PRO-iBIOSPHERE
WWW.PRO-iBIOSPHERE.EU

Coordination & policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability & Dissemination

SEVENTH FRAMEWORK PROGRAMME

Project Acronym:	**pro-iBiosphere**
Project Full Title:	**Coordination & policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability & Dissemination**
Grant Agreement:	**312848**
Project Duration:	**24 months (Sep. 2012 - Aug. 2014)**

D2.3 The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools)

Deliverable Status:	**Final**
File Name:	**pro-iBiosphere_WP2_NBGB_VFF_30082013.pdf**
Due Date:	**31 August 2013 (M12)**
Submission Date:	**30 August 2013 (M12)**
Dissemination Level:	**Public**
Task Leader:	**Quentin Groom (NBGB)**
Authors:	**Q.Groom (NBGB), D.Agosti (Plazi), A. Güntsch (FUB-BGBM), P. Hovenkamp (Naturalis), E. Kralt (Naturalis), D. Mietchen (MfN), A. Paton (RBGK), S. Sierra (Naturalis)**

European Commission

Consisting of:

Naturalis	Naturalis Biodiversity Centre	Netherlands
NBGB	Nationale Plantentuin van België	Belgium
FUB-BGBM	Freie Universität Berlin	Germany
Pensoft	Pensoft Publishers Ltd	Bulgaria
Sigma	Sigma Orionis	France
RBGK	The Royal Botanic Gardens Kew	United Kingdom
Plazi	Plazi	Switzerland
Museum für Naturkunde Berlin	Museum für Naturkunde Berlin	Germany

Revision Control

Version	Author	Date	Status
1.0	Quentin Groom (NBGB)	3 June 2013	Start of drafting
2.0	Quentin Groom (NBGB)	7 August 2013	First draft completed
3.0	Anton Güntsch (FUB-BGBM), Peter Hovenkamp (Naturalis), Eva Kralt (Naturalis), Daniel Mietchen (MfN), Soraya Sierra (Naturalis).	23 August 2013	Corrections and additions
4.0	Eva Kralt (Naturalis)	27 August 2013	Draft
5.0	Quentin Groom (NBGB), Soraya Sierra (Naturalis), Eva Kralt (Naturalis), Donat Agosti (Plazi)	28 August 2013	Draft
6.0	Quentin Groom (NBGB), Soraya Sierra (Naturalis), Donat Agosti (Plazi), Alan Paton (RBGK)	29 August 2013	Final Draft reviewed
6.1	Quentin Groom (NBGB)		Final Draft updated
FF			Converted to PDF

Table of Contents

Executive Summary

Objective

Taxonomists provide all the names and organisation for the life on Earth. The objective of this report is to summarise, from the perspective of taxonomists, the requirements and obstacles of a fully digital workflow for taxonomic publication.

Approach

In this document we examine the strengths and weaknesses of existing software solutions for taxonomic publication; obstacles to the adoption of these tools by taxonomists; and, some of the benefits to taxonomists who adopt a digital approach. The report is based on a workshop organised by the pro-iBiosphere project in May 2013 attended by about 100 people; a questionnaire distributed among taxonomists and related professionals, which received 220 responses; a literature review; and, direct conversations with taxonomists.

Results

There is considerable variety of biodiversity informatics software, much of it created by the taxonomists themselves for individual projects. Support for international standards for data fields and data exchanges is patchy, and there are few examples of a seamless flow of data from one system to another. Knowledge of biodiversity informatics systems among taxonomists is poor, and there is still considerable scepticism within the community regarding electronic publication and open access to literature. Nevertheless, there are examples where well designed software and strong community groups have combined to create exemplary databases of taxonomic information.

Recommendations

1. Focus on usability and interoperability of software, not just functionality.
2. Promote information technology within the taxonomic community.
3. Provide clear direct benefits in software for taxonomists, as well as downstream users.
4. Set a realistic minimum level of IT-literacy that is necessary to function as a professional taxonomist, and to incorporate that level into curricula and professional training.
5. Tackle the social obstacles to IT adoption, such as the disconnect between taxonomists and the users of taxonomy.
6. Stop using journal impact factors to assess the value of taxonomic works, but use measures such as book sales and web hits, which reflect the value of the work to users, rather than to other taxonomists.
7. Raise the profile of descriptive standards within taxonomy. For example, by creating translations for the Taxonomic Database Working Group (TDWG) world geographical scheme for recording plant distributions.
8. Mandate institutional data archival policies.
9. Research data has to be open access.
10. Ensure firm, long term commitment of institutions to digital taxonomic infrastructure.

Introduction

Open, comprehensive, authoritative and accessible biodiversity information is required for wildlife conservation, biological research and natural resource management. It is only through access to such information that effective evidence based policy can be formulated. An ideal vision of biodiversity informatics is where information is reliably available to the biological community and to the general public in such a way that it is discoverable, interlinked, both human and machine readable, and reusably licensed.

Biodiversity information originates from many disciplines, such as, ecology, conservation, biogeography, forestry, fishing, etc. However, central to all these is taxonomy (i.e., the science of naming, describing and classifying organisms. It includes all plants, animals and microorganisms of the world). Taxonomists define the conceptual units of biodiversity and, for the vast majority of species, their names are the gateway to scientific study of an organism.

Taxonomists have traditionally published the results of their work on paper. Indeed, until very recently, the international Codes of Zoological and Botanical Nomenclature demanded paper publication of new taxon descriptions (International Commission on Zoological Nomenclature, 1999; McNeill *et al*., 2012). Not only has this left us with a vast corpus of legacy literature, but it also means that the workflows and approaches to taxonomic research are all geared towards the publication of research on paper. Indeed, even though most new taxonomic publications are available in a digital format (usually PDF) in addition to paper, these digital documents are normally not semantically enhanced, annotatable or linked to other resources. Ironically, today's taxonomists store their data in digital formats, such as in spread sheets, only to print it on paper, from where it then needs to be re-digitised before adding to aggregators and other publications.

A different vision of taxonomic research is one where new taxonomic data are always stored digitally over the whole research cycle; from the beginning of projects, through to publication, then to reuse (Penev *et al*., 2010). New information will be given unique identifiers and any pre-existing data is assembled and referenced by its unique identifier. These identifiers will help provide the verifiability for users and acknowledgment for contributions. Publications will be generated from these data in semantically enhanced documents that are fully linked to the sources of their data. Platforms that can support such a vision of modern taxonomy include the Common Data Model (CDM) datastore and editor (wp5.e-taxonomy.eu), the Pensoft Writing Tool (www.pensoft.net/services-for-journals) and the Scratchpads system (scratchpads.eu). These platforms illustrate the possibilities of future taxonomy in collaborative data gathering, semantically enhanced publications and linked information in multiple media.

In fashioning an internet based vision of taxonomy, we need to understand both the user and creator communities of this knowledge. Many of the requirements for such a system are common to both communities, for example, they both want discoverable and free access to taxonomic literature. However, taxonomists also want funding, academic freedom and recognition for their work, as well as a "back-office" system in which work-in-progress can be stored confidentially. These requirements are not always compatible with user's demands for comprehensible, comprehensive, stable and up-to-date information. Balancing the demands of creators and users may be challenging, though in some respects this might be more easily achieved using digital media than on paper. For example, digital information can, in response to end-user specifications, be more easily repurposed and presented in different formats than can paper-based data.

The user community's requirements for taxonomic information has been reported in depth in the report on user feedback, pro-iBiosphere deliverable 2.2. Here, we will concentrate on the requirements of taxonomists and other stakeholders that create taxonomic knowledge. Yet, it is important to acknowledge that there are end users of taxonomy whose requirements are important if taxonomy is going to remain relevant.

In order to put into practice the vision of a fully integrated biodiversity informatics environment, taxonomy needs to operate in a setting that allows the development of such an environment. It requires much openness of publication and it will also require changes in the way taxonomists are evaluated and how they are trained.

In February 2013, pro-iBiosphere organised a two day workshop in Leiden on "e-platforms & e-tools for taxonomy". The first day consisted of talks and demonstrations of a wide range of digital tools for taxonomists. The second day consisted of training, where four parallel sessions were run, giving hands on experience on a number of these technologies. The purpose of the workshop was to promote digital taxonomic tools amongst taxonomists and to bring taxonomists and software producers together to allow an interchange of ideas and experience.

About eighty scientists from six continents attended, including taxonomists of botany, zoology and mycology. Their knowledge of information technology also varied. Some were being introduced to these technologists for the first time, others were long time users, while still others designed and built these systems. This mix of technical aptitudes was important to encourage reciprocal knowledge exchange from the producers and users of the software. It is important to foster debate on these technologies so that all the stakeholders in the international taxonomic community have a common vision of the future and a voice in its development.

A questionnaire was organised around this event on taxonomist's use of digital technologies. Details of the questionnaire have been published in deliverable D3.1 (Towards a Best Practices Guide on editorial policies, section 1.1.) The questionnaire was circulated to the workshop participants, and to several major European taxonomic institutions and taxonomic newsgroups. We received 220 responses from taxonomists and associated professions from thirty-three countries (see Appendix 1).

