This version edited: 22nd August 2014

The Shifting Paradigm of Biological Identification

A report for the Tomorrow's Biodiversity Project

August 2013

Dr Richard Burkmar
Biodiversity Project Officer
Field Studies Council
Head Office
Montford Bridge
Shrewsbury
SY4 1HW

richardb@field-studies-council.org
Tel: (01743) 852125

Tomorrow's Biodiversity Project
funded by the Esmée Fairbairn
Foundation

Contents

1 General introduction

The identification of biological specimens remains central to the study of ecology and our understanding of the complex roles of biodiversity in the functioning of the earth's biosphere. Without the ability to identify, and differentiate between, the taxonomic units of the biosphere, we cannot understand the ecosystem services on which mankind depends. As ecosystem services continue to degrade (Millennium Ecosystem Assessment, 2005; UK National Ecosystem Assessment, 2011; Natural Capital Committee, 2013) our understanding is no longer just an intellectual challenge – as it was for John Ray, Gilbert White, Charles Darwin and Alfred Russel Wallace – but a key human endeavour on which our future depends.

The so-called 'taxonomic impediment' recognises a bottleneck in the development of our understanding of the biosphere resulting from a shortage of skills in taxonomy and systematics (Reaka-Kudla et al., 1996; House of Lords Science & Technology Committee, 2008). The taxonomic impediment relates not only to alpha taxonomy (the discovery, description, naming and classification of new species) but, just as crucially, the application of existing taxonomic knowledge to the collection of primary biodiversity data from the field; in other words, the skills of biological recorders.

Tackling the taxonomic impediment to the collection of primary biodiversity data can involve any, or all, of the following elements:

- producing more biological recorders;
- improving the skills of biological recorders;
- enhancing support networks for biological recorders;
- developing and improving tools for managing biological records; and
- developing and improving tools and resources for biological identification.

The rapid development and adoption of information technology over the last forty years or more has affected almost every aspect of human endeavour, including taxonomy and systematics. There are now many more potential approaches by which we can deliver the elements listed above and help to remove the taxonomic impediment.

This report briefly reviews the impact of the computer revolution on taxonomy and systematics in general and, in particular, on biological recording in the UK – looking in greatest detail at the development of new tools and resources for biological identification and considers whether we are taking full advantage of the opportunities presented by emerging information technology platforms, standards and paradigms.

The report was written in August 2013, but this version incorporates minor edits made in August 2014. It is distributed under a Creative Commons Attribution license which means that you can freely distribute it or derive other work from it, as long as the Field Studies Council is credited for the original creation.

2 The evolution of eTaxonomy in biological recording

E. O. Wilson's widely-quoted line – "*imagine an electronic page for each species of organism on Earth, available everywhere by single access on command*" – is not sufficient, by itself, to convey the grandness of his vision of an interconnected cyber-taxonomy for all life on earth (Wilson, 2003). Wilson imagined a host of diverse but complementary and interconnected information systems that together would enable and comprise an 'infinitely expansible' resource for every species, summarising "*everything known about the species' genome, proteome, geographical distribution, phylogenetic position, habitat, ecological relationships and, not least, its practical importance for humanity*".

In the ten years since Wilson's expression of this vision, substantive progress has been made towards assembling the information technology infrastructure that can help realise it. Initiatives like Encyclopedia of Life (EOL, 2013), Catalogue of Life (Catalogue of Life, 2013), Global Biodiversity Information Facility (GBIF, 2013), and IdentifyLife (IdentifyLife, 2013) have all contributed to the realisation of a vibrant 'eTaxonomy' (Zauner, 2009).

But eTaxonomy is more than just the elements which comprise the end product – the comprehensive encyclopaedia of life on earth – envisioned by Wilson. It is the totality of the information technology tools, systems, networks, standards and data that enable it. The following areas of development are all integral and major influences on the evolving eTaxonomy:

- systems for collecting, collating and managing biological records;
- molecular identification techniques - DNA barcoding;
- automated species identification systems;
- computer-based biological identification tools and resources; and
- virtual support communities.

In this report the term 'computer-based biological identification' is used to refer to computer programs that help people to make biological identifications, e.g. by creating or implementing identification keys, and does not encompass programs which facilitate the other areas of eTaxonomy listed above. A discussion of developments in computer-based biological identification and virtual support communities forms the major part of this report, but it includes a brief introduction to the other facets of eTaxonomy which lend vital context to that discussion.

2.1 Biological record management systems

There can be very few active biological recorders in the UK who have not been positively affected by the application of information technology to the collection and management of biological records over the last few decades. PC-based tools like Recorder (JNCC, 2013), MapMate (MapMate, 2013) and even spreadsheets like Microsoft Excel® have transformed the way that biological records are managed and exchanged. More recently, internet-based tools like RODIS (Cheshire rECOrd, 2013), Living Record (Living Record, 2013), BirdTrack (BTO, 2013) and iRecord (BRC, 2013) and now smartphone apps like BirdTrack Mobile (BTO, 2012), Record Wildlife (Record Wildlife, 2013) and iRecord Ladybirds (CEH, 2013) have further enriched the record entry and management toolset.

The UK biodiversity informatics network has grown over the same period and is comprised of, among others, National Recording Schemes & Societies, Local Environmental Records Centres and the national Biological Records Centre, all, to a greater or lesser extent, exchanging biological records and contributing primary biodiversity data to the NBN Gateway – the British node of the Global Biodiversity Information Facility (GBIF).

The journey of the UK biodiversity informatics network to where it is today has not always been easy and there remain substantial problems to overcome. One reason has been a lack of focus on consensus-built standards around which technical innovation can occur. When it comes to biological records management in the UK, we have not lacked for excellent technical innovation (as the previous examples testify), but the totality of that innovation has not been greater than the sum of its parts; in fact you might reasonably argue that it has amounted to less than the sum of its parts. Different systems implement radically different data models and exchange formats and use incompatible record identifiers. One upshot of this is that records are duplicated – with different identifiers – in different systems where each can start a 'life' of its own. Inevitably this leads to record duplication. No one knows – even to the nearest few million – how many of the 85 million records on the NBN Gateway are duplicates.

Latterly there has been some coalescing of development effort around *de facto* standards promoted by the NBN (NBN, 2013b, 2013a), but there remains a lack of leadership within the UK biodiversity informatics network in terms promoting consensus-built standards and interoperability. This is significant in the context of this article because of its relevance to the discussion on standards around computer-based biological identification which follows.

