)。通过这种方式能够清楚地向使用者表明, 哪些数据是研究者自己的数据, 哪些数据是其他网络来源的。当然, S2管理员也可以对这些外源数据审核和校验后, 将它们存储到本地数据集中。比如来自网络的大量图片信息, 其鉴定的准确性可能存在各种各样的问题, 但也许图片质量高, 或是记录了非常少见的一种植物及其关键的形态特征或生境。S2管理员可以对这些图片重新鉴定整理后归入到自己的数据集中。

2.2 多语言环境/修订/跟踪

研究工作的国际化需要以不同语言向同行、合作者或期刊杂志传播相关的研究成果。比如《中国植物志》和*Flora of China*就是比较典型的例子。前者是中文, 是中国植物家底资料的奠基性工作, 在国内有非常广泛的用户群体, 而*Flora of China*不仅反映大量类群的最新研究成果, 而且能被国际同行方便地利用。由于传统印刷出版模式的制约, 很难将这样的志书同时以多种语言形式呈现给读者, 并且保持信息的动态更新。一些研究者认为, 分类学的印刷版成果在出版的时候就已经“过时了”(Kress, 2004; Clark *et al.*, 2009)。同样, 世界在线植物志的编研也面临类似的问题。然而, 多语言内容和数据动态更新的问题在S2上得到了很好的解决。

传统的系统通常只能对应用程序的界面(如标题或者导航菜单)提供不同语言形式, 对具体数据内容常需要依赖Google translator这样的第三方翻译系统, 而这些自动翻译系统在专业领域的应用效果目前还非常不理想(Marhold *et al.*, 2013)。S2对多语言内容的处理方式是: 手工翻译内容, 系统对不同语言内容分别存储, 然后通过多语言处理模块、版本控制和文件比较功能实现对这些多语言数据内容的对照、历史跟踪和修订管理。

以物种的描述性信息(形态描述、地理分布、保护状态和经济利用等)为例(附图3), 用户在编辑器中输入相关内容, 同时可以选择该条信息的语言类型。保存该信息时, 系统会提示是否将该条信息翻译为其他语言版本。当用户浏览或查询相关信息时, 系统会自动根据其计算机操作系统和浏览器的语

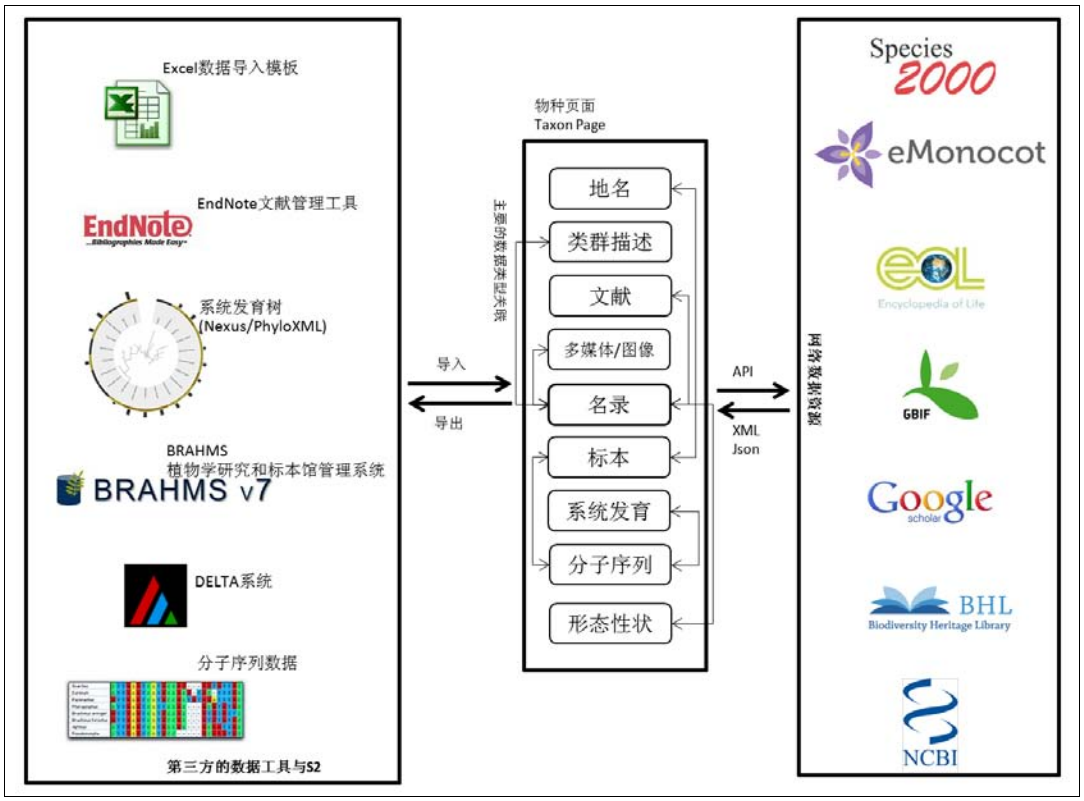


图1 Scratchpads 2.0系统基本数据类型和流程(箭头表示数据类型间的关联关系)
Fig. 1 Basic content types and workflow in Scratchpads 2.0. Arrows indicate the relationship between different data content

言版本，判断其最可能的语言环境，并将该语言环境的信息反馈给用户。如果有其他语言形式的相应内容，系统会在页面提供导航。而且，多语言内容的处理由用户灵活控制。比如，可以设定哪些内容需要或不需要进行多语言版本的对照。

传统的系统数据动态更新大多是通过时间戳和作者标记来进行的，从中可以了解该项信息的创建时间、最近的修改时间以及修改者。但具体对这些信息到底作了哪些修改，以及它修改前的内容是什么，大多数系统无法实现对信息内容本身的动态追踪和管理。分类学研究中，对基础数据的编辑和修订是一项非常琐碎和耗时的工作，它可能是仅仅修改某个学名的一两个字母、名称作者的引证格式(缩写还是全拼)，也可能是物种形态描述的某些特征，或者是某些分布记录等等。而且修订和校正是一个不断反复的过程。S2考虑到分类学研究实践中的这种具体需求，通过计算机领域广泛使用的文件比较(diff)(http://en.wikipedia.org/wiki/Diff)和版本(version)功能来实现数据内容的动态追踪。diff提供的是对单项信息逐行的追踪比较，它能够显示用户对某项信息的详细修改记录(可以具体到修订的单个字符)，以不同颜色来标记修改的状态(标记哪些内容是新增加的，哪些是删除或修改过的)。version则是通过对某一类信息集合的追踪。比如有关一个物种的描述信息可能存储在多个字段和表中，通过version功能能够将这些信息在某个时间的修订作为某个时间的版本存储下来，将来对它的更改则生成与相应时间和操作相关的版本。