Information from the questionnaire was also supplemented with information from interviews that were conducted at the Royal Museum for Central Africa, Belgium and the Royal Belgian Institute of Natural Sciences and group discussions at the National Botanic Garden of Belgium. Useful information was also gathered from the participants at the pro-iBiosphere workshops and from scientific social media (e.g., www.researchgate.net).

User acceptance of information technology can be broken down into two elements, i.e., perceived usefulness and perceived ease of use (Davis, 1989). Here, we will report these two elements from the perspective of the taxonomic community. How usefulness is, at present, perceived; what advantages software brings; how it could be improved; and, what limitations exist. Ease of use varies widely between different IT systems. In this report, we focus on the general perception of the community and consider general improvements that all systems could make to overcome the reluctance among taxonomists to use such tools.

Taxonomists often debate the causes and consequences of the so-called taxonomic impediment (e.g., Godfray *et al*., 2007). That is the slowness and inefficiency of taxonomic research, compared with the large number of undescribed species believed to exist. It is as yet unclear how information technology will help resolve this impediment; however, information technology has already had an impact on accessibility and discovery of literature; communication between taxonomists; the speed of publication; and, data gathering. Yet, in all these areas there is room for improvement.

In this report we examine the incentives to and attitude of taxonomists towards further digitisation of their profession. Emphasis is given to, the (i) advantages and disadvantages taxonomists perceive to digitisation; (ii) practical aspects preventing further digitisation; (iii) possible solutions to tackle these challenges; (iv) aspects of open access to scientific literature and to data; (v) availability of software tools; (vi) taxonomic collaboration; and (vii) uptake of software by the community.

Taxonomist's Attitude Towards Open Access

Open Access to Literature

The internet has caused a paradigm shift in the way we want to access scientific literature. It is widely recognised that open access to published science promotes scientific discovery and innovation, but also fosters trust and understanding of science among the general public.

There have been many recent developments in the area of open access publishing, and it is interesting to appreciate the perspective of copyright and open licensing from a taxonomist's perspective. We asked respondents to our questionnaire "*What sort of value do you gain from restricting reuse of you publications?*" (Fig. 1). The fact that only a quarter of the respondents completed this question suggests that there is little awareness of these issues in the community. The scope of copyrights and the meaning of Creative Commons licensing schemes is not always clear to scientists.

Many respondents seemed to believe that copyrights and closed access schemes helped them ensure citation of their work and data. There seems little recognition that the data reported in publications is not copyrightable (Agosti & Egloff, 2009). There were also a substantial number of respondents who mentioned that they did not want anyone else to use their data before they did. Clearly, these responses are not indicative of a culture of openness and sharing.

Revenue from publications did not seem important to the majority of people and probably only relates to those respondents publishing large, bound texts. Certainly, the revenue generated from such volumes can never cover the complete costs of writing and research. However, they do usually cover the publication costs.

The attitude of taxonomists from developing countries to open access was rather ambivalent: they appreciated the greater availability of taxonomic literature, but were worried that the cost would be too high. There seemed little or no awareness of either journal fee waivers for developing countries or low and no-cost options. They felt that they were unlikely to publish in open access journals themselves, particularly when many of these journals have either a low or no impact factor. Developments in open access journals are changing rapidly and it is likely that the rather negative opinion towards open access will lessen as more high quality journals adopt open access and prices fall.

The reasons that respondents to the questionnaire gave for choosing closed licensing for the publication of their research are summarised in Figure 1. Respondents were asked "*What sort of value do you gain from restricting reuse of you publications*?". Only forty-six (out of 220) respondents answered this question.

Figure 1. Reasons given by taxonomists for choosing closed licensing for the publication of their research.

Open Access to Data

One person's literature are another person's data. Taxonomists may want to give people free access to their publications to get maximum exposure of their work. However, they are often facing conflicting interests. One is that when they give away the data they will no longer be appropriately credited for their work. This challenge is particularly acute in taxonomy, as many data users are not aware of the amount of work that is involved in collating reliable and useful data in many taxa (e.g., by providing georeferences and identifications). It is, therefore, unrealistic to expect taxonomists to provide full open access to their data without some way of negotiating a proper recognition for their contribution. On the other hand, openness and the verifiability of science are essential to good scientific practice. Scientists should be able to back up their conclusions with the data that supports them.

As yet, the performance of taxonomists is largely assessed by their published output, not by their creation of data. In our questionnaire, the majority of respondents (>70%) were evaluated on their publication record in some form. In the case of alpha taxonomists[1], 94% are evaluated by their publication output. Many (46%) taxonomists are assessed by the impact factors of the journals they publish in, which, though increasingly discredited as an index of quality of science (e.g., Alberts, 2013), still appears to be an important metric for the evaluation of scientists (Campbell, 2008; Simons 2008). Until recently; the publication and the data underlying it were not seen as separate entities. Data were gathered for a specific reason and once the results were published, little thought was given to data reuse or even data archiving. Unless data are considered separately from the publications, credit cannot be given for its creation. Initiatives such as Dryad (datadryad.org), Figshare (figshare.com) and the Biodiversity Data Journal (biodiversitydatajournal.com) all aim to provide avenues through which scientists can publish their data in a citable manner, yet these initiatives are quite recent, and there is no evidence yet of rapid adoption by the taxonomic community. Yet, these publication channels may be an opportunity for many taxonomists and parataxonomists[2] to leverage the work they have done in data creation.

Nevertheless, it is clear from conversations with scientists that many get prestige from publishing in high impact factor journals and that they prioritise phylogenetic and ecological studies above purely taxonomic works, because they can publish these in high impact factor journals. Therefore, it seems likely that, even if they gained credit for the sharing of data, it would not be a strong incentive to openness.

Scientists can only be given credit for their data if those data are traceable. While most modern publications have a unique identifier (usually a DOI); there is not yet an equivalent for observations and specimens. This topic is at present being discussed

[1] Alpha taxonomists are those persons specifically involved in the description of organisms, rather than being involved in the related disciplines such as phylogenetics.

[2] Parataxonomists work alongside taxonomists in the collection, preservation, cataloging and study of biodiversity, but are generally less qualified to do the more formal scientific taxonomy. Nevertheless, they often have considerable skill in the identification and discovery of specimens.

as part of pro-iBiosphere's activities (see Best practices for stable URIs). Furthermore, in cases where researchers are aggregating millions of observations in their research, there is no feasible mechanism to credit all of the contributors of those data.

Large scale infrastructural repositories could help by providing novel metrics for the use of provider's data. Yet there is a trade-off between the costs of implementing such a system and the benefits to providers. Certainly, it is unlikely that it would be possible any time soon to have such fine granularity of contribution statistics that individual researchers could receive a citation index for their contribution to mass aggregators such as Global Biodiversity Information Facility (GBIF). Nevertheless, if institutions gained benefit from contributing their data, those institutions may find ways to credit and reward their contributors internally.

Access to scientific data and data reuse is a recent issue that has received growing attention as large data aggregators have developed. One of the deliverables of pro-iBiosphere is a report on "*D2.4.1 Draft policy on Open Access for data and information*". Few of the institutions we contacted during the writing of this report have defined data policies dealing with these issues, though it is clear that such policies are needed at both institutional and national levels. Such policies should give clear guidance to scientists on their attitude to data archiving, openness and standards and should outline clearly the responsibilities involved. In July 2013 the pro-iBiosphere project attended the European Commission's Public consultation on open research data. It was clear that projects proposals containing data management plans on Open Access might be required in the near future. A recent press release from the commission stated "The Commission will make open access to scientific publications a general principle of Horizon 2020. As of 2014, all articles produced with funding from Horizon 2020 will have to be accessible".

The internet has also broadened our view of where data can come from. Technology now enables us to "scrape" data from websites and publications. Mashups can be created from disparate sources of information, providing novel views of the world and unexpected results. At the pro-iBiosphere workshops in Leiden 2013, a presentation on the "mining of phylogenetic trees by extracting data from the figures in scientific publications" was given by one of the participants (Ross Mounce). Such kind of data-mining has great potential for efficiently reusing scientific data. Still, our ability to do this with the corpus of scientific text is severely restricted both by the limited access to literature and by the available formats. Even though the data in publications is not protected by copyright, the prevalence of subscription-based publications means that it is often not possible to access such data without considerable payments to the publishers (Hagedorn, 2011, Poisot, Mounce & Gravel, 2012,).

Conclusions

- Taxonomists need a better understanding of copyrights and how it influences their work and scientific impact.
- Taxonomists need to increase their awareness of the value of their work outside their own field.
- Taxonomists need convincing arguments that open access is beneficial for science and ultimately for them.

- Developing systems that can give credit to the sources of biodiversity data will encourage openness, but one should not overstate the degree of incentive that this provides.
- All institutions and large projects should develop data policies, which outline the responsibilities of scientists towards data.

Status of Biodiversity Informatics Resources for Taxonomic Publication

The pro-iBiosphere *deliverable (D2.1.1) on "On-going Biodiversity Related Projects, Current e-Infrastructures and Standards"* gives status reports on most of the leading IT projects related to taxonomy (Agosti *et al.*, 2013). However, here we will focus on the status of some of these platforms of relevance to the creation of taxonomic information by taxonomists.

Biowikifarm

Biowikifarm is a wiki-style system based upon mediawiki installations. Its user niche is in the creation of unstructured or loosely structured webpages that can be edited frequently and easily. The MediaWiki approach is already known to many users and this system has good long-term support and active development.