2.2 DNA barcoding

Of 86 papers and abstracts presented at a major European conference in 2010 called *Tools for Identifying Biodiversity* (Nimis and Lebbe, 2010), 18 concerned themselves principally with DNA barcoding. DNA barcoding refers to the technique of identifying biological specimens and describing new taxa by examining sequences of molecular DNA (Hebert et al., 2003; CBOL Plant Working Group and Janzen, 2009; Casiraghi et al., 2010).

The initial excitement that accompanied early developments in DNA barcoding generated as much heat as it did light. Many taxonomists and biological recorders responded defensively to the attendant, and frequently outlandish, hype around it (e.g. Ebach and Holdrege, 2005). But as the technology has developed over the last 10 years, and both its potential and limitations have become more widely appreciated (e.g. Shen et al., 2013), more realistic expectations are emerging. The technique is becoming a useful complement to traditional techniques for classifying species and identifying specimens, but has not, and probably never will, completely supplant any existing skills or techniques (Sbordoni, 2010).

A few examples can illustrate the potential utility of DNA barcoding in the collection of primary biodiversity data. The work of Cubelio et al. (2010) suggests that it could help to distinguish between commercially important crabs in India, including species where the adults are cryptic. Galimberti et al. (2010) found barcoding to be a useful technique for separating cryptic taxa of *Myotis* bats. Grube and Muggia (2010) propose DNA barcoding as a potentially tractable method for identifying the algal and cyanobacterial photoautotrophs in lichens. Bertolani et al. (2010) showed how DNA barcoding can complement other techniques used to classify and identify tardigrades.

Ji et al. (2013) demonstrated that 'metabarcoding' – a technique for separating 'operational taxonomic units' (OTUs) from DNA soup derived from relatively unprocessed samples or 'environmental DNA' – can serve to provide estimates of alpha and beta diversity for invertebrates that are comparable to those obtained from more time-consuming identification through traditional morphological comparisons. However, at this point it time, the resolution of OTUs to named taxa was far lower with the metabarcoding technique than with the traditional techniques. In other words although it provided reliable estimates of diversity, it is not currently a sufficiently good technique for identifying all the species present.

DNA barcoding has clear potential to help biological recorders get to grips with cryptic taxa, but it seems likely that we are still some way off the point where the technology will have matured enough to be readily accessible to most of them. But when that does eventually happen, it will illuminate the work of many biological recorders and make them more productive; a far cry from making their skills redundant.

2.3 Automated species identification

Of the 86 papers and abstracts presented at the 2010 conference *Tools for Identifying Biodiversity* (Nimis and Lebbe, 2010), 8 concerned themselves principally with automated species identification. The term 'automated species identification' is not well-defined, but here we use it to refer to computer-based identification systems (other than those based on DNA sequencing) that aim to automatically assign biological specimens to known taxa with no, or minimal, human interaction (Gaston and O'Neill, 2004).

Automated species identification is an area of rapid technical development which includes assigning specimens to known taxa based on photographic images or audio recordings – techniques of great potential value to biological recorders. *Non-automated* techniques for identifying organisms based on photographs and audio recordings are well-established. The iSpot project (iSpot, 2013) is a widely-used web resource where digital photographs of organisms can be posted and identified by a very broad community of users, including many experts. Assigning observations of bats to taxa by analysing recordings of their ultrasonic vocalisations is now a long-established primary identification technique (Bach et al., 1977). The effectiveness techniques such as these depends on three limiting factors:

1. the technical quality of the photograph/recording (e.g. resolution and clarity);
2. the appropriateness of the photograph/recording (it should include the salient features for identification); and
3. the availability of someone with an appropriate level of expertise to examine the photograph/recording.

In most cases the third of these – the availability of appropriate expertise – is the limiting factor. Education and simple technical solutions can address the first two far more easily than the third. But developments in machine learning and both image and sound processing show encouraging signs that, for some identification problems, automated species identification could help address the problem of lack of available expertise.

Although work on automated species identification has been ongoing for more than 15 years (an early example was the DAISY system, O'Neill et al., 1997), progress has been slower than for DNA barcoding and most of the literature, to date, describes experimental and evaluation work rather than production systems for biological identification. The technology has been applied to a diverse range of taxonomic groups and a wide range of visual and acoustic features of those groups, including, for example, mites by 'landmark' recognition (Becerra and Valdecasas, 2004), moths by wing pattern, size and shape (Mayo and Watson, 2007), crickets, cicadas and katydids from sounds produced by stridulation (Ganchev and Potamitis, 2007), marine polychaetes from body shape (Nikolaou et al., 2010), small mammals from size and shape of footprints (Geng et al., 2012), trees from leaf shape and venation (Zhang et al., 2008), birds from their songs (Daou et al., 2012) and bats from their ultrasonic echolocation calls (Agranat, 2013).

Rapidly advancing technology and analytical techniques (e.g. artificial neural networks) have delivered improving success rates with these identification techniques (e.g. Mayo and Watson, 2007; Ganchev and Potamitis, 2007) and this trend looks set to continue. As Gaston and O'Neill (2004) pointed out: "*Vision and enterprise are perhaps more limiting at present than practical constraints on what might possibly be achieved*."

3 Traditional keys and field guides for biological identification

Griffing (2011) showed that the development of traditional paper-based keys started at least 320 years ago when, in 1689, Richard Waller made a presentation to the Royal Society entitled "*Tables of the English Herbs reduced to such an order, as to find the name of them by their external figures and shapes.* " The subsequent development and theory of traditional paper-based keys is comprehensively described elsewhere (e.g. Pankhurst, 1978; Tilling, 1984). An excellent overview of the principles of biological identification keys and the terminology used is given by Hagedorn et al. (2010). The major types of key and their main features are briefly reprised here to put the discussion of computer-based biological identification resources, and their features, into context.

3.1 Single-access keys

The traditional paper-based dichotomous key remains a fundamental tool in the armoury of many biological recorders. Identification of a specimen is achieved by starting at the root of the key and choosing between pairs of descriptions, known as couplets, which normally consist of contrasting states of one or more morphological characters. Each description in a couplet is also known as a 'lead'. Depending on which lead in a couplet most closely matches the specimen, the user is directed to another couplet in the key or to the identity of the specimen. A useful way of conceptualising such a key is as a decision tree and dichotomous keys are occasionally presented as such (Tilling, 1984).