S2的这种动态修订管理功能对我们的研究实践有非常重要的价值。比如，可以利用它来比较同一个物种在不同植物志中描述的差异；将我们的数据内容按周期存储为不同版本，来反映不同时间周期所完成的具体工作和动态。因此，S2并不是为某个特定的数字化项目而设计的平台，而是为研究者的日常研究工作服务，研究者通过这个平台将日常研究和数字化应用有机地结合在一起。

2.3 使用授权/版权/团队协作

S2是基于社区开源软件Drupal 7基础上的二次

开发和集成。因此它在使用上遵守GNU通用公共授权协议(GNU General Public License)。也即任何组织和研究者都可以自由免费使用该软件, 甚至按照自己的需要进行修改。通过S2所建立的站点及其内容由使用者或使用者所属的研究机构所有, 并保证其数据的科学性和合法性。并且, S2使用者可以在遵守知识共享协议(Creative Commons Licences, http://creativecommons.net.cn/)的前提下, 建立对自己的站点内容复制、发布、编辑等信息使用的授权协议。S2软件开发者只对系统所涉及的技术问题负责, 推荐站点拥有者使用限制最小的CC-BY署名协议 (http://creativecommons.net.cn/licenses/meet-the-licenses/), 以使站点内容的自由发布和传播最大化。因此, S2对于非盈利性的科学研究机构和个人来说, 能够降低研发和应用成本。

分类学“团队项目”(Knapp, 2008)的特点意味着它需要研究者通过相互合作的方式来产生有影响力的成果, 比如《中国植物志》、*Flora of China*和将来的世界在线植物志。以志书的编研为例, 在实施过程中需要建立编研指南, 内容包括分类概念、名称、文献和标本的引证规范等诸多细节问题。但是在项目的具体实施过程中, 因为编研者众多, 出现的各种问题无法及时更正, 当我们将大量的结果汇总到一起后可能会发现, 并不是所有的信息都是规范和有效的。

通过S2网络系统, 不仅能够对项目的实施设立统一的规范和标准, 还能够对众多研究者的编研活动进行动态追踪和管理。快速而有效的沟通和反馈机制的建立一方面是通过它的网络化, 另一方面是通过它内置的一些模块, 如基于角色(role)的用户管理。S2默认设置了管理者(administrator)、维护者(maintainer)、贡献者(contributor)和编辑者(editor)四种角色。管理者拥有系统设置、权限分配、用户管理、模块的开启和关闭、数据管理、访问控制等最高权限。维护者的角色是负责系统的运行和性能的调节, 比如决定在什么时间和按什么周期对系统运行产生的一些“脏数据”进行清理。这个角色一般分配给有技术基础的用户。贡献者和编辑者的权限主要是数据管理, 适合专业研究人员。S2的角色和权限管理是细粒度水平的, 能够详细到哪个用户可以对哪些具体数据进行相关操作(如增加、修改、删除和发布)。而且, 系统针对匿名访问和注册用户分别设置不同的访问权限, 即哪些内容只对注册用户可见。

可见, 像志书编研这类多人参与的项目, 利用S2平台的优势是非常明显的。不管项目的参与者身处何地, 他所需要的只是一台可以联网的计算机, 就可以和项目其他参与者进行快速的沟通交流和协同编研工作。系统通过署名、时间戳、diff和version等多项功能来标记不同参与者的实际贡献和进展情况, 作为项目管理者也能够快速知晓各个类群的编研动态和进展, 并将项目的动态信息反馈给编研者。世界茄科在线(表1)和世界单子叶植物在线(eMonocots, http://e-monocot.org/), 向我们展示了世界不同地区的科学家如何通过S2进行在线协同研究。世界茄科在线曾于2004–2009年获得美国国家自然基金的资助, 它的目标是对有重要经济、食用和园艺价值的茄科植物进行世界性修订, 与最新的系统发育框架(分子系统学)进行整合, 并通过网络平台的形式进行数据收集、分享和研究成果的发布工作。eMonocots是由NERC资助, 英国皇家植物园邱园组织世界单子叶植物专家共同建立的, 最终目标是建立世界单子叶植物在线植物志(邱园Paul Wiki博士, 个人交流)。

S2还利用了社交网络(social web)的功能来促进研究信息的社会化传播和利用。比如可以通过博客(Blog)、RSS订阅、OpenID(以用户为中心的数字身份识别框架: http://zh.wikipedia.org/wiki/OpenID)和分享模块(Jiathis/Sharethis)将相关的信息通过电子邮件、新闻(Newsletter)、链接等多种形式推送到腾讯、新浪、人人和微博等社交网络。

2.4　数据论文和发表

对科学研究成果开放式获取(Open Access)的讨论(Leptin, 2012)提出了开放科学的理念(Vision, 2010)。开放科学倡导无障碍地使用各项科研成果, 并且关注研究成果和数据发表过程的开放和透明。最近几年, 科研数据通过互联网正式发表已成为科学界的一个趋势(Costello, 2009; Smith, 2009; Vision, 2010; Brach & Boufford, 2011)。

由于出版模式的制约, 论文所基于的大量原始数据经常不包括在论文中, 或者是以补充性材料(supplimentary material)方式来公布剪裁过的信息。近20年来, 分类学研究论文由于所谓的“低影响”、篇幅长等原因, 已经不作为许多主流期刊的刊登内

容。但分类学本身是一个数据密集型(data intense)学科(Bowker, 2000), 它的绝大多数研究成果的数据本质是显而易见的。例如植物志、类群修订、专著、植物名录、分布图集、鉴定手册等, 其中所包括的植物学名、中文名、地理分布、形态描述、检索表、标本引证等信息, 每一部分都可以对应到数据中的一条或多条记录。在研究实践中, 分类学家经常需要借助专门的工具来处理这些格式化的信息, 以符合论文的出版要求。在这样的背景下, 有人提出用数据论文(Chavan & Penev, 2011)的概念来解决基础研究过程中原始数据的发表问题。

S2在这方面无疑是实践领域的先驱。它通过发表模块(publication module)和在线写作辅助工具(publication write tools)将存储在S2中的数据以数据论文的形式向Pensoft(Erwin *et al.*, 2011; Smith & Penev, 2011)出版商管理下的*ZooKeys*、*PhytoKeys*、*MycoKeys*和*Biodiversity Data Journal*等在线刊物提交论文, 以实现网络环境下从研究数据的收集、整理、分析到发表的完整流程(Penev *et al.*, 2010a, b)。Penev(2009)、Blagoderov(2010)和Smith(2011)分别介绍了通过S2生成和提交论文、在线交互检索表, 以及生成IUNC红色名录评估报告的详细流程和技术细节, 数据发表相关模块的研发也是最近S2团队关注的重点。