All sorts of useful biodiversity data is produced in an unstructured format, even though it may be worked into a more structured format later and Biowikifarms provides a resource where these data can be developed.

As an example, Biowikifarms is used by pro-iBiosphere because it allows members of the whole consortium to contribute, view and edit pages on our meetings and projects. However, the system is very versatile and can be used for the creation of identification keys, writing books, checklists, glossaries and online floras.

Pensoft Publishing Tool

The Pensoft publishing tool aims to be "*a collaborative article authoring and publishing platform*". Users are guided through the process of creating a peer reviewed publication on a single online platform.
Users are presented with a template for publishing and can use the platform collaboratively with their co-authors. By structuring the process of documentation from the beginning of the writing processing the procedure of semantic mark-up and linking is automated and controlled by the author. Upon publication, data are automatically share with aggregating database. For example occurrence data is shared with GBIF and taxon treatments to PLAZI and EOL.

Taxonomic Editor and Common Data Model

The Taxonomic Editor is an interface to highly atomised databases of taxonomic information implemented as a CDM server. This system has been developed at the Department of Research and Biodiversity Informatics, Botanischer Garten und Botanisches Museum Berlin-Dahlem (FUB-BGBM). It allows users to create highly detailed accounts of taxa, nomenclature, bibliographies etc. The data in the CDM store can either be imported from XML or Darwin Core Archive formats or can be added directly to the CDM store using the Taxonomic Editor. This editor allows users to browse taxonomies, rearrange taxonomies, edit synonyms and add information of many different kinds on the nomenclature, traits, phenology, distributions, etc.

Data from CDM stores can then be published on the internet using a Drupal portal, or published to paper. The format of the printed output is specified by Extensible Stylesheet Language Transformations (XSLT) transformation of XML output of the CDM. This creates XSL-FO documents that can be converted directly to PDF.

Linnaeus II

Linnaeus II is a desktop application for the preparation and publication of taxonomic information. This application is produced by ETI Bioinformatics (www.eti.uva.nl), which was founded in 1990 as an independent organisation, but has now merged with Naturalis Biodiversity Centre (www.naturalis.nl). Although outputs from their system can be published on paper, they are best-known for the production of CD-ROMs, websites and smartphone apps for the identification of biota. Their CD-ROM catalogue currently lists 113 different titles covering a wide variety of different topics, including the identification of plants, fungi, mammals and even the dead bird remains for the accurate determination of birds involved in aircraft collisions.

The Linnaeus II software is free, but not open source. ETI is funded through various projects, sales of its products and through paid application support. Their merger with Naturalis indicates that their products are sustainable. Though it is not clear how this may influence their priorities.

ETI are soon to release a new product, Linnaeus Next Generation. This is an internet based system to eventually replace Linnaeus II. Users can collaboratively write a publication through a browser based application. This is intended to produce the same products as Linnaeus II, such as books, websites, CD-ROMs and apps. However, there are many advantages for users to working online, including collaboration, central maintenance and storage, connections to web services and real time updates of websites. Linnaeus' data are stored in a mysql database, but the data can be exchanged with the CDM and Scratchpad systems.

Lucid

Lucid is a product of QBIT, based at the Queensland Alliance for Agriculture and Food Innovation Institute (www.cbit.uq.edu.au). Lucid focuses on the aspect of identification enabling users to create expert systems for the identification of biota, although its use is not restricted to biota. Lucid keys are made for online and internet distribution and there are over 200 available. For example there is an online key to the commonly cultivated palms (itp.Lucidcentral.org/id/palms/palm-id).

Lucid is a commercial, closed source product. A single license is sold for $599.00 (€426 on 19/06/2013). It is implemented in Java, so can be used on all operating systems. Lucid can import and export data in Software Design Description (SDD) format for exchange with other software systems.

Lucid's users are predominantly based in Australia and the USA.

Scratchpads

Scratchpads are an internet based content management system based upon the Drupal framework, but tailored to the needs of the biodiversity community. Users can create Scratchpads along any theme in biodiversity, and so far, more than 500 sites have been created. Some features that make them popular among taxonomist are ease of use, good application support, versatility, considerable investment in training and a low initial investment (they are free to the creator and user).

The Scratchpads team is in the process of migrating its sites from the original version to Scratchpads 2. This upgraded system incorporates the change from Drupal 6 to Drupal 7 and enhances the front-end usability and backend scalability. It is important to the long term sustainability of platforms that they are kept up-to-date with the latest enhancements, even though this can be a time consuming and expensive process. Nevertheless, maintenance is made easier in the case of Scratchpads, because of the broad user base of Drupal that is far larger than the Scratchpads community.

Multi-access Keys

Software systems for the creation of multi-access keys have the longest history of any IT system for taxonomists. The Pankey and Delta systems for the creation of digital keys date from the 1970s (Pankhurst, 1971; Watson & Milne, 1972). They are closed source, even though Delta was free to download. Pankey remained a DOS program and Delta's development halted in the 1990s, even though both were in regular use and there was demand for development. Finally, in 2011, the Open-Delta project was funded by the Atlas of Living Australia to create an open source version of the Delta system (code.google.com/p/open-delta/). Meanwhile, numerous alternative systems for creating digital keys have been created, including Navikey (www.navikey.net), XPER2 (lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2), Key to Nature (www.keytonature.org.uk), and Lucid (www.Lucidcentral.com). These include open and closed source programs, some are free, some use international standards, others are proprietary. This diversity of options is good for competition in the field, but could lead to indecision. Several respondents to the questionnaire, including taxonomists interviewed mentioned that the range of choices for software systems, can be bewildering for them, and hence, this can make challenging for them to make a choice on what system to use. There is also a reluctance to learn how to use new software systems and a perception that systems are constantly changing, so that the investment in learning is not rewarded.

It is challenging to tell which model of software sustainability will be most successful. Open source solutions in this field have not been around for a long time and closed source solutions have a patchy record for sustainability.

Bespoke Systems [FishBase]

Many researchers and institutions choose to create their own IT tools for taxonomy. In our questionnaire, 36% of respondents answered yes to the question "*Have you or your institution created your own e-tools for taxonomy?*". The reported reasons for this follow three main themes:

- No suitable tools were available when they started their project.
- Other tools do not fit their needs and they want to be able to customise their own solutions.
- They valued their independence and did not want to rely on external support.

Therefore, taxonomists perceive several advantages in creating their own software. Nevertheless, there are several disadvantages for the taxonomic community as a whole:

- A lack of common standards for data formats and data exchange.
- Wasted effort as similar solutions are created in parallel.
- There is no competition among solutions to stimulate the good ones and weed out the bad ones.

Nevertheless, bespoke systems have low initial costs in terms of time and resources and they can adapt quickly to changed priorities. An example of a successful and evolving bespoke system is FishBase.

FishBase is an online authoritative, comprehensive resource on fish. FishBase is run by a consortium of nine research institutions in Europe, Asia and North America. FishBase uses uncomplicated open source technology to present its database online, i.e., Hypertext Preprocessor (PHP) and MySQL. These are widely used, popular systems and easy to program.

In the taxonomic workflow, FishBase supports data aggregation and discovery. It has been successful in uniting the global community of ichthyologists and creating an authoritative resource. It has done this through the cooperation and long term commitment of the major interested institutions.

Part of FishBase's success is the strong community of ichthyologists that use and contribute to the database. In the period Jan 2012 - March 2013 it received 355,000 unique visitors to its website each month and had over 2,000 contributors.

The strong global community backing FishBase is reflected by its multilingual interface that has been translated into 19 languages, including Chinese, Hindi, Russian, Farsi and Arabic. This support for many languages is particularly important for fish where vernacular names are widely used. For the 32,500 species in FishBase they have collected 299,700 common names, all with a supporting reference.

FishBase encourages users to cite the primary sources of information within FishBase. Therefore, it is an important tool for the discovery of information. Yet owing to this policy, its citation rate will not reflect its utility to researchers. To prove its value, it has to rely on website statistics. A downside of this is, that if other websites use FishBase information under the CC-BY-NC licence, FishBase will lose page hits that it would otherwise receive, thereby undermining the justification for its *modus vivendi.* This is a common challenge for aggregating websites. They may want to act as backbone resource providing authoritative support to other more specialised websites, but when they do this, it can conflict with their need for a prominent web presence.

FishBase could be stronger on standards. Data can be exported to XML, but not to any international standards. FishBase maintains its own separate taxonomy, which largely follows the taxonomy of the Catalogue of Fishes, but is not tied to it. Consensus of the taxonomy is large made on pragmatic grounds, though they chiefly follow the latest revisions of particular groups. Literature references are only available as character strings, rather than full atomised records. Specimen and observation information is sourced directly from various institutional databases or indirectly from GBIF. To avoid duplication, institutions with their data on GBIF do not also provide their data directly. There are no standard unique identifiers for specimens or observations, though catalogue numbers from each institution are preserved.

Conclusions

- Marketing of software is needed to explain the uses and advantages of the multitude of software options for taxonomists.
- Bespoke systems have the advantage of being rapid, cheap and adaptable, but at the expense of standardisation.
- Support and cooperation of the community is important for the long term usage and sustainability of systems.