Dichotomous keys are a type of 'single-access' key because there is a single starting point at the root of the key (although in practice an experienced user of a key might jump to a particular couplet when they know, from experience, that a particular type of specimen will take them there). But not all single-access keys are strictly dichotomous – sometimes 'couplets' may actually consist of three or more leads (polytomous).

There are two major ways of presenting single-access keys on a printed page: 'parallel' and 'yoked'. In parallel keys, the leads of a single couplet are presented one after the other. An advantage is that the contrasting leads can be easily compared, but a disadvantage is that a lead will often need to direct the user to a couplet some distance away in the key. A well-known example of a parallel key is *Field Flora of the British Isles* (Stace, 1999), sometimes known as 'Baby Stace'. Parallel keys are sometimes referred to as 'bracketed' keys.

In the yoked style of presentation, leads of couplets are split so that leads of successive couplets which result in an identification can be grouped together. An advantage of this style is that for some identifications (preferably those most frequently required) all the positive leads which result in that identification are presented together, one after the other. A disadvantage is that it can be difficult to contrast the leads of a single couplet which can be widely separated in the printed key. A well-known example of a yoked presentation of a single access key is *Grasses: A Guide to Their Structure, Identification, Uses, and Distribution in the British Isles* (Hubbard and Sampson, 1968), also known simply as 'Hubbard'. Yoked keys are sometimes referred to as 'indented' keys, but the term can be confusing since many parallel keys also employ indentations for clarity.

3.2 Multi-access keys

'Multi-access' keys operate in a very different way. In a multi-access key, the user can consider the morphological characters in the key in any order, matching them against the character states from their specimen. One way to visualise this is as a character matrix (or spreadsheet) with each column in the matrix representing a character, each row a species and each cell the state of the corresponding character for the corresponding species. (In fact the creation process of single-access keys also often starts with such a matrix.) By noting (perhaps marking) the character states which match the specimen in hand, a user should see which of the rows in the key (corresponding to a species) most closely agrees with the specimen. Note that there could be more than one potential match.

The many advantages of multi-access keys over single-access keys) are summarised in box 1. A disadvantage of multi-access keys is that all but the simplest are problematic to produce and use on the printed page. An example of a paper-based multi-access key is found in *Keys to the Families of British Spiders* (Jones-Walters, 1989). Multi-access keys are also sometimes referred to as 'free-access' keys or, wrongly, as 'synoptic' keys. (The terms 'synoptic' and 'artificial' are used, more correctly, to contrast types of key that either reflect a taxonomic relationships in their decision trees (synoptic) or ignore taxonomic relationships, in favour of ease of use, in their decision trees (artificial) (Hagedorn et al., 2010b).)

Box 1. Advantages of multi-access keys over single-access keys (not computer-based).

- ***Control.*** *Multi-access keys allow the user more control over the identification process. In contrast to a single-access key, where characters must be considered in the order presented by the key, the user can choose what character to consider next when using a multi-access key.*
- ***Uncertainty.*** *In a multi-access key, the user can select more than one state for a character to reflect uncertainty about the value of the character state. The end result may be that more than one potentially matching taxon is 'keyed out', but this helps to avoid the tendency of single-access keys to force the user's hand down one route or another, even when they are uncertain.*

- ***Incompleteness.*** *Dealing with incomplete information. If a user of a single-access key comes upon a couplet describing a character which is not evident from the specimen (e.g. because of an incomplete specimen or a specimen at a stage of development not appropriate to the character), he/she is effectively stymied. The only possible action is to subsequently follow both leads with all the problems that entails. And if the situation occurs again, further on in the key, the key becomes almost unusable. But with a multi-access key, inappropriate characters can just be ignored.*

Before the widespread availability of computers, many interesting – even ingenious – mechanical ways were devised to implement multi-access keys including punched cards (termed 'polyclaves') and plastic overlays (Tilling, 1984; Walter and Winterton, 2007). But the very complexity of these methods was a weakness and parts could get lost or wear out, so multi-access keys never realised their full potential and biological recorders were not, by and large, exposed to their many potential advantages.

3.3 Field guides

Field guides to the identification of animals, plants and fungi can be considered a form of biological identification resource distinct from keys. The variety of field guides is huge but a common feature is the use of images as the primary means of conveying information on the morphology of each taxon covered. Most support the images with some textual information which can highlight the key identification features and variously include information on distribution, phenology, ecology, etc. Although much of the technique when using a field guide relies on visual comparison of a specimen against the images, many of the best guides also include limited keys to assist in the process of selecting appropriate images for comparison. Stevenson et al. (2003) includes an excellent review of the history of the field guide format.

3.4 Producing keys and field guides

The construction of an effective key or field guide owes as much to art as it does to science. Keys and field guides are an interface between people and complex taxonomic knowledge bases, their purpose to provide clear pathways through the knowledge base to correct determinations. For any average-sized group of taxa, hundreds of different keys and/or field guides could be constructed, all with this single purpose in mind. Many would be radically different from each other and others more subtly so, but even subtle differences can have a huge impact on their usability and effectiveness.

Writers of keys and field guides require feedback from potential users in order to settle on something which works for most people most of the time and building a testing and feedback loop into the construction phase is vital. Furthermore, a commitment to produce future editions to deal with taxonomic revisions, improvements to the knowledge base and usability issues that surface after publication is important if a key or field guide is to remain relevant and usable into the future (Tilling, 1987; Lawrence and Hawthorne, 2006).

Costs and difficulties associated with the production and distribution of paper keys and field guides make the testing, revision and publication process a challenging one, especially given the limited audience for many of these resources, but there are examples where this practice is successfully embedded in the development of identification resources, for example The Field Studies Council's AIDGAP series (Aids to Identification in Difficult Groups of Animals and Plants) (Tilling, 1987).

4 Computer-based biological identification

The history of computer-based biological identification is almost as long as that of electronic computing itself. Almost as soon as computers became widely available to the academic community, ecologists recognised their potential as tools for biological identification. The roots of computer-based biological identification can be traced back to the 1960s and tools of types we would recognise today began to appear in the 1970s (Pankhurst, 1991; Edwards and Morse, 1995).