在以DNA分子序列数据为主的领域实际上已经具有了类似数据论文的实践。由于DNA分子信息开始广泛应用于物种及其相关问题的研究时, 计算机和互联网技术已经有了长足的发展, 分子系统学研究的蓬勃兴起也催生了像NCBI这样大型、普及的在线数据管理和应用系统(Agosti, 2003; Benson *et al.*, 2011)。大多数研究者已经默认在使用来自NCBI的数据时要在论文中明确引证和说明, 并将自己的序列数据向系统提交, 已经形成了滚雪球式的良性循环。DNA条形码研究也有类似的发展趋势。当研究者提出这个研究方向后不久(Hebert *et al.*, 2003), 它的基础信息管理和应用系统就应运而生, 如iBOL(http://ibol.org/)、CBOL和BOLD(http://www.barcodinglife.com/)。因此, S2在基于名录、标本、文献、图像等数据类型的网络化发表方面迈出了可喜的一步。

3 系统管理、维护和开发

3.1 安装运行/管理维护/分布式站点/使用申请

S2应用了大量开源软件领域的主流产品和新技术, 比如PHP编程语言、Drupal 7内容系统、MySQL数据库系统、Linux主机、Apache和Nginx网络服务器, 以及Memcache和Varnish缓存。S2使用PHP作为系统主要开发语言, 通过Git系统对应用程序的源代码进行管理。

使用者可将应用程序(Application)、数据库(Database)分布于异地的多个Windows或Linux系统主机, 并且多个应用程序可以共享或独占应用程序代码库(图2)。S2可与目前主流的数据存储系统如MySQL、MSSQL、Oracle、Postgresql, 甚至最新的非SQL(NoSQL)存储系统如MongoDB(http://www.mongodb.org/)结合使用。因此, 它为使用者在原有软硬件基础上升级, 并建立异地备份和镜像网站提供了灵活的选择方案。当用户希望使用S2来构建自己的系统时, 可以通过S2注册系统(http://get.scratchpads.eu/)进行申请, 系统会自动建立一个基于 http://xxx.myspecies.info 或 http://xxx.taxon.name 的二级域名系统。这个过程通过Aegir(http://www.aegirproject.org/)分布式管理系统来实现, 也即用户对系统的申请、建立和管理基本上是一个自动化的过程(图2)。虽然S2可以运行在Windows系统, 但我们推荐使用更为高效的LAMP(Linux-Apache-MySQL-PHP) 或 LNMP(Linux-Nginx-MySQL-PHP)的组合方式。S2的安装可以通过Profile模块进行个性化定制, 即用户在安装的时候就可以决定系统默认的语言环境开启的功能模块、系统的使用界面等多项参数。我们推荐使用Drush(http://www.drush.org/)来进行S2系统的管理。Drush可以运行在任何类Unix系统(包括MaxOS和Linux)环境下。在终端(terminal shell)通过简单的命令行来执行包括应用程序的安装、更新和备份, 数据迁移(Migration)、用户管理、主题设置、系统测试数据的自动生成等一系列复杂的任务。

3.2 模块和扩展

S2使用模块(Module)化的方式来组织整个应用程序。其优势在于具备足够的灵活和可扩展性, 非常适合程序功能、界面、需求存在诸多不确定的因素和需要经常变化的情况, 尤其适用于科研网络建

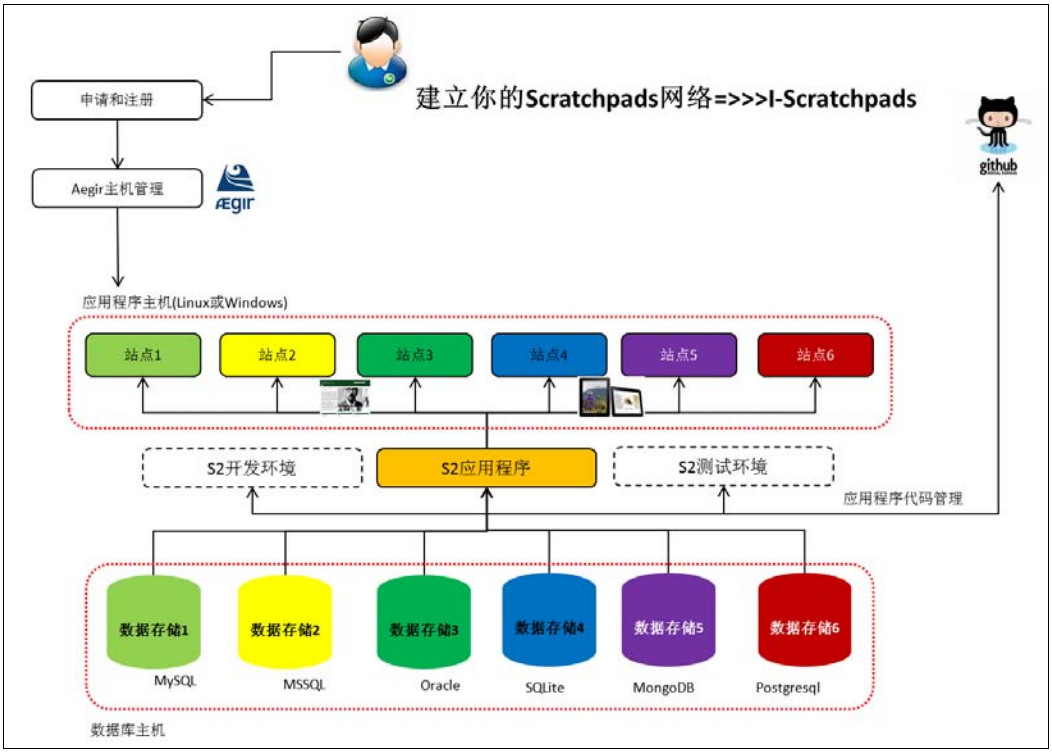


图2　Scratchpads 2.0系统申请注册流程及分布式应用的架构示意
Fig. 2　Workflow of application for Scratchpads 2.0 and distributed infrastructure illustration

设。模块化方式也体现了技术领域的一句俗语: “不要去重新发明轮子”。

因为S2构建在Drupal 7基础上, 除了S2本身与生物多样性研究相关的90多个模块外, 还可以使用由3万多名开发者贡献的2万多个模块(https://drupal.org/)。我们以视图(views)和字段(field)模块为例进一步说明。