Advantages and Disadvantages of Digitisation to Potential Users of Digital Tools for Taxonomy

…To Alpha Taxonomists

The primary role of alpha taxonomists is the collection and dissemination of information about biodiversity, particularly the recognition of and the distinction between different taxa. They are potentially important users of these tools, both as creators and consumers of digital information. Taxonomists will benefit from a fully digitised workflow in many ways, for example:

- Greater exposure of their work.
- Facilitated citability and better recognition.
- Greater accessibility to their own data, which will be repurposable.
- Improved networking with other taxonomists and users of their publications.
- Easier access to other open data, such as bibliographies, biodiversity observations and specimen images.

Yet the route towards fully digitised biodiversity information is not a simple one, and to understand the technical and social aspects of working with digital taxonomic tools, it is informative to reflect on particular usage cases. For example, The *Nouvelle Flore de Belgique* has been published by the National Botanic Garden of Belgium since 1973. Its 6th edition has recently been published, and the Garden is planning the 7th edition. However, the authors of the 6th edition have retired, and a new team has been put in place to develop the next edition. This new edition will have many changes, including new taxonomic treatments, new keys, illustrations and a completely revised taxonomy to bring it up-to-date with the consensus of the Angiosperm Phylogeny Group.

The Garden and authors want a digital system for writing their Flora; however, they have many requirements for such a system. The authors want a system that allows:

- Working collaboratively and simultaneously. Some authors work in different institutions, and some of them may prefer to work from home.
- The ability to import the previous edition of the Flora into the system as a basis to work from.
- The addition of annotations (comments) to the document. These annotations would not be published, but will help the editors track the reasons why text was changed.
- The simple taxonomic rearrangement and renaming of taxa.
- The parallel writing of a bilingual Flora, preferable where the two versions can be seen next to each other for ease of translation.
- Able to produce a high quality printed output.
- The simple reuse of data for subsequent editions and potentially websites and apps.
- The automatic validation of data.
- The system would keep previous versions of the documents for future reference.

Given these demands, various proposals were submitted to and by the authors as possible solutions. Table 1 gives an overview of these solutions and their ability to fulfil the requirements of the authors. Clearly, no solution completely satisfies all the requirements of the users.

There are features of many IT platforms that, at this time, were *not* thought of as important by the authors of the Flora. These included:

- The ability to have a linked website that can publish taxonomic treatments as soon as they were prepared.
- The automatic linking of treatments either to specimens in our virtual herbarium, to observational data (e.g., GBIF), to nomenclatural data (e.g., The International Plant Names Index) or to bibliographical references (e.g., Biodiversity Heritage Library, Citebank).
- The federation of content from providers such as the Encyclopaedia of Life.

- The ability to create matrix keys, rather than standard dichotomous keys.
- Publishing under a Creative Commons Licence.
- Comments and feedback from users.
- The use of digital media such as videos and audio, except perhaps for a small quantity of photographs.
- Character matrices and automated structured descriptions.
- Forums and blogs.

Comparing the requirements of the authors and the solution provided by the providers of digital taxonomic systems, there is clearly a dichotomy. On the one hand, authors are fixed into a traditional printed media publishing model where copyright restricts reuse, whereas the system providers have an open model where the authors build upon available digital content and their additional content is in turn reused by the community. This difference in approach needs bridging if we are to build an integrated Open Biodiversity Knowledge System (i.e., i-Biosphere).

From meetings with the authors, it is clear that there are several reasons for their preference for a traditional print model for their Flora.

- Print provides a long lasting legacy for the authors.
- For the Botanic Garden, the Flora is a flagship product and a website is not perceived with such regard.
- Print is cheaper than the development and long term maintenance of a website.
- Book sales provide income for the Garden.
- Books are more robust and reliable when used in the field.
- Many users of floras are not IT literate.
- There is a reluctance to learn how to use new software systems and a perception that systems are constantly changing, so that the investment in learning is not rewarded.

This case study exemplifies many, if not most, of the difficulties of encouraging users to adopt digital media for the creation and sharing of taxonomic knowledge. These difficulties include the lack of a funding model for internet publication, the IT fluency of taxonomists and the recognition of contribution.

Table 1. Trade-offs among envisaged software/platforms for creating the new Flora of Belgium

Possible solution	Suitable for collaborative works	Importing original text	Annotations	Renaming and rearrangement of taxonomy	Bilingual writing	Printed output	Repurposing of data	Automatic validation	Versioning	Cost
Word processor (e.g., Microsoft Word)	Poor	Simple	Easy	Difficult	Possible with two windows open	Easy	No	No	Only manually	Moderate
Google Docs	Good	Simple	Easy	Difficult	Possible with two windows open	Easy	No	No	Only manually	Low
Scratchpads	Good	Moderate difficulty	Only as comments at the end of a page	Simple	Yes	Perhaps possible with an export	Yes	To some extent	Yes	Low
CDM Editor	Good	Moderate difficulty	Only as attached notes	Simple	To some extent	Yes	Yes	To some extent	No	Low

Linnaeus	Good	Moderate difficulty	Only as attached notes	Simple	To some extent	Yes	Yes	To some extent	No	Moderate
Biowikifarm	Good	Moderate difficulty	Possible as footnotes	Fairly simple	Possibly with two windows open	Perhaps by using mediaWiki extensions	To some extent	No	Yes	Low
Custom Database	Good	Moderate difficulty	Not without considerable programming	Fairly simple	Yes	Easy	Perhaps	Perhaps	Not without considerable investment in programming	High

…To Ecologists and Conservationists

Ecologists and conservationist are large consumers of taxonomic information, and creators of biodiversity information and knowledge on distributions, populations, behaviour and habitat. They require information that is authoritative, citable and available in an easily digestible format.

Their work often combines information they have gathered together locally, with more general information from which they draw conclusions about the biota worthy of conservation and the habitats of their region.

They require high quality, up-to-date data and information to inform conservation assessments, management plans, and other measures. For example, the International Union for Conservation of Nature (IUCN) Red List assessment procedure requires accurate demarcation of a taxon's circumscription; information on the past and current distribution, population estimates and an assessment of threats (IUCN, 2013). The red list assessor may collect much of the current information, but the taxonomic literature is a very important source of information.

The roles of taxonomists, ecologists and conservationists are not clearly separable and the data from each discipline percolates into the others. Taxonomists (i) use the field knowledge and observations of ecologists and conservationists and (ii) provide identification skills to ecologists and conservationists. Therefore, the use of and production of digital biodiversity data by ecologists and conservationists has great value.

…To Citizen Scientists & Parataxonomists

Citizen scientists & parataxonomists provide much of the biodiversity data that is at present being collected in developed countries. This is particularly true for information on large and charismatic flora and fauna, but is also true of many small, obscure groups. Use of information technology among citizen scientists varies enormously, from those who use sophisticated databases and networked information systems to those whose work is completely paper-based. Their use of information technology tools is nevertheless growing rapidly.

There are numerous examples of projects engaging the public in biodiversity monitoring and data collection. There are too many to list here, but specific examples relevant to taxonomy are the transcription of herbarium specimen information (e.g., Herbarium@Home and Les Herbonautes); collection of observations (iSpot, waarneming.nl, telmee.nl, Waarnemingen); the transcription of field notebooks (Wikisource) and in the collection of phenotypic data (Evolution Mega Lab).

Most citizen scientists are unlikely to be users of digital tools for taxonomy, but a small subset conducts research and publishes to a professional standard. These self-financed, self-motivated parataxonomists make up a large contribution to biodiversity informatics; examples include the Flora of Zimbabwe, The Flora of North-east England and the Verspreidingsatlas. Such researchers could be important users of these technologies; however, they are self-funded and value their independence. They would benefit from a greater standardisation of systems and conformation to standards and in return, science would benefit from their expert knowledge.

A shining example of a successful IT system among amateur naturalists is the program Mapmate® from the UK. This program sells for about €37 and is closed source, it allows users to enter simple biodiversity records and map them. Its simplicity certainly contributes to its success, but its unique feature is that users can synchronise their database with each other. This is not just an import/export system, but a true synchronisation of data that allows disparate users to update their own databases with other people's corrections, additions and deletions. Approximately 15,000 licences have been sold, of which about 10,000 are currently active. As of March 2013, it is estimated that 1 billion record exchanges have occurred through the Mapmate system.

Many large non-governmental organisations (NGOs) in the UK use Mapmate, including the Royal Society for the Protection of Birds and the Botanical Society of the British Isles. These societies maintain large central hubs of Mapmate data, which enable greater access nationally and secure backup for the users. The Mapmate hub of the Botanical Society of the British Isles contains 13,733,706 records as of July 2013. While the success of this simple platform is undoubted, the system does not exchange data in any agreed standard format and neither do its taxa lists agree with any international standard. What makes this program so successful is its usefulness and ease of use to the users.

An unforeseen consequence of the synchronisation system of Mapmate, is that it has facilitated openness in the biodiversity community who now regularly exchange records. An advantage of a synchronisation system, as opposed to a common database is that users have the choice to be open with their data, and data flow can be uni- or bidirectional. Therefore, users have the security of "owning" their own data and some control of which data they share. The users themselves, in general, are most concerned with their local 'patch', where they do their data collection, but the system allows data users at a national and global level to make use of their data (Ellis, Waterton & Pacha, 2007).