4.1 Programs to facilitate the creation of paper keys

The first developments in computer-based biological identification were, naturally enough, computer programs to facilitate the creation of traditional paper-based keys. In the simplest of these systems, dichotomous keys are encoded in a digital format understood by a program responsible for laying out and printing the key (Morse et al., 1968). One advantage of separating the knowledge representation (the digital encoding of the key) from the presentation of the key (layout and printing) is that new versions of the key, incorporating changes, can easily be produced by editing the underlying data representation before re-running the program. Another is that different printed layouts can easily be applied to the same key. It should be possible, for example, to produce either a yoked or parallel printed key (or both) from the same underlying data. Note, that it is still the responsibility of the designer of the key – not the computer program – to specify the decision tree of keys produced in this way.

The 1970s saw the arrival of the first computer programs which could themselves examine a character-attribute data matrix and automatically create and evaluate alternative decision trees (Pankhurst, 1970; Hall, 1970; Morse, 1971). These programs evaluate decision trees based on heuristics defined by the programmer, or the user in more interactive systems, such as "*a couplet should divide the remaining taxa into two groups which are as close to each other in size as possible*", and various weightings describing, for example, reliability and convenience for each morphological character (represented in the knowledge base). These programs should not take control out of the hands of key designers who have the ability to override decisions made by the program (Dallwitz, 1974; Pankhurst, 1988). Their great advantage is that the key designer is relieved of the laborious task of evaluating hundreds of possible decision trees from a character-attribute matrix and can concentrate instead on evaluating and refining those that are most effective.

4.2 Computer-based single-access keys

The next logical development was 'interactive' or 'online' single-access keys; in other words, keys that instead of being printed are used directly on the computer. It was the advent of the Hypertext Transfer Protocol (HTTP) – the basis of the World Wide Web – which really made this possible in the 1990s. We take them for granted now, but hyperlinks – words or images on a web page that link to other web pages – were revolutionary when introduced and there could hardly have been a better innovation for implementing interactive keys. The first and most obvious use for a hyperlink is to transport a user to the next relevant couplet (or a summary of an identified species) when a lead is selected. But there are many advantages of implementing single-access keys interactively on computers (Wright et al., 1995; Farr, 2006) and these are summarised in box 2. An example of a relatively early computer implementation of a single-access key is Taxakey (Chesmore and Yorke, 1997).

Box 2. Advantages of interactive computer-based over paper-based keys.

- ***Vocabulary.*** *Many keys require a specialist vocabulary which is sometimes specific to a taxonomic group. When a user of a key encounters an unfamiliar word, they have to go to a glossary on another page to look it up. This can make a key awkward to use, particularly in the early stages of learning. In an interactive key, words that require explanation can be linked to another page (or more normally nowadays a 'pop-up' box) containing an explanation. Consequently the disruption caused by unknown words in interactive keys is minimised.*
- ***Illustrations.*** *Pictures, be they diagrams, line drawings or photographs, can greatly increase the usability and effectiveness of a key (Leggett and Kirchoff, 2011), but space for illustrations within the body of a paper key is limited. Illustrations can be included at the back of a key but this, as with a glossary, has obvious drawbacks. An interactive key can include direct links to pages or pop-up boxes that host illustrations relevant at any relevant point.*
- ***Currency.*** *Computer-based keys, particularly those presented online, are normally more easily updated than their paper counterparts, holding out the possibility that keys can be quickly updated to keep abreast of changes in taxonomy, new research etc.*
- ***Access.*** *Computer-based keys are more easily distributed than paper-based keys – by publishing to the internet, a key can be quickly available anywhere in the world with an internet connection.*
- ***Backtracking.*** *Inevitably mistakes are made when using a key. Backtracking to the couplet at which the user feels an error might have occurred is problematic with a paper key. In computer-based keys, a 'back button' can quickly be used to backtrack through a key.*

4.3 Computer-based multi-access keys

Arguably the biggest impact of computers on biological identification keys has been in the way they have transformed the delivery of multi-access keys. The advantages of implementing single-access keys on computers also apply, to a large extent, to multi-access keys, but there are many more besides. The difficulties of implementing multi-access keys without computers have already been alluded to: laying out and using multi-access keys on a printed page is problematic for all but the simplest of them and ,mechanical implementations, such as punched cards, are cumbersome or inconvenient. Computer-based systems do not suffer from the same space or usability problems and they have also enabled entirely new features that weren't practical with paper or mechanical delivery methods (box 3).

Box 3. Advantages of computer-based multi-access keys.

- ***Specimen description****. Computers offer a host of user-friendly interfaces for inputting relevant character states for a specimen and these can easily be modified or characters excluded from consideration altogether. Some of these features apply to other mechanical or paper-based methods, but none can match the flexibility and ease of use offered by computers.*
- ***Numeric characters****. With paper-based keys, numeric characters (e.g. measurements) from a specimen must be compared to ranges in the key (which are often given as two ranges – a normal range and an 'exceptional' range). With computer-based multi-access keys, numeric*

measurements from a specimen can be entered directly. The computer will do the work of comparing against various ranges.

- ***Best character**. The discriminating power of a particular character changes depending on the character states which have already been entered for a specimen. The computer can advise the user which of the remaining characters (for which a state has not been entered) has the most discriminating power at any particular point.*
- ***Error tolerance**. The computer can be instructed to ignore a user-specified number of mis-matches between entered character states and those for individual taxa in the knowledge base before eliminating them from further consideration.*
- ***Similarity ranking**. The computer can calculate, at any given point, a similarity index (using one of a variety of statistical methods) between the specimen and each taxon in the key's knowledge base (for the character states entered). This can be an alternative strategy to eliminating taxa which do not match character states entered by the user.*
- ***Probabilistic determination**. Related to similarity ranking; computers can embrace the concept of uncertainty by calculating, for every taxon in the knowledge base, a likelihood that it matches the specimen.*
- ***Instant feedback**. Every time a new character state is entered (or altered) the computer can provide instant feedback to the user on eliminated/potential taxa and/or similarity indices or likelihood values (see above).*
- ***Extra features**. Computers can add value to multi-access keys in a number of ways, for example by comparing character states in the knowledge base, two taxa can be contrasted and the most important distinguishing characters highlighted; images can be flexibly presented and juxtaposed against each other to highlight key differences and similarities; for any selected taxon, the computer can indicate which are the most diagnostic character states.*

The sum total of the advantages (actual or potential) of computer-based multi-access keys over traditional paper-based single access keys is expressed by the combined boxes 1-3. An early example of a computer-based multi-access key for British *Carex* sedges was written for IBM-compatible PCs and BBC microcomputers in the early 1990s (Legg, 1992). Other early systems include Delta Intkey (Dallwitz, 1992) and Cabikey (Chesmore and Yorke, 1997).