视图模块的主要作用是: 用户可以自己定义系统采用哪些数据, 以什么样的形式, 在页面的哪个位置显示, 以及用户的访问权限。字段模块的主要功能是: 创建用户所需要的数据类型, 并与系统已有数据间建立关联关系。以建立分子序列数据类型为例, 只需要4个基本步骤就能完成分子数据存储模块的“开发”: (1)创建一个新的内容类型, 我们可以定义为molecular sequence; (2)定义该数据类型需要存储的信息有哪些, 比如序列的类别是ITS还是MatK; (3)定义该数据内容与已有数据的关联关系, 比如分子序列数据应该和凭证标本有最直接的关联关系; (4)定义该数据类型的表现形式和访问权限。因此, 对于熟练的S2系统管理者来说, 甚至不需要写一行PHP代码, 仅通过系统配置, 就可以建立功能丰富强大的分布式应用程序。

除模块所提供的功能外, S2通过应用程序接口与第三方程序或工具进行协同工作, 比如分类学研究中常用的植物学研究和标本馆管理系统BRAHMS、DELTA系统和系统发育分析软件Mesquite, 以及通过Taverna(http://www.taverna.org. uk/)工作流管理系统与R软件包(http://www.r-project.org/)和OpenModeller(http://openmodeller.sourceforge.net/)等第三方分析工具的集成。在这方面, S2秉承的理念是: 以S2的分布式数据存储为核心, 通过API机制和工作流系统, 尽可能整合已有的第三方数据源和分析工具, 从而形成一个以用户及其数据为核心, 灵活和多元化的应用系统。

3.3　标准和协议/系统服务

数据标准和协议是生物多样性系统建设初期需要关注的问题之一。S2充分利用了涉及物种名录、描述、标本、图像元数据、系统发育在内的多项行业标准(表2)。比如物种描述, 它采用了TDWG推荐的SPM模型标准, 对每一项信息都在用户界面

表2 Scratchpads 2.0目前移植的相关数据标准
Table 2 Data and metadata standards implemeneted in Scratchpads 2.0

数据和元数据标准 Data and metadata standards	说明 Notes
BibTeX, RIS, XML bibliographic citations (export can also be in XML, RTF or Tagged Field)	与各种桌面和在线文献管理工具，如Endnote, ReferenceManager, BitText和Google Scholar等进行双向数据交换
CSV / XLSX (spreadsheet content)	与电子表格数据如Excel、Access等进行数据交换(需要使用指定的数据模板格式)
Darwin Core Archive and selected extensions	达尔文数据标准，标本数据交换标准
EXIF, XMP image metadata	图像元数据交换标准
ITIS taxon metadata standard (for taxon names and hierarchies)	ITIS分类群元数据标准(适用于类群学名和分类树)
LUCID (for taxonomic keys)	与鉴定工具LUCID系统的数据交换
Nexus (character data, export only)	与系统发育分析系统或性状矩阵软件的数据交换格式Nexus
Species Profile Model (for taxon descriptions)	TDWG推荐的结构化的物种描述模型标准
XML (export only, selected content in the TaxPub schema, http://www.ncbi.nlm.nih.gov/books/NBK47081/)	NCBI所推荐的与NCBI相关期刊中物种描述的信息标准

有详细的描述和说明(表3)，这样开发者和使用者都能快速了解各项信息存储的位置和用途。S2还对广泛应用的达尔文核心标准(Wieczorek *et al.*, 2012)进行了改进，提出了DwC-A(Darwin Core Archive) (Baker *et al.*, 2014)标准。其优点是可以使用单一的数据文件将存储在S2系统中目标物种的分类、标本、物种描述等相关信息传输到第三方的数据集成器(aggregators) (如EOL和eMonocot门户)。对于高效的应用系统来说，数据服务总是双向的，即它不仅能够消化来自其他网络应用程序的数据，而且能够向其他应用程序提供数据服务。S2已经整合了BHL(http://www.biodiversitylibrary.org/Tools.aspx)、EOL(http://eol.org/api/docs/pages/1.0)、GBIF(http://ispecies.blogspot.com/2007/08/maps-and-google-tweak.html)和Morphbank(http://services.morphbank.net/mb)等主流生物多样性基础信息系统的数据服务。形态性状、分子序列的数据则可以从DELTA、OpenDELTA和Mesquite等软件导出为通用的Nexus和PyloXML(Stoltzfus *et al.*, 2012)数据格式，然后导入系统。系统发育树使用jsPhyloSVG (Javascript Libaray for Visualizing Interactive and Vector-based Phylogenetic Trees)插件，以Newick和Nexus格式，由用户上传类群的系统树并与其他分类数据进行整合。

4 结束语

互联网信息技术的发展对科学研究的影响深远。专著和植物志的编研是系统植物学领域最重要的科学实践。如果说早期Frodin(2001)对网络化植物志的编研还只是一个构想的话，现在，将分类学研究活动移动到互联网环境(Moving to Web)，实现世界在线植物志科学目标的诸多技术已日渐成熟。从大范围来看，分类学研究的范例(Paradigm: 科学研究的本质和目标)并没有发生根本的变化(Stuessy, 2009)，推动它不断向前发展的则是方法论范畴的不断变革和创新。也即，我们不断地尝试使用新方法来描述和验证我们对地球上生物有机体的观察和判断，并应用新的手段来组织和管理(分类)这些基础信息。因此，分类学研究与网络信息技术的结合是时代发展的必然。

分类学研究的继承性特点决定了任何新的研究成果总是需要追溯以往的事实和证据。从地理范围来说，这些事实和证据可能散布在世界各地的标本馆和图书馆，以不同的语言形式存在；从保存的方式来说，它们可能是标本馆的模式标本、图书馆的专著或期刊杂志中的文章，也或是数据库中的记录；从时间来说，它甚至可追溯到17世纪林奈时代的著作。因而，分类学研究所依赖的事实和证据高度片断化和低效重复利用的现状，及传统成果出版模式所导致的“信息孤岛”问题已是业界的共识(Godfray, 2002; Godfray *et al.*, 2007; Thomas, 2009; Parr *et al.*, 2012; Scotland & Wood, 2012)。而Web2.0互联网研究环境的建立被认为是这些问题的重要解决方案(Deans *et al.*, 2012; Leptin, 2012; Parr *et al.*, 2012; Marhold *et al.*, 2013)。