While the distributional data that Mapmate stores is only peripheral to the field of taxonomy, its success is a model of how simple technologies can transform the way people gather data and collaborate. Its use of synchronisation, rather than import-export, is something that could be used more widely in the field of biodiversity data.

…In Developing Countries

Wherever you are in the world, internet based tools for taxonomic publications have the same potential for cheap, collaborative, integrated publication. What makes these tools particularly relevant in the developing world is that there are few alternatives. Availability of basic taxonomic literature is scant where such literature exists, but often comprehensive treatments are unavailable for many groups. For example, the Republic of the Congo has no flora whatsoever and the flora of the neighbouring Democratic Republic of Congo is only ¾ complete. Even in countries where taxonomic works exist, they are often large bulky tomes that are restricted to a few libraries. It is not that information on the plants of these countries does not exist. However, it is widely scattered in many diverse publications in multiple languages and using different taxonomies.

Internet based systems are much more suited to access in these regions. Even though internet access can be erratic and electronic devices can be unreliable, a paper-based solution is not an option. There is also the problem of accessibility to collections, which are mainly located in the northern hemisphere and particularly in Europe. Providing digital images and data from these herbaria is the only cost effective way to allow general access to these collections. With respect to plants, considerable progress has been made under the Global Plants Initiative (plants.jstor.org).

The level of computer fluency varies widely across the world, as does the accessibility of hardware, software and broadband internet connections. At the pro-iBiosphere workshop on "*e-tools and e-platforms for taxonomy*" organised in February 2013 (in Leiden), there were representatives from a wide variety of non-European countries including Brazil, Brunei, Cameroon, Equatorial Guinea, Ethiopia, Gabon, Indonesia, Iraq, Malaysia, Mexico, Nepal, South Africa, Suriname and Zimbabwe. Clearly, there is interest in these taxonomic software systems from a wide variety of countries in different stages of development.

Conclusions

- Digital tools for taxonomy have advantages to a wide variety of stakeholders.
- Taxonomists still lean towards a print publishing model and changing this attitude may be a slow process.
- Software with a good ease of use and high perceived usefulness can be faster adopted by the community.

Collaboration

For biological, historical and economic reasons, the major biodiversity collections are located in the temperate countries of the northern hemisphere, whereas the majority of undescribed taxa are located in the tropics and southern hemisphere. This is one reason that taxonomists seek collaborations outside their countries. Indeed, in our questionnaire, taxonomists reported that they tended to have more collaborators in foreign countries than at home (Fig. 2).

There are probably several other reasons for this, which may include:

- Funding rules often stipulate that consortia must be from different countries.
- Taxonomists are in competition with their local colleagues for funding, whereas foreign partnerships may improve their chances of funding.
- Alliances with foreign research groups provide resources unavailable locally.
- Collaboration reduces the legal and language barriers to the collection of biota and use of resources from a foreign country.

While collaboration presents many opportunities, it requires fast and efficient communication and a good project management. The number of collaborators reported by the respondents of the questionnaire is summarised in Figure 2. Respondents were asked separately how many collaborators they had within their own country and with foreign countries.

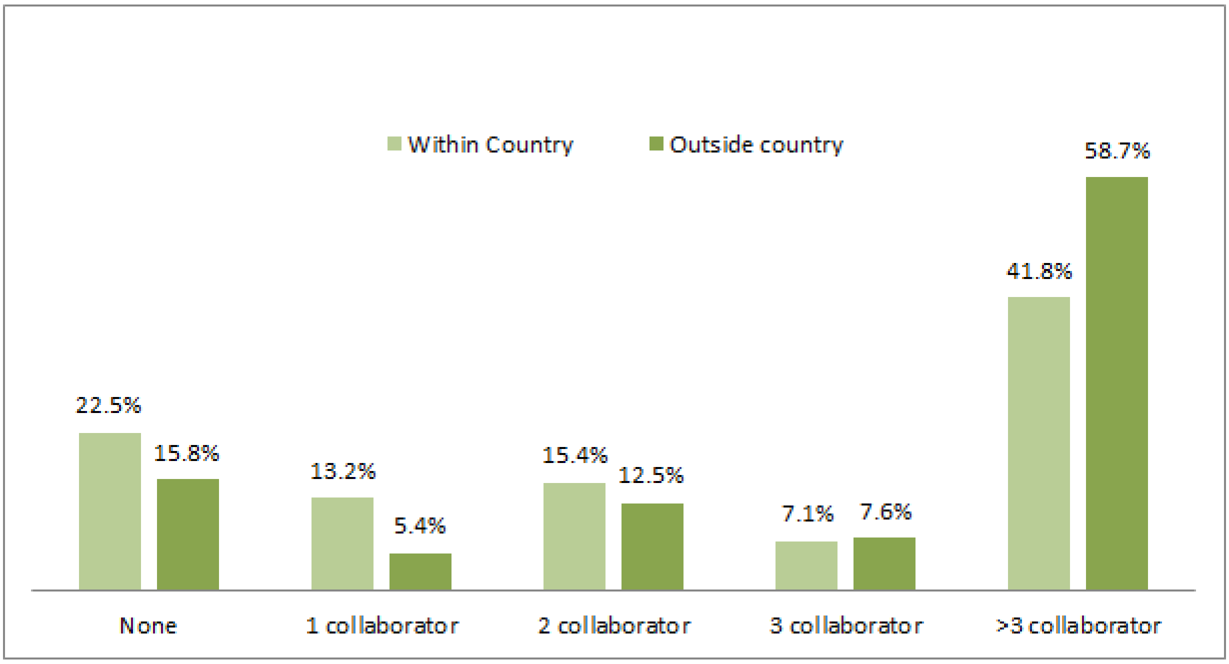
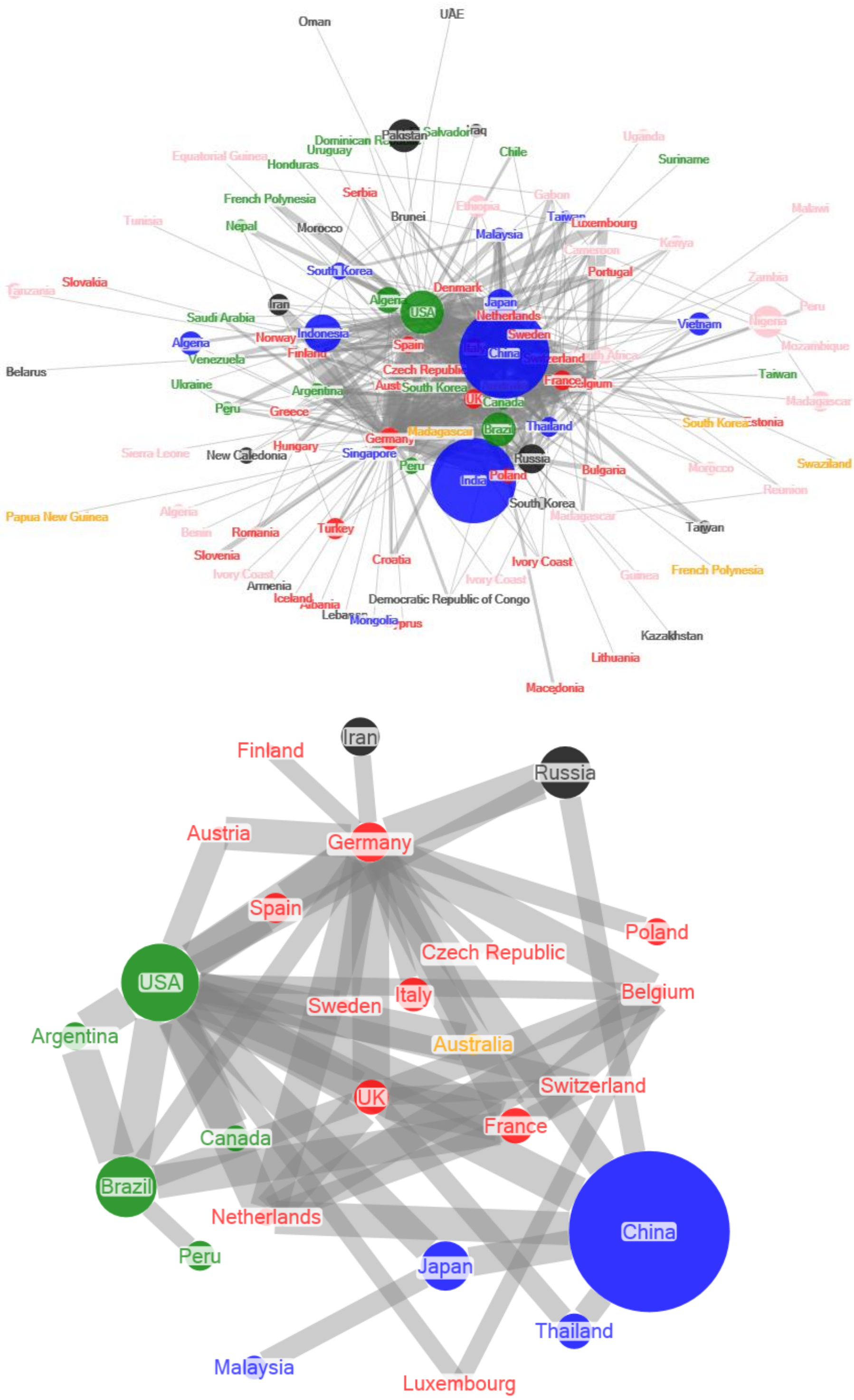


Figure 2. Number of collaborators reported by the respondents of the questionnaire.