4.4 Progress and development

Considering that the potential of computer-based biological identification has been actively explored for upwards of 40 years, the penetration of this technology into the everyday toolbox of biological recorders has been remarkably poor (Morrison, 2011, 2012). Most biological recorders still rely on paper-based identification resources even though, for many identification problems (but not all), the facilities offered by computer-based resources would be superior.

It is interesting to compare the relative lack of progress in computer-based identification resources to the greater progress made with systems for biological records management described previously. One reason for this difference lies in the availability, or otherwise, of alternative systems. Before the advent of computerised biological record management systems, biological records were, for the most part, only shared when they were collated to produce documents such as county floras. But beyond this, records were not used and re-used as they are in today's computerised systems. The comparatively greater development and uptake of biological records management systems does not necessarily imply that biological recorders are universally happy with these systems, it’s just that their arrival did not present an alternative to another paradigm, rather it presented an entirely *new* paradigm that offered benefits impossible to ignore.

On the other hand, computer-based identification tools evolved from pre-existing paper-based tools which served the requirements of biological recorders very well for many years and continue to do so. So computer-based identification tools have to prove themselves alongside some well-established alternatives and a tradition in which the use of a technically complex paper-based dichotomous key has been regarded a badge of expertise. Consequently the natural inertia of people to experiment with new ways of doing things has been accommodated.

Another reason for the relatively slow development of computer-based biological identification compared to biological records management is that the former is more demanding of the technology required to implement it. Biological records management is implemented with database and data-exchange technologies that have been the fundamental building blocks of the Information Technology revolution from the start and has never been seriously limited by the available technology.

On the other hand, the ambition of computer-based biological identification tools has often exceeded the capability of the technology. Key technical limitations have been the sophistication of Graphical User Interfaces (GUIs) and the portability of hardware. Starting with the arrival of hypertext (with the World Wide Web) and continuing with advances in graphics and touch-screen technology, the limitations of GUIs have all but gone and in recent years the miniaturisation and portability of hardware has been revolutionary. There remain few, if any, technical barriers to the effective delivery of computer-based biological identification tools and there are encouraging signs that progress is now accelerating (Nimis and Lebbe, 2010).

An impediment to progress in both fields contrasted above has been the slow development of standards that promote interoperability and collaborative effort. The result is that innovations and developments in computer-based biological identification, as with biological records management, have often happened in silos and the tools for developing and running interactive keys have been tightly bound with the knowledge bases that underlay them, meaning that keys developed on one system cannot be easily used in another. This has stifled advancement and resulted in ‘re-invention of the wheel’ in many instances. For potential users of computer-based interactive keys, the impression has been one of a confusing and immature market. This changing situation is examined later in this report.

4.5 Current computer-based biological identification tools

4.5.1 Current accessibility

Lists of computer-based identification tools have been published elsewhere (e.g. Walter and Winterton, 2007; Dallwitz, 2011), but these lists quickly become out of date as new initiatives occur and as the products of older initiatives become unavailable or otherwise irrelevant. Although not exhaustive, table 1 lists some of the computer-based identification tools encountered today – both in the market place and in the literature. Preference has been given to those programs or tools which are accessible or otherwise significant.

Table 1. Recent and current computer-based biological identification programs. The table indicates the extent to which each supports the Structured Data Description (SDD) interchange format and whether or not the program is commercial or free.

Name	Notes	SDD support	Cost
Delta Key	Gnerates paper-based single-access keys. Developed by Mike Dallwitz since the 1970s. Unclear if still maintained. Doesn't seem to be available for 64 bit computers. delta-intkey.com	None	Free
Delta Intkey	Builds/plays interactive multi-access keys. Originator of the DELTA data format. Developed by Mike Dallwitz since the 1990s. delta-intkey.com	None	Free
Lucid3	Builds/plays interactive multi-access keys. Developed by the Queensland Alliance for Agriculture and Food Innovation (University of Queensland). An earlier version of the software can be downloaded and used for free, though it does not include support (e.g. help files). www.lucidcentral.com	Support for SDD export & import	Commercial (but see notes)
Lucid Phoenix	Builds/plays interactive single-access keys. Developed by the Queensland Alliance for Agriculture and Food Innovation (University of Queensland). www.lucidcentral.com	None	Commercial
3I Interactive keys	Builds/plays interactive multi-access keys. Developed by a Dmitry Dmitriev in the USA since around 2006. imperialis.inhs.illinois.edu/dmitriev	Some support for SDD export	Free
Linnnaeus II	Builds/plays interactive single-access keys flexibly from taxonomic databases. Developed by ETI Bioinformatics for primary taxonomists and systematists. Free to contributors to the ETI World Biodiversity Database, otherwise commercial. www.eti.uva.nl/products/linnaeus.php	Support for SDD export	Commercial (but see notes)
Xper²	Builds/plays interactive multi-access keys. Developed at the Laboratoire Informatique et Systématique. Associated with the KeyToNature project. lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2	Support for SDD export & import	Free
iSpot Bayesian keys	Plays interactive multi-access keys through the iSpot website. Users can work with iSpot to develop new keys. www.ispot.org.uk/keys	None	Free
FRIDA	Builds/plays interactive single-access keys flexibly from taxonomic databases. Developed since 2003 by the Department of Life Sciences of the University of Trieste (Italy). Linked to Dryades and KeyToNature projects. www.dryades.eu & www.keytonature.eu	Some support for SDD export	Unclear
Open Key editor/player	Builds/plays interactive single-access sub-keys for specified sub-sets of taxa from a master key. Developed as part of the KeyToNature project.	Support for SDD export & import	Free
Biowikifarm Wiki Keys	Format to Build/play interactive single-access keys directly in web pages implemented in the MediaWiki platform. Developed as part of the KeyToNature project.	Some support for SDD export	Free
XKey	Builds non-interactive single access keys. Developed by a collaboration of Spanish academics (Delgadocalvoflores et al., 2006).	Native support for SDD	Unknown
Navikey	Java-based tool to play interactive multi-access keys encoded in the DELTA data format. http://www.navikey.net/	None	Free
PanKey	Suite of tools for identification including interactive multi-access key generation. Supports the DELTA data format. www.exetersoftware.com/cat/pankey/pankey.html	None	Commercial
XID	Builds/plays interactive multi-access keys. keys. xidservices.com	None	Commercial
X:ID	Online tools to build/play interactive multi-access keys. Developed for the uBio project. http://www.ubio.org/index.php?pagename=XID/key	None	Free
SLIKS	Plays interactive multi-access keys in web browsers. Developed by Gerald F. Guala. stingersplace.com/SLIKS	None	Free
MEKA	Builds/plays interactive multi-access keys. Developed at University of California, Berkeley. ucjeps.berkeley.edu/meacham/meka	None	Free