由于网络化提供了高效率低成本的成果发布

表3　Scratchpads 2.0中物种描述信息的分类标准(Species Profile Module, SPM)和说明

Table 3　Example standard of Species Profile Module implemeted in Scratchpads 2.0

序号 No.	主题术语 Items	解释和说明 Description
1	关系 Associations	描述捕食者–被捕食者; 寄主–寄生虫; 传粉, 共生, 互惠, 共栖, 杂交等生物相互关系的信息 Predator–prey; host–parasite, pollinators, symbiosis, mutualism, commensalism, hybridization…
2	行为 Behaviour	描述生物有机体对生物或非生物环境的反应 Cover actions and reactions of organism in relation to its biotic and abiotic environment
3	保护状况 Conservation status	描述物种目前或将来面临绝灭的可能性 A description of the likelihood of the species becoming extinct in the present day or in the near future
4	周期性 Cyclicity	描述物种周期性的特征, 比如物候 A state or condition characterised by regular repetition in time
5	细胞学 Cytology	描述物种的细胞生物学, 例如结构、功能等 Cell biology formation, structure and function of cells
6	特征摘要 Diagnostic description	描述物种与其相似物种的区别性特征 Distinguishing feature of this taxon from its closest relatives
7	疾病 Diseases	描述物种的疾病危害 Diseases of organisms
8	散布 Dispersal	描述物种的散布策略和机制 Dispersal strategies and mechanisms
9	分布 Distribution	描述物种按行政区域或生物地理区域界定的地理分布范围, 可以是全球性分布, 也可以是局部尺度上的分布情况。Cover ranges, e.g., a global range, or a narrower one; may be biogeographical, political or other (e.g., managed areas like conservencies); endemism; native or exotic; ref Darwin Core Geospatial extension.
10	演化 Evolution	描述物种的系统演化信息 Phylogenetic information relating to the taxon
11	一般描述 General description	广泛描述物种的综合信息 A comprehensive description of the characteristics of the taxon
12	遗传学 Genetics	有关物种的遗传信息, 包括染色体信息 Including karyotypes
13	生长 Growth	描述物种的生长速率、参数等 Rate; parameters; allometries.
14	生境 Habitat	描述物种的栖息环境, 包括地表分类(陆地、海洋)以及气候条件、忍耐幅、水平和垂直梯度上的特征 Include realm (e.g. Terrestrial etc) and climatic information (e.g. Boreal); also include requirements and tolerances; horizontal and vertical distribution.
15	立法 Legislation	描述有关该物种的立法情况 Legal regulations or statutes relating to the taxon
16	生活周期 Life cycle	专性发育转换 Obligatory developmental transformations
17	寿命 Life expectancy	描述物种的平均寿命 The average period an organism can be expected to survive
18	相似物种 Look alikes	描述与该物种相似的其他物种, 例如在入侵种群落中 Other taxa that this taxon may be confused with. Common in invasive species communities.
19	管理 Management	描述与物种的立法相关的管理情况, 比如CITES名录 A statement about the level of need to manage a taxon which can be related to a piece of legislation, e.g., a CITES list.
20	迁移 Migration	描述物种定期从一个地点移动到另外一个地点的情况, 例如动物繁殖期间的移动 Periodic movement of organisms from one locality to another (e.g., for breeding).
21	分子生物学 Molecular biology	描述物种的基因、蛋白质、生物化学(例如毒理) Include genomic, proteomic and biochemistry (e.g. toxicity)
22	形态学 Morphology	描述物种的形态特征(也包括解剖特征) The appearance of the taxon; e.g. habit; anatomy (the branch of morphology that deals with structure of animals)
23	生理学 Physiology	描述物种的生理学过程 An account of the physiological processes
24	居群生物学 Population biology	描述物种的丰度信息 Include abundance information
25	管理步骤 Procedures	如何处理某一类群, 已知的威胁是什么? Deal with how you go about managing this taxon; what are the known threats to this taxon?
26	生殖 Reproduction	描述物种的生殖策略、信号和限制条件 Reproduction cues, strategies, restraints.
27	风险评估 Risk statement	描述物种的入侵风险和影响 Include invasiveness, impact
28	体积 Size	描述物种的体积, 例如周长、长、体积、重量 Average size, max, range; type of size (perimeter, length, volume, weight ...)
29	分类群生物学 Taxon biology	描述分类群的生物学特征 An account of the biology of the taxon
30	威胁 Threats	描述物种面临的威胁 The threats to which this taxon is subject
31	趋势 Trends	描述物种的居群是否稳定? 是在增长还是下降等信息 An indication of whether a population is stable, or increasing or decreasing.
32	营养策略 Trophic strategy	描述物种在食物网中的位置和食物偏好 Include nutritional aspects, diet, position in food network.
33	用途 Uses	描述物种与人类利用的关系: 参考“经济植物” Relationships to humans; ref Cook “Economic Botany”

和传播机制, 它已逐渐成为研究工作必备的基础设施之一。被认为是系统植物学研究领域最严肃和高标准的专著也在考虑网络化的问题(Marhold *et al.*, 2013)。与此同时, 科学研究数据的大量积累使研究活动在时空广度以及深度上日益增加。强调“数据挖掘”(data-driven discovery)为基础的Big New Biology研究(Costello, 2009; Thessen & Patterson, 2011)要求科学家充分利用已有的数据在大的时间和空间尺度分析生命现象。这正如Smith (2009)曾指出的: “如果说网络信息技术是推动科学研究组织方式发生变革的引擎的话, 那么大量的数据信息将是引擎的燃料”。如何保证燃料的可持续产生和高效利用, 发挥它在科学研究中的最大价值? 首要的问题是我们如何用科学的方法来管理它们。从这个意义上说, 在生物多样性基础研究中网络信息技术代表的不是单纯的技术性工作, 它更代表着我们在大数据(Big Data)(Lynch, 2008; Mitch, 2008; 李国杰和程学旗, 2012; 盛杨燕和周涛, 2013)和大科学时代观念和思维方式的变革, 这或许是我们科研创新的原始动力和源泉。

S2系统的设计并不是简单地认为网络信息技术的应用是将已有的旧知识以改头换面的形式呈现给用户, 而是将研究者的实践活动放在网络大背景, 来有效解决数字化和实践研究的分离问题。过去研究者通常只关注论文, 而原始数据经常由其他数字化项目来完成。S2则将研究者从数据收集、管理、分析和发表的完整工作流程都放在网络背景下考虑, 将研究者的日常工作和数字化有机结合起来。它也有效解决了个体研究者在网络化中的角色和定位问题。因为任何S2的使用者都拥有最大权限来决定研究数据使用方式和机制。比如, 他可以决定系统的哪些数据可以公开, 哪些数据可以被其他研究项目和个人所利用, 而不再仅仅是过去数字化实践中的数据提供者和校验者。S2准确的角色定位、学科业务需求的深度挖掘和优越的技术实现, 决定了它能够在相对较短的时间获得大量用户认可。

参考文献

Agosti D (2003) Encyclopedia of life: Should species description equal gene sequence? *Trends in Ecology and Evolution*, **18**, 273.

Baker E, Rycroft S, Smith VS (2014) Linking multiple biodiversity informatics platforms with Darwin Core Archives. *Biodiversity Data Journal*, **2**, e1039. doi: 10.3897/BDJ.2.e1039.

Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB, Scotland RW (2010) Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences, USA*, **107**, 22169–22171.

Benson DA, Karsch-Mizra CI, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. *Nucleic Acids Research*, January; 39 (Database issue): D32–D37. Published online 2010 November 10. doi: 10.1093/nar/gkq1079.

Blagoderov V, Brake I, Georgiev T, Penev L, Roberts D, Rycroft S, Scott B, Agosti D, Catapano T, Smith VS (2010) Streamlining taxonomic publication: a working example with Scratchpads and ZooKeys. *ZooKeys*, **50**, 17–28.

Bowker GC (2000) Biodiversity datadiversity. *Social Studies of Science*, **30**, 643–683.

Brach AR, Boufford DE (2011) Why are we still producing paper floras? *Annals of the Missouri Botanical Garden*, **98**, 297–300.

Brake I, Duin D, Van de Velde I, Smith V, Rycroft S (2011) Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure. *ZooKeys*, **150**, 177–192.

Chavan V, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, **10**, S2.

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

Clark BR, Godfray HCJ, Kitching IJ, Mayo SJ, Scoble MJ (2009) Taxonomy as an e-Science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **367**, 953–966.

Costello MJ (2009) Motivating online publication of data. *BioScience*, **59**, 418–427.

Costello MJ, May RM, Stork NE (2013a) Can we name Earth's species before they go extinct? *Science*, **339**, 413–416.

Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE (2013b) Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology and Evolution*, **28**, 454–461.

Deans AR, Yoder MJ, Balhoff JP (2012) Time to change how we describe biodiversity. *Trends in Ecology and Evolution*, **27**, 78–84.

Duin D, van den Besselaar P (2011) Studying the effects of virtual biodiversity research infrastructures. *ZooKeys*, **150**, 193–210.

Erwin T, Stoev P, Georgiev T, Penev L (2011) ZooKeys 150: Three and a half years of innovative publishing and growth. *ZooKeys*, **150**, 5–14.

Fontaine B, van Achterberg K, Alonso-Zarazaga MA, Araujo R, Asche M, Aspöck H, Aspöck U, Audisio P, Aukema B, Bailly N, Balsamo M, Bank RA, Belfiore C, Bogdanowicz

W, Boxshall G, Burckhardt D, Chylarecki P, Deharveng L, Dubois A, Enghoff H, Fochetti R, Fontaine C, Gargominy O, Lopez MSG, Goujet D, Harvey MS, Heller K-G, van Helsdingen P, Hoch H, De Jong Y, Karsholt O, Los W, Magowski W, Massard JA, McInnes SJ, Mendes LF, Mey E, Michelsen V, Minelli A, Nafria JMN, van Nieukerken EJ, Pape T, De Prins W, Ramos M, Ricci C, Roselaar C, Rota E, Segers H, Timm T, van Tol J, Bouchet P (2012) New species in the Old World: Europe as a frontier in diodiversity exploration, a test bed for 21st century taxonomy. *PLoS ONE*, **7**, e36881.

Frodin DG (2001) *Guide to Standard Floras of the World*. Cambridge University Press, Cambridge.

Godfray HCJ (2002) Challenges for taxonomy: the discipline will have to reinvent itself if it is to survive and flourish. *Nature*, **417**, 17–19.

Godfray HCJ, Clark BR, Kitching IJ, Mayo SJ, Scoble MJ (2007) The web and the structure of taxonomy. *Systematic Biology*, **56**, 943–955.

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**, 313–321.

Johnson NF (2012) A collaborative, integrated and electronic future for taxonomy. *Invertebrate Systematics*, **25**, 471–475.

Knapp S (2008) Taxonomy as a team sport. In: *The New Taxonomy* (ed. Wheeler Q), pp. 33–53. CRC Press Taylor & Francis Group, Boca Raton, London, New York.

Kress WJ (2004) Paper floras: how long will they last? a review of flowering plants of the Neotropics. *American Journal of Botany*, **91**, 2124–2127.

Leptin M (2012) Open access—pass the buck. *Science*, **335**, 1279.

Li GJ (李国杰), Chen XQ (程学旗) (2012) Big Data: a significant strategic area in future science, technology, and economy development—research status and thinking of Big Data. Bulletin of the Chinese Academy of Sciences (中国科学院院刊), **27**, 11.

Lynch C (2008) Big data: How do your data grow? *Nature*, **455**, 28–29.

Maddison DR, Guralnick R, Hill A, Reysenbach A-L, McDade LA (2012) Ramping up biodiversity discovery via online quantum contributions. *Trends in Ecology and Evolution*, **27**, 72–77.

Marhold K, Stuessy T, Agababian M, Agosti D, Alford MH, Crespo A, Crisci JV, Dorr LJ, Ferencová Z, Frodin D, Geltman DV, Kilian N, Linder HP, Lohmann LG, Oberprieler C, Penev L, Smith GF, Thomas W, Tulig M, Turland N, Zhang XC (2013) The future of botanical monography: Report from an international workshop, 12–16 March 2012, Smolenice, Slovak Republic. *Taxon*, **62**, 4–20.

Mayo SJ (2008) Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis. *Kew Bulletin*, **63**, 1–16.

Mitch W (2008) Big data: Wikiomics. *Nature*, **455**, 22–25.

Moore W (2011) Biology needs cyberinfrastructure to facilitate specimen-level data acquisition for insects and other hyperdiverse groups. *ZooKeys*, **147**, 479–486.

Parr CS, Guralnick R, Cellinese N, Page RDM (2012) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology and Evolution*, **27**, 94–103.

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. *Trends in Ecology and Evolution*, **25**, 686–691.

Penev L, Kress WJ, Knapp S, Li DZ, Renner S (2010a) Fast, linked, and open—the future of taxonomic publishing for plants: launching the journal PhytoKeys. *PhytoKeys*, **1**, 1–14.

Penev L, Roberts D, Smith VS, Agosti D, Erwin T (2010b) Taxonomy shifts up a gear: new publishing tools to accelerate biodiversity research. *ZooKeys*, **50**, i–iv.

Penev L, Sharkey M, Erwin T, van Noort S, Buffington M, Seltmann K, Johnson N, Taylor M, Thompson F, Dallwitz M (2009) Data publication and dissemination of interactive keys under the open access model. *ZooKeys*, **21**, 1–17.

Platnick NI (2013) The information content of taxon names: a reply to de Queiroz and Donoghue. *Systematic Biology*, **62**, 175–176.

Scotland RW, Wood JRI (2012) Accelerating the pace of taxonomy. *Trends in Ecology and Evolution*, **27**, 415–416.