To examine the scope of collaboration within the taxonomic community we took data from the questionnaire on the origins and collaborating countries of the respondents (see Appendix 1). These were supplemented with co-authorship data from twenty-two taxonomic journals from botany, mycology and zoology published during 2012. These were assembled into an international collaboration network showing the collaborative links between countries (Fig. 3a). Clearly, taxonomy is a large integrated network of researchers. The centre of this network is a highly connected web containing most of the EU-countries, while the periphery of the network mainly contains the poorer countries. They have smaller research communities, which explains their lack of extensive collaboration. Figure 3b shows a simplified version of the network, filtered to show only the countries with the largest number of collaborations. Several things are evident from these graphs:

- Despite their geographic positions and political alliances, there is only one network of taxonomists.
- European countries provide many of the important links in the networks. Of the top ten countries with the highest centrality in the network, seven are European.
- The number of collaborations a country is involved in is not in proportion to the size of its population, but appears to be strongly related to the number, size and age of its biodiversity collections. While most of the countries with large numbers of collaborative partnerships have large historic biodiversity collections, China and Brazil stand out among the emerging countries.



Figures 3a & b. The collaboration network of international taxonomy. Different colours distinguish the continents. The area of the circle is proportional to the population of the country (population statistics taken from Wikipedia, 2013). The width of the connection is proportional to the number of collaborative links between those countries.

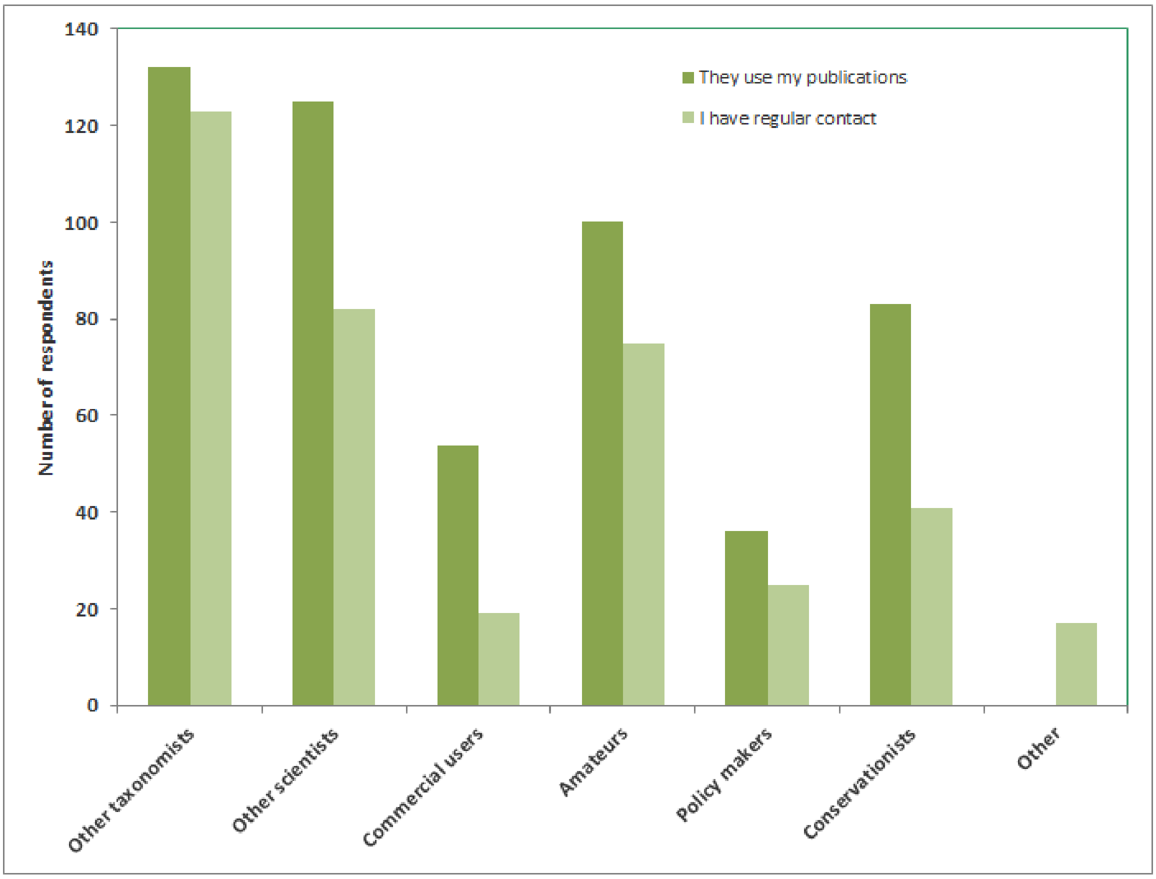


Figure 4. The demographics of the networks of taxonomists.

Respondents to the questionnaire were also asked about their contacts with other professions. The intention of this was to assess the level of contact taxonomists had with the users of their products and potential sources of income for taxonomists. One outcome of the questionnaire was that taxonomists have few contacts with commercial users and policy makers which are, ironically, the two groups that might be able to fund taxonomy. Indeed, it is evident from Figure 4 that taxonomists mainly communicate with other taxonomists and other scientists, and that it is these groups that they perceive as being the main users of their work. Indeed, as their work is largely assessed by their publications in the scientific literature, it is inevitable that they will target their research towards other scientists.

Conclusions

- International collaboration is highly important to taxonomy.
- Software for taxonomists should support collaboration.
- European taxonomic institutions should support systems that encourage international collaborations to foster their central position in international taxonomy.

Barriers to the Adoption of Software Tools by Taxonomists

Traditionally, taxonomy has been firmly based on paper publication and in standard editorial processes. To leverage the advantages of digitised taxonomic information, there will have to be a cultural change in the way many taxonomist work. It is also important that, in particular, taxonomists from developing countries are not left behind. Taxonomy is a truly international science, and to include as much taxonomic expertise as possible from the community, requirements of IT literacy should not be unreasonably high.

In our questionnaire, respondents were asked for their reasons for not using software tools for taxonomy (Figure 5). According to the respondents, the main barriers to their adoption of these tools are a lack of training, a shortage of application support and the time required to learn a new system. These were closely followed by the lack of recognition from internet publications.

Yet all of the other proposed reasons had support from some respondents. In addition, respondents were given the opportunity to supply unstructured responses. These comments underscored the importance of training, but an additional issue is that researchers were daunted by the array of options available. In some cases they expressed that they felt that once they started using a system they would be unable to change in the future.

Indeed, although data import and export in standard formats has been implemented by most of these tools, this functionality is often rather basic. For example, none of these systems allow synchronisation, which is, coordinating updates, additions and deletions between software systems. Current, import/export functionality on most of these systems does not allow two systems to be kept harmonised.

Standards, compatibility of systems and data exchange formats are important subjects for the migration of legacy systems. Data migration is rarely an easy process, but it is important for the rapid adoption of new software systems.

Other issues that came up in these comments were the use of jargon by the IT community; the frequent bugs in software programs and the overly high complexity of programs. The usability of systems is important to encourage user acceptance of software.

Although not explicitly stated by the community, it also seems that many of these software tools do not fit into current workflows for taxonomy and are not seen as providing clear benefits to the taxonomists themselves.

Respondents to the questionnaire were asked what prevented them from using software tools for taxonomy. The answers to this question have been summarised in Figure 5.