The systems listed in table 1 are not equally accessible to all biological recorders. Accessibility depends both on the limitations of the technology and on whether the software is to be used primarily to make use of existing interactive keys, or to develop new resources.

Potential accessibility issues include:

- software availability & cost;
- supported platforms;
- availability of technical support;
- key (knowledge base) availability; and
- interoperability.

With prices normally in the order of several hundred pounds, commercial systems can be prohibitively expensive for many biological recorders. The price might be worth paying for some if the taxonomic coverage and quality of the keys available closely match their needs, but with the market still at a relatively immature stage, this is frequently not the case. If the primary purpose is to develop new resources, perhaps for others to use, and costs can be covered or recouped, then this may be less of an issue.

Not all of the non-commercial tools listed in table 1 can be easily obtained, for example those developed through funded programmes might not be maintained beyond the life of those programmes. In addition the systems were developed for a diverse range of technical environments, e.g. 3I Interactive must be implemented on a web server (Dmitriev, 2006) and Biowikifarm Wiki Keys operate on the MediaWiki platform (Hagedorn et al., 2010a). These kinds of technical environments are more accessible to teams engaged in taxonomic research and development than most biological recorders and they are clearly developed with that kind of audience in mind.

The level of technical support offered (e.g. documentation, forums and personal help), varies considerably amongst the tools listed. A certain standard of technical support normally accompanies a commercial product but, especially for non-commercial products, it is best not to take anything for granted. Someone wishing to learn how to use free software should be prepared to do some digging around first in order to see what kind of support is available to them.

The availability of existing keys is of major importance to someone who wants to use interactive keys for their own biological recording and, realistically, they will probably be steered towards a particular product by the availability of a key that interests them. A related issue is the extent to which the software and the keys (the knowledge bases) are independent of each other. In other words, the degree to which the software is capable of using keys developed on other systems and *visa versa*. This is particularly relevant for someone who wants to develop new keys since they may wish the products of their work to be available via software other than that which they used to develop it. There is a common standard for encoding knowledge bases for interactive keys called Structured Data Description (SDD) – discussed more later – and the extent to which software supports SDD (indicated in table 1) gives a good indication of how independent the software and keys are from each other.

Surprisingly, given the dominant position that printed keys still enjoy, developments in the area of using computers to aid the production of paper keys have been few and far between. This lack of progress was noted by Edwards and Morse (1995) and little seems to have changed since. Mike Dallwitz's original Key program is still available but, it seems, not in a version that will run on modern 64 bit Windows computers. A promising recent development called XKey (Delgadocalvoflores et al., 2006) appears to have sunk without trace. So even though these types of tool have the longest history in the development of computer-based biological identification, today there does not appear to be an accessible tool (free or commercial) whose purpose is to enable quality printed keys to be flexibly produced (e.g. in both yoked and bracketed styles) from an underlying knowledge base.

There have been a number of developments around interactive single-access keys, including Linnaeus II, FRIDA, OpenKey and Biowikifarm Wiki Keys (table 1). In general, these systems, although mostly non-commercial, are less accessible than some of those for interactive multi-access keys.

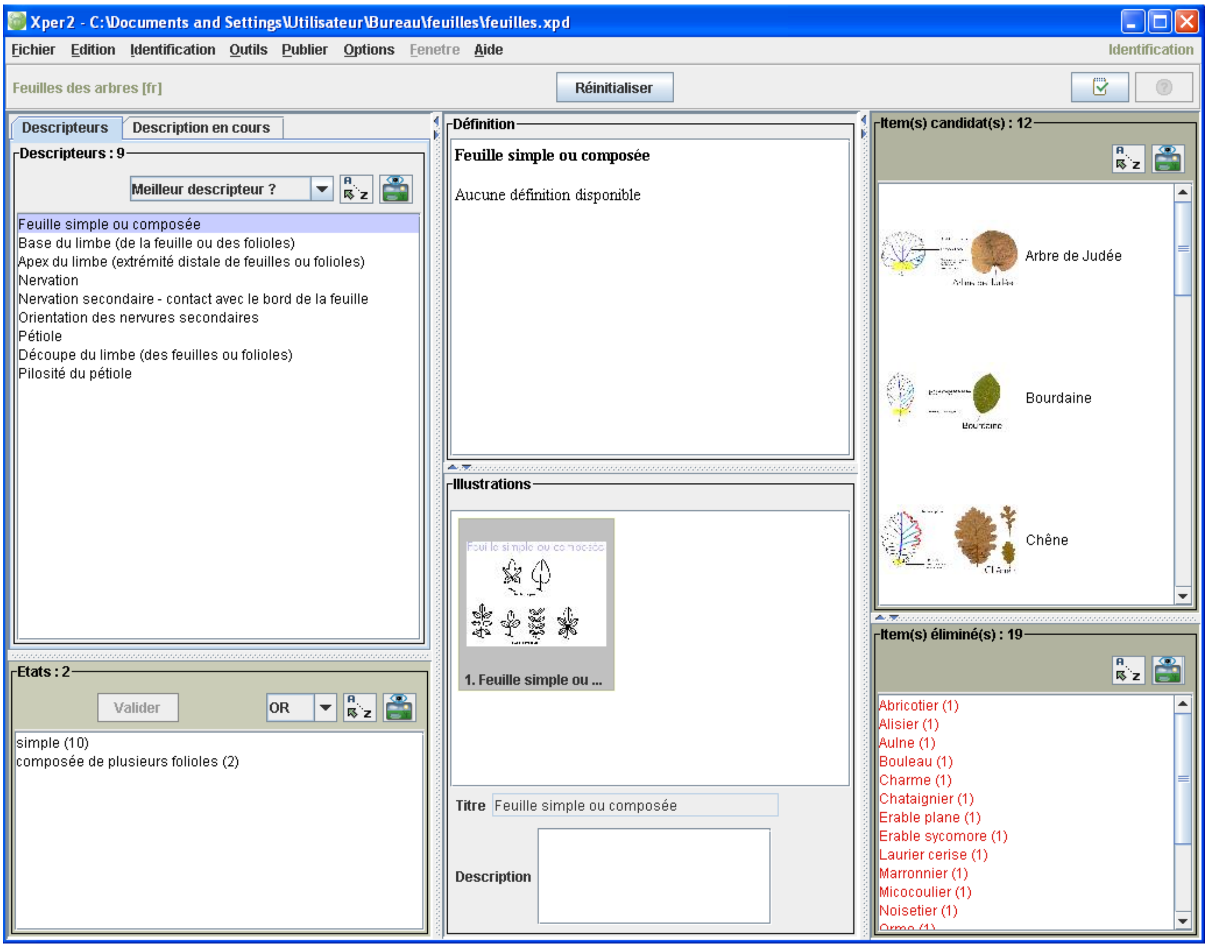
Interactive multi-access keys are currently receiving most attention, perhaps understandably so since they can realise the greatest advantages of computer-based biological identification. Reasonably well-established systems like Delta Intkey and Lucid will be familiar to some and other significant developments include 3I Interactive keys, Xper^2 and iSpot Bayesian keys (table 1). A number of these developments, e.g. Delta Intkey, Xper^2 and iSpot, are non-commercial and provide accessible platforms for those interested in exploring the technology.