Sheng YY (盛杨燕), Zhou T (周涛) (2013) *Big Data Time: A Revolution that Will Transform How We Live, Work, and Think* (大数据时代: 生活、工作与思维的大变革). Zhejiang People's Publishing House. (in Chinese)

Smith V (2009) Data publication: towards a database of everything. *BMC Research Notes*, **2**, 113.

Smith V, Penev L (2011) Collaborative electronic infrastructures to accelerate taxonomic research. *ZooKeys*, **150**, 1–3.

Smith V, Rycroft S, Brake I, Scott B, Baker E, Livermore L, Blagoderov V, Roberts D (2011) Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys*, **150**, 53–70.

Smith V, Rycroft S, Harman K, Scott B, Roberts D (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. *BMC Bioinformatics*, **10**, S6.

Stoltzfus A, O'Meara B, Whitacre J, Mounce R, Gillespie E, Kumar S, Rosauer D, Vos R (2012) Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes*, **5**, 1–15.

Stuessy TF (2009) Paradigms in biological classification (1707–2007): Has anything really changed? *Taxon*, **58**, 68–76.

Thessen A, Patterson D (2011) Data issues in the life sciences. *ZooKeys*, **150**, 15–51.

Thomas C (2009) Biodiversity databases spread, prompting unification call. *Science*, **324**, 1632–1633.

Thomas WW, Forzza RC, Michelangeli FA, Giulietti AM, Leitman PM (2011) Large-scale monographs and floras: the

sum of local floristic research. *Plant Ecology and Diversity*, **5**, 217–223.

Vision T (2010) Open data and the social contract of scientific publishing. *BioScience*, **60**, 330–331.

Whitlock MC (2011) Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution*, **26**, 61–65.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: an evolving community-developed biodiversity data standard. *PLoS ONE*, **7**, e29715.

Winker K, Withrow JJ (2013) Natural history: small collections make a big impact. *Nature*, **493**, 480.

Yesson C (2007) How Global Is the Global Biodiversity Information Facility? *PLoS ONE*, **2**, e1124.

(责任编委: 马克平　责任编辑: 时意专)

附录 **Supplementary Material**

附图1　Scratchpads 2.0可视化的模块管理(A)和数据导入界面(B)

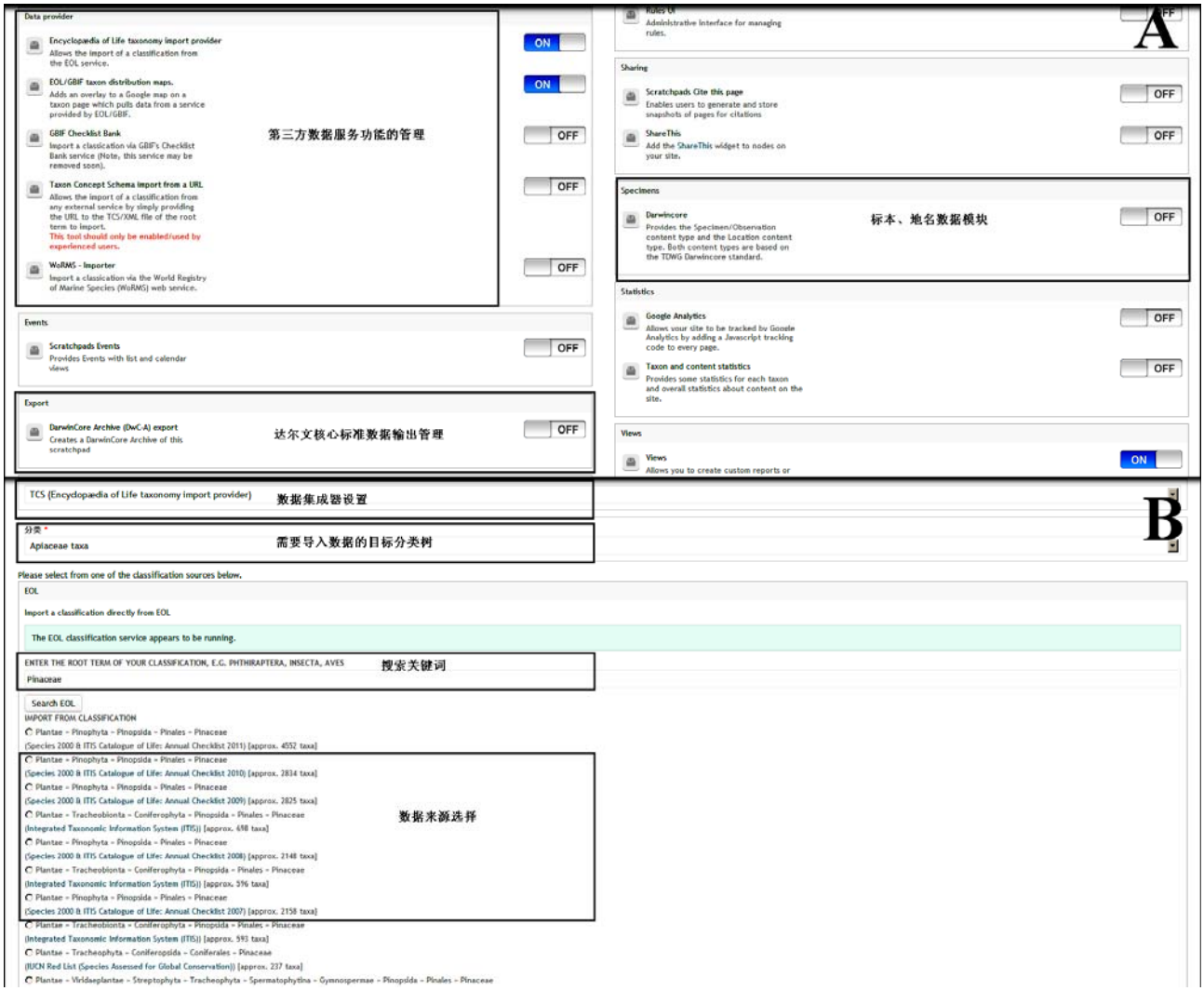
Fig. S1　Visualized interface of module management (A) and automatically taxonomic data import (B) in Scratchpads 2.0
http://www.biodiversity-science.net/fileup/PDF/w2014-012-1.pdf

附图2　Scratchpads 2.0可视化的分类术语编辑界面

Fig. S2　Visualized interface of taxonomic terms editor in Scratchpads 2.0
http://www.biodiversity-science.net/fileup/PDF/w2014-012-2.pdf