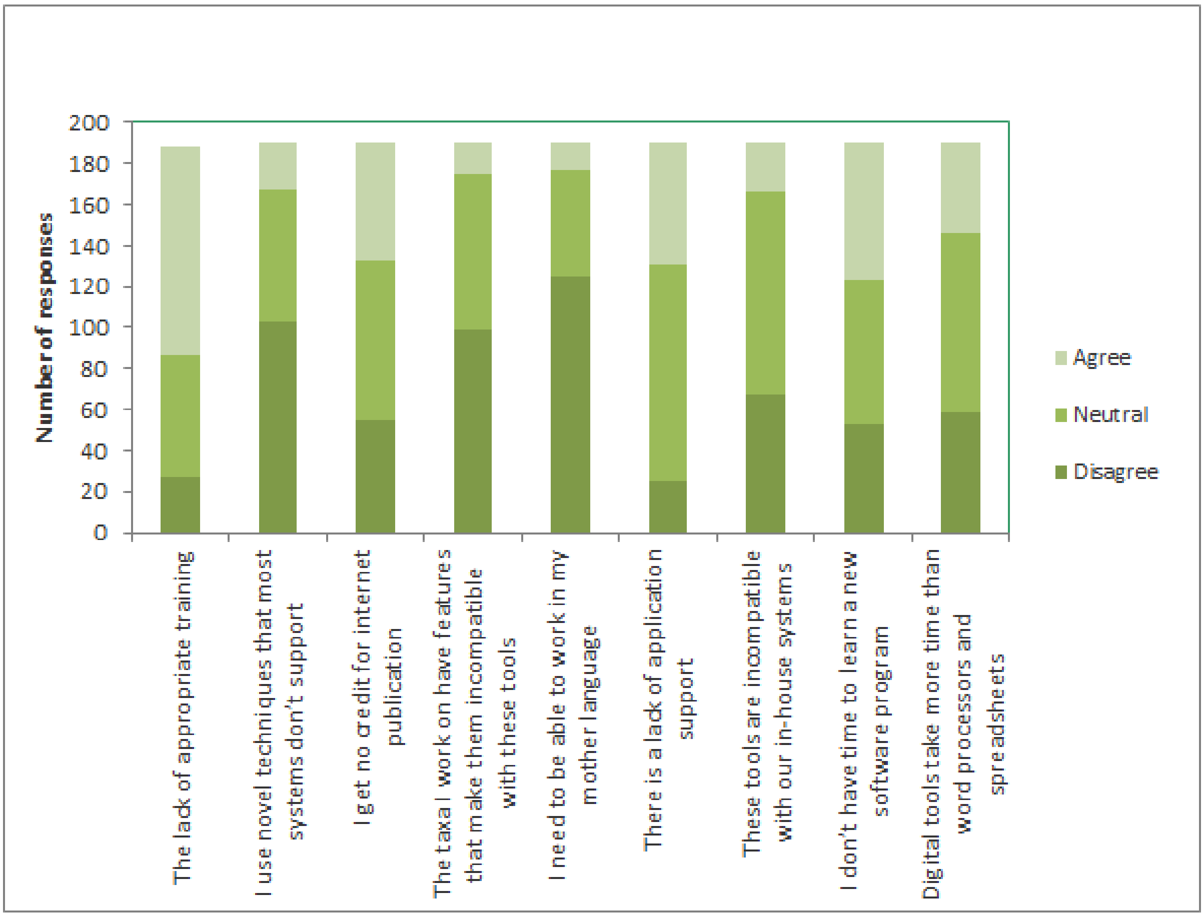


Figure 5. Reasons for not using software tools for taxonomy.

Ignorance of Software Systems

Do taxonomists know which tools are available and what they can be used for?. Awareness of digital tools for taxonomy is not particularly good (Fig. 6). Scratchpads and Delta have the highest recognition. The Scratchpads team has expended a significant effort into promotion and training of users, including promoting an international group of Scratchpad Ambassadors and running training courses (Van de Velde, 2013). Delta in contrast has benefited from its longevity. While it has not been so actively promoted, it was produced at a time when few other IT products were available for taxonomists. The low recognition for Biowikifarm is due to its recent development. Lucid is probably better known in Australia, from where it originates.

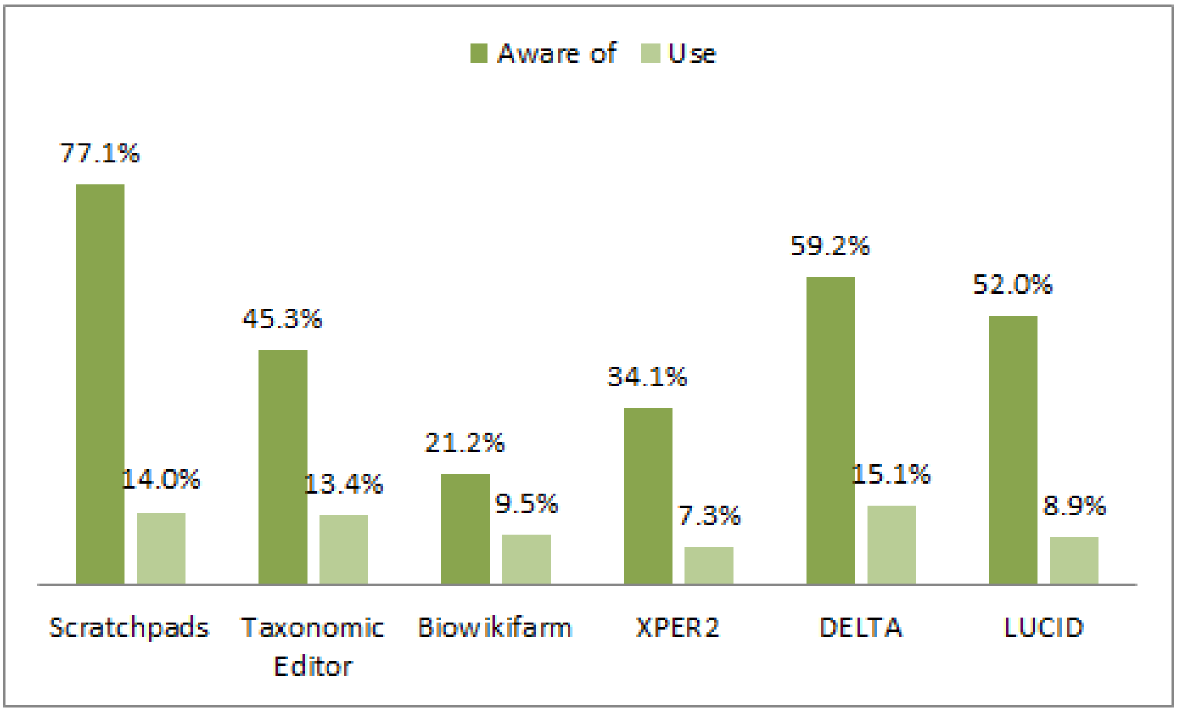


Figure 6. The percentage of respondents that know of, and use different IT systems for taxonomists. NB: Some users use multiple systems.

According to the answers received, a quarter of taxonomists are unaware of the best-known taxonomic software (Fig. 6). In conversations with botanists, even those few who have heard of these systems are unaware of their uses and features. Furthermore, few had tried using these systems, even though the majority are free.

Inexperience

With the increasing digitisation of all science, it is expected that IT skills among taxonomists will become increasingly important and that the training of taxonomists should reflect this. Until recently, the stereotype of a taxonomist was a person either collecting in remote wild places or looking down a microscope at a specimen. However, increasingly, taxonomists are required to handle large digital datasets of DNA, morphology, images, nomenclature and spatial information.

Language

Language did not seem to be of primary concern to the respondents in our questionnaire (Fig. 5). Interestingly, all respondents who thought language was important were European. It seems likely that language is important to those taxonomists writing for a general readership in their own country, whereas, taxonomists writing for a global scientific readership are comfortable writing in English. Therefore, it is likely that language is of much more importance to the users of taxonomic works than to the creators. There is an obvious conflict; it helps taxonomy and science generally if the scientific community uses a common language. However, this isolates the taxonomists from potential users of their work.

FishBase is a good example of an online multilingual taxonomic work, where the international community of ichthyologists has worked together to provide translations.

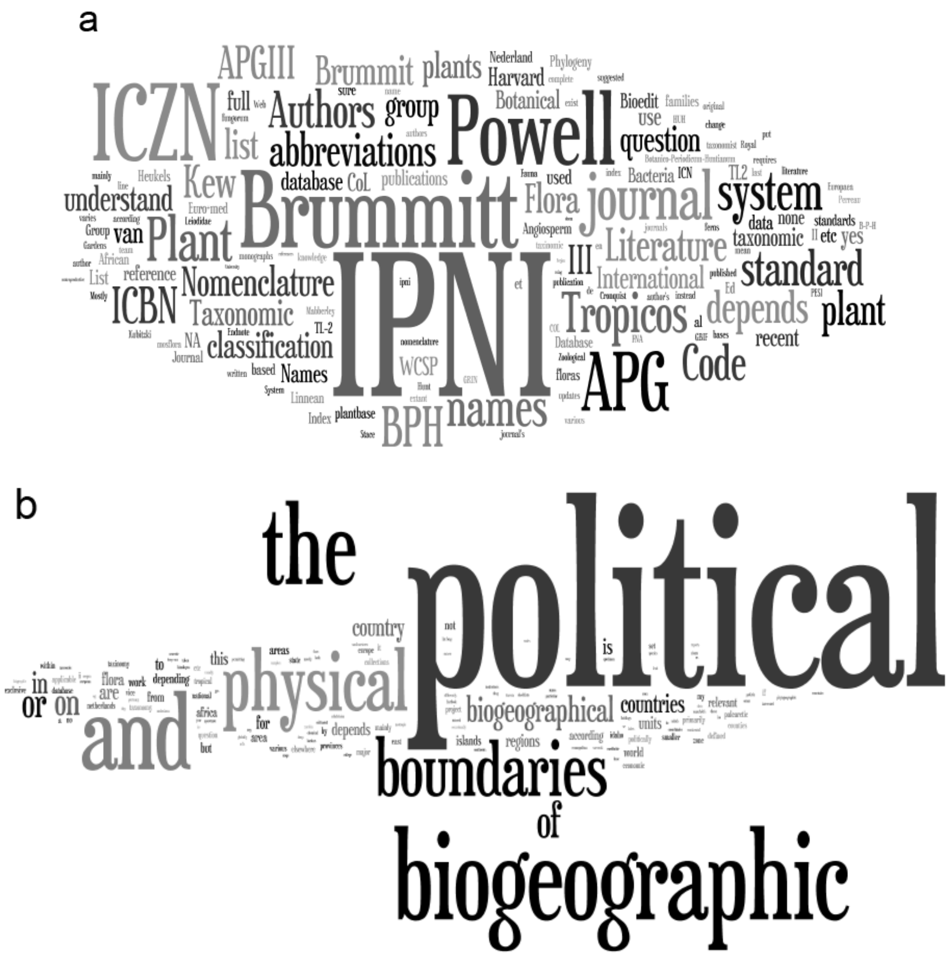
Conclusions

- Further work needs to be done on the promotion of IT tools to taxonomists.
- IT skills should be taught to trainee taxonomists.
- A realistic minimum level of IT-literacy that is necessary to function as a professional taxonomist should be defined and incorporated into curricula minimum requirement to employ taxonomists.
- IT Platforms that support different languages and translations will help to connect taxonomists to their users.