4.5.2 Examples of accessible tools: Xper^2

As an example of an accessible program for creating and running interactive multi-access keys, Xper^2 is worth describing in more detail.[1] Xper^2 is developed by the Laboratoire Informatique et Systématique – a consortium of biologists and computer specialists, based at the University Pierre et Marie Curie, who develop technology and keys for interactive biological identification (Ung et al., 2010; LIS, 2013a). It is completely free to download and runs as a standalone program on Windows, MacOS and Linux. The program supports several languages including French, English, Spanish, Chinese and Portuguese. Existing keys (knowledge bases) are available for download (LIS, 2013b) and although this is currently quite a small selection, the software supports import and export of knowledge bases encoded in the SDD format, so there are many more which could, potentially, be used. There is a reasonably comprehensive set of written guidance including FAQs and tutorials and a mailing list (note that many contributions to the latter are in French).

---

[1] Subsequent to the production of this document in August 2013, a new internet-based implementation of Xper was released – Xper^3: http://www.xper3.com/



Feuille simple ou composée
Base du limbe (de la feuille ou des folioles)
Apex du limbe (extrémité distale de feuilles ou folioles)
Nervation
Nervation secondaire - contact avec le bord de la feuille
Orientation des nervures secondaires
Pétiole
Découpe du limbe (des feuilles ou folioles)
Pilosité du pétiole

Feuille simple ou composée

Aucune définition disponible

Titre: Feuille simple ou composée

simple (10)
composée de plusieurs folioles (2)

Arbre de Judée
Bourdaine
Chêne

Abricotier (1)
Alisier (1)
Aulne (1)
Bouleau (1)
Charme (1)
Chataignier (1)
Erable plane (1)
Erable sycomore (1)
Laurier cerise (1)
Marronnier (1)
Micocoulier (1)
Noisetier (1)

Figure 1. Screenshot of Xper² demonstrating an interface that is typical of a number of interactive multi-access key programs including Delta Intkey and Lucid Player.

Figure 1 shows a screenshot of the Xper² program running in identification mode (as opposed to editing mode). This kind of interface is typical of several high-profile interactive multi-access key programs that run as standalone computer programs including Delta Intkey and Lucid Player. In all three of these programs the main window includes four areas: one for character selection, one for specifying character state, one to display a list of candidate taxa and one to display a list of eliminated taxa. (In figure 1 these areas are top-left, bottom-left, top-right and bottom-right respectively).

Xper² has two central areas; one for showing 'definitions' (top-middle) and one showing images (bottom-middle). If a user clicks on a character or a character state, then definitions of these characters/states (if provided by the key) are displayed in the definition area. If a user clicks on a taxon, the full description of that taxon is displayed in the definition area. Similarly, images associated by the key with characters, character states, or taxa are displayed in the image area whenever the user selects a relevant item.