附图3　Scratchpads 2.0可视化的物种描述编辑(A)和内容管理(B)界面

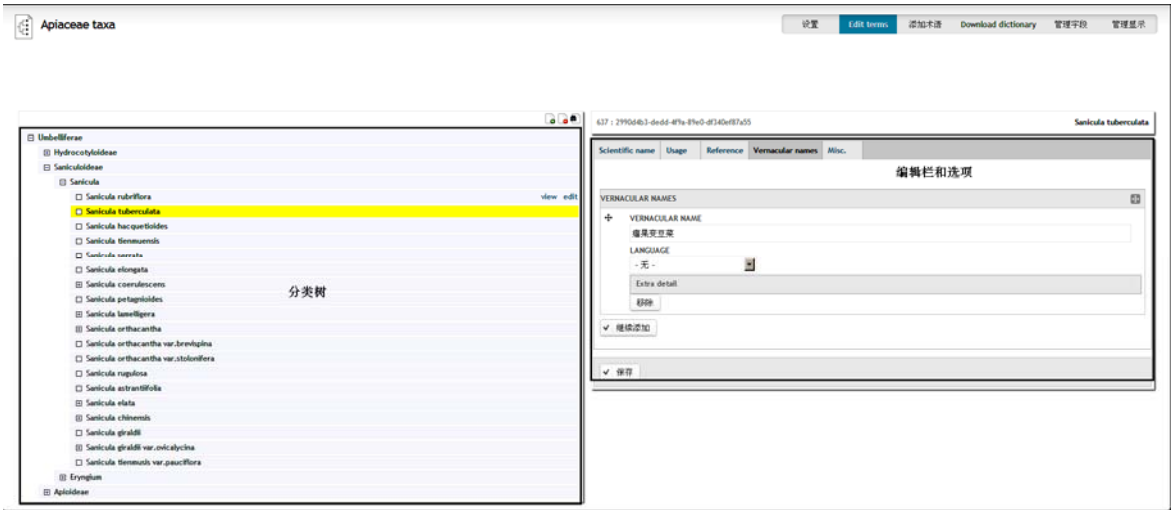
Fig. S3　Visulized interface of species description editor (A) and content management (B) in Scratchpads 2.0
http://www.biodiversity-science.net/fileup/PDF/w2014-012-3.pdf



附图1 Scratchpads 2.0可视化的模块管理(A)和数据导入界面(B)

Fig. S1 Visualized interface of module management (A) and automatically taxonomic data import (B) in Scratchpads 2.0

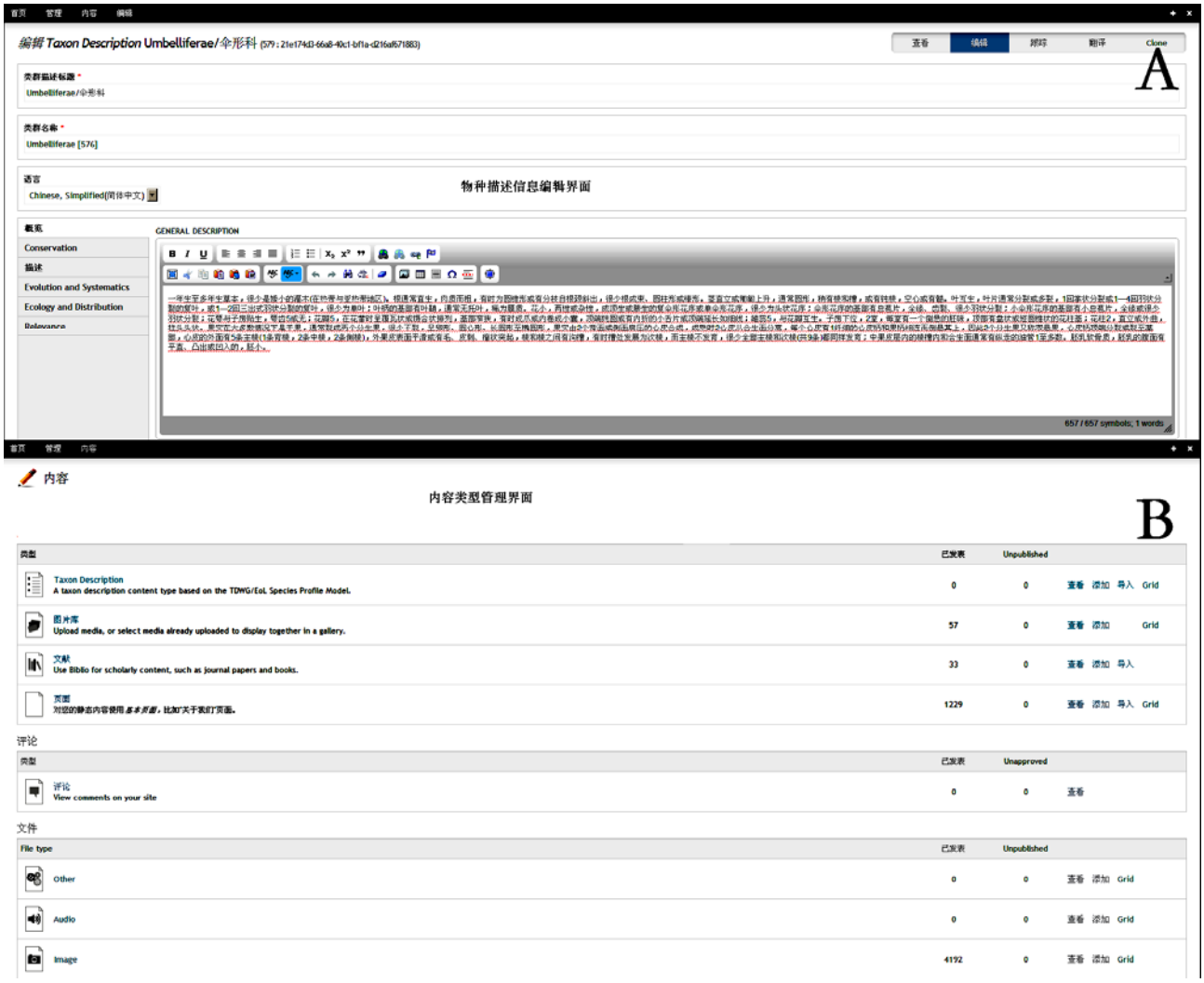
http://www.biodiversity-science.net/fileup/PDF/w2014-012-1.pdf



附图2　Scratchpads 2.0可视化的分类术语编辑界面

Fig. S2　Visualized interface of taxonomic terms editor in Scratchpads 2.0

http://www.biodiversity-science.net/fileup/PDF/w2014-012-2.pdf



附图3　Scratchpads 2.0可视化的物种描述编辑(A)和内容管理(B)界面

Fig. S3　Visulized interface of species description editor (A) and content management (B) in Scratchpads 2.0

http://www.biodiversity-science.net/fileup/PDF/w2014-012-3.pdf