The use of Global Standards

Linked digital information is dependent on having global standards. Therefore, it is important that taxonomists are aware of and consistently use standards. Various standards exist in taxonomy and they are used to varying degrees.

We created word maps to illustrate the contrast between the taxonomists' responses to the use of nomenclatural and geographic standards (Fig. 6a & 6b). For nomenclatural standards taxonomist clearly know what global standards there are and where they can find standardised information. For example, the standardised authority names of Brummitt and Powell (1992) as well as The International Plant Names Index feature prominently (Fig. 6a). In contrast, true standards are barely mentioned by respondents to the question on geographic standards. Indeed, political boundaries are most conspicuous (Fig. 6b), but there is obviously little recognition that political boundaries are not standardised. Among respondents, there is neither awareness of the Taxonomic Database Working Group's world geographical scheme, nor how this relates to the scheme they are using.



Figures 7a. & b. Word maps of the responses to our questionnaire on the use of (a) nomenclatural and (b) geographic standards.

Figure 7. A word map of the responses to the question "What standards, if any, do you use for these [vegetation, ecology/biome, geology, life-form, morphological terms] classifications?"

Regarding the standardisation of other terms, taxonomists use a wide variety of other systems. Some are specific to particular taxonomic works, while others are either country or language specific. Standardisation of all taxonomic terms is not entirely necessary, but it does help users understand text and helps data reuse. Therefore, the prominence of the work 'none' in the word map in Figure 7 suggests that further standardisation of terms would be possible without stifling innovation.

Conclusions

- Nomenclature is well standardised within taxonomic literature but few other areas are.
- There is plenty of scope for further standardisation of taxonomic term usage.

Areas That Need Further Study

Collection Management Software

Specimens are fundamental to taxonomy, particularly those that are designated as nomenclatural types. Specimen management systems are used at the beginning of the taxonomic workflow and the start of the creation of new digital information.

Collection management software was frequently mentioned in conversations with taxonomists and in our questionnaire. This software domain was not well represented in the pro-iBiosphere workshop organised in February 2013. Its focus was on the publication aspect of taxonomy.

The databases used in collections management are as heterogeneous as any other field. Indeed, specimen databases were one of the earliest uses of databases in taxonomy. Therefore, historically, many systems were created in-house, before commercial options were available. Others have become tied into proprietary legacy systems. Such databases are a major topic of conversation in forums such as the Consortium of European Taxonomic Facilities (www.cetaf.org). The two most frequent gripes about such systems are their incompatibility with other systems and their poor usability. Yet alternatives are not without problems and data migration is difficult, expensive and time consuming.

Nonetheless, the limitations of the current systems are well known to institutional managers and modernised versions of various platforms are being developed. For example, a consortium of museums and botanical gardens has formed to steer the development of Brahms Version 8 (herbaria.plants.ox.ac.uk/bol/brahms/Pages/v8). This development has a closed source and software licensing model for sustainability, but there are open source systems such as *Darwin,* the collection management system of the Royal Belgian Institute of Natural Sciences (darwin.naturalsciences.be).

Management of Observations

Taxonomists often make and use observational data to supplement their collection's information. There is probably even less standardisation of software in this field than with collections management software. Each country has its own solutions tailored to their own preferences. Spreadsheets and home-made databases are the norm, not the exception. Needless to say, there is very little coordination between the systems in terms of vocabulary and exchange standards. The challenge of the diverse vocabulary is something that has been studied in depth by the GBIF (2011). Software and standards for the management of observations is a rapidly developing field and something that taxonomists should be aware of, though it does fall out of the scope of this document.

References

Alberts, B. (2013). Impact Factor Distortions. *Science* **340(**6134) : 787. doi: 10.1126/science.1240319

Agosti, D., and Egloff, W. (2009). Taxonomic information exchange and copyright: the Plazi approach. *BMC research notes*, **2**:53. doi:10.1186/1756-0500-2-53

Agosti, D., Penev, L., Catapano, T., Eckert, S., Georgiev, T., Groom, Q., Güntsch, A., Hagedorn, G., Hovenkamp, P., Kirkup, D., Kralt, E., Mietchen, D., Miller, J. and Sierra, S. (2013). D2.1.1 Report on ongoing biodiversity related projects, current e-infrastructures and standards (p. 70). Retrieved from http://wiki.pro-ibiosphere.eu/w/media/c/c9/Pro-iBiosphere_WP2_PLAZI_D2.1.1_VFF_30062013.pdf. (accessed Aug 7, 2013).

Arvanitidis C., Faulwetter S., Chatzigeorgiou G., Penev L., Bánki O., Dailianis T., Pafilis E., Kouratoras M., Chatzinikolaou E., Fanini L., Vasileiadou A., Pavloudi C., Vavilis P., Koulouri P. and Dounas C. (2011). Engaging the broader community in biodiversity research: the concept of the COMBER pilot project for divers in ViBRANT. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. *ZooKeys* **150**: 211–229. doi:10.3897/zookeys.150.2149

Brummitt, R.K. and Powell, C.E. ed. (1992). Authors of Plant Names: a List of Authors of Scientific Names of Plants, with Recommended Standard Forms of their Names, Including Abbreviations. Royal Botanic Gardens, Kew. ISBN 0-947643-44-3.

Campbell, P. (2008). Escape from the impact factor. *Ethics in Science and Environmental Politics*, **8**: 5–7. doi:10.3354/esep00078

Catapano, T., Hobern, D., Lapp, H., Morris, R.A., Morrison, N., Noy, N., Schildhauer, M. and Thau, D. (2011). Recommendations for the Use of Knowledge Organisation Systems by GBIF. Released on 04 Feb 2011. Copenhagen: Global Biodiversity Information Facility, 49 pp., Retrieved from http://links.gbif.org/gbif_kos_whitepaper_v1.pdf (accessed Aug 7, 2013).

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, **13**: 319–340.

Ellis, R., Waterton, C. and Pacha, M. (2007). Assembling Nature: The Social and Political Lives of Biodiversity Software. Lancaster.

Godfray, H.C.J. (2007). Linnaeus in the information age. *Nature*, **446**: 259–260. doi:10.1038/446259a

Hagedorn, G., Mietchen, D., Morris, R. A., Agosti, D., Penev, L., Berendsohn, W. G., and Hobern, D. (2011). Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. *ZooKeys*, (150), 127.

International Commission on Zoological Nomenclature (1999). International Code of Zoological Nomenclature (Fourth Edition). The International Trust for Zoological Nomenclature. ISBN 0853010064.

IUCN Standards and Petitions Subcommittee (2013). Guidelines for Using the IUCN Red List Categories and Criteria. Version 10. Prepared by the Standards and Petitions Subcommittee. Downloadable from http://www.iucnredlist.org/documents/RedListGuidelines.pdf (accessed Aug 7, 2013).

McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Marhold, K., Prado, J., Prud'homme Van Reine, W.F., Smith, G.F., Wiersema, J.H. and Turland, N. J. (2012). International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). International Association for Plant Taxonomy. ISBN 0-853010-06-4.

Pankhurst, R.J. (1971). Botanical keys generated by computer. *Watsonia*, **8**: 357–368.

Penev, L., Roberts, D., Smith, V., Agosti, D. and Erwin, T. (2010). Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research. *ZooKeys*, **50**: 1–4.

Poisot, T., Mounce, R. and Gravel, D. (2012). Moving toward a sustainable ecological science: don't let data go to waste! figshare. http://dx.doi.org/10.6084/m9.figshare.693745 Retrieved 09:00, Aug 27, 2013 (GMT).

Simons, K. (2008). The misused impact factor. *Science*, **322**, 165. doi:10.1126/science.1165316

Van de Velde, I. (2013). Ambassadors network is fully operational. http://vbrant.eu/sites/vbrant.eu/files/M3.15_Ambassadors_network_%20is_fully_operational.pdf (accessed Aug 7, 2013).

Watson, L., and Milne, P. (1972). A flexible system for automatic generation of special-purpose dichotomous keys, and its application to Australian grass genera. *Aust. J. Bot.* **20**: 331–352.

Wikipedia contributors, "List of countries by population," *Wikipedia, The Free Encyclopedia.* http://en.wikipedia.org/w/index.php?title=List_of_countries_by_population&oldid=565727819 (accessed July 25, 2013).

Acknowledgments

Many thanks to all the people who helped in the preparation of this report, particularly Isa Van de Velde, Philippe Kok, Tobias Musschoot, Patricia Mergen; and the taxonomists and herbarium staff of the National Botanic Garden of Belgium.

Appendix 1

Demography of the respondents to the Questionnaire