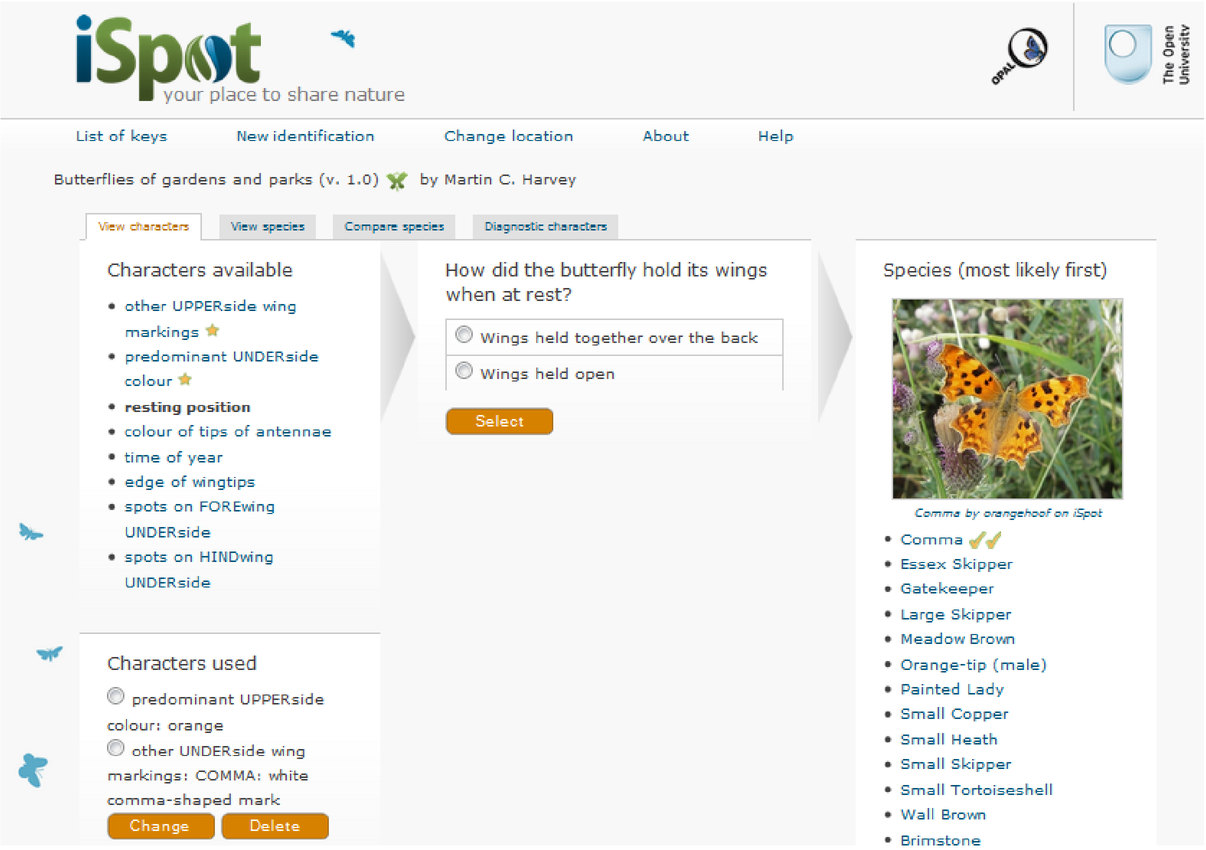
The basic procedure when using Xper² for identification, which is similar to that for several programs, is straight forward. The user first opens the relevant knowledge base (key) using the File menu. Initially, all the taxa described by the key are listed in the candidate taxa area. Now the user selects a character from those displayed in the character selection area. (The order in which the characters are listed is determined by selecting from a number of options, one of which is 'best character' which lists them in order of their discriminating power.) When a character is chosen, the various options for specifying a relevant state for it are shown in the character state area. When the user specifies a state for the character and clicks the 'Submit' button, any taxa not consistent with that character state are removed from the candidate taxa list and added to the eliminated taxa list. Specification of states for different characters continues until only one taxon is left in the candidate taxa area, or the user is unable to specify the states of any more discriminating characters.

At any stage, the user can change the state – or unset the state – of any character previously entered. A taxon is only removed from the candidate list once the number of mismatches with the specified character states exceeds a user-specified threshold. By default, this threshold is zero but it can be easily changed. This allows a degree of error tolerance. The number of mismatches for all taxa – from both candidate and eliminated lists – is always displayed in parentheses after the name of the taxon in the list.

When run in 'edit' mode, Xper^2 can be used to edit existing knowledge bases or create new ones from scratch. A short tutorial is available to guide the user in the creation of new knowledge bases and the tool is very well-constructed, intuitive and easy to use for this. Knowledge bases available in SDD format can be imported an Xper^2 knowledge based can be exported to SDD (import and export available from the File menu).

4.5.3 Examples of accessible tools: iSpot Bayesian keys

The iSpot Bayesian keys facility is a web-based implementation of multi-access interactive keys developed and maintained by a team based at the Open University (originally funded through the Open Air Laboratories – OPAL – project). As a web-based implementation, it is accessible from many different platforms (including web-browsers on smart-phones). A number of existing keys are available from the website and potential authors are encouraged to create and submit new ones. There is no key-building software; instead, knowledge bases are constructed as taxon-character matrices, to a defined format, using a spreadsheet and submitted to iSpot for conversion to a format that the iSpot Bayesian key software can use. There is currently no SDD import/export functionality, so the ability to run iSpot keys on other software (e.g. Xper^2) is limited. There are online instructions on how to use the keys, though guidance on their construction is currently hard to find.



Butterflies of gardens and parks (v. 1.0) by Martin C. Harvey

Characters available

- other UPPERside wing markings ★
- predominant UNDERside colour ★
- **resting position**
- colour of tips of antennae
- time of year
- edge of wingtips
- spots on FOREwing UNDERside
- spots on HINDwing UNDERside

How did the butterfly hold its wings when at rest?

○ Wings held together over the back

○ Wings held open

Characters used

○ predominant UPPERside colour: orange

○ other UNDERside wing markings: COMMA: white comma-shaped mark

Species (most likely first)

Comma by orangehoof on iSpot

- Comma ✓✓
- Essex Skipper
- Gatekeeper
- Large Skipper
- Meadow Brown
- Orange-tip (male)
- Painted Lady
- Small Copper
- Small Heath
- Small Skipper
- Small Tortoiseshell
- Wall Brown
- Brimstone

Figure 2. Screenshot from a web-browser showing an iSpot Bayesian key for common butterflies.

In the screenshot of an iSpot interactive key in figure 2, four main areas are shown on the screen. Two areas on the left show the characters against which the user can specify states (top) and characters for which states have already been specified (bottom). In the middle area the user can specify a state for the currently selected character. At any time during the identification process, the user can change or unset the state for a character previously specified. So far, this is similar to the previous Xper^2 example but the main difference lies in the list of candidate taxa on the right. Rather than have two lists – one for candidate taxa and one for eliminated taxa – iSpot only has a single list; taxa are never completely eliminated to a separate list but are instead ranked according to the probability that they match the specimen described by the user. Those with particularly high probabilities are marked with anything from one to four ticks (two in the example in figure 2) – the more ticks, the higher the probability.

In iSpot the probabilities are calculated using Bayesian statistics (Rosewell and Edwards, 2009). One interesting feature is that the ‘prior probabilities’ for each candidate taxa can be adjusted based on data from the NBN Gateway describing their known distribution and abundance if the user chooses to specify a location for their specimen (which can be done by clicking on a map).

The procedure for using an iSpot key is not dissimilar to that previously described for Xper^2. Characters are selected, in any order, and states for them specified. iSpot adjusts the ranking (and the probability ticks) of the candidate taxa list as each character state is specified, modified or removed. The identification process is complete, at the user’s discretion, when there is a candidate taxa with a high probability of matching the specimen, or when the user can add no further relevant character states.

4.6 Common standards, interoperability and development models

The data description format adopted to encode taxonomic knowledge bases for use as, or for the generation of, biological identification keys on computers is critically important. The technical details of these formats aren't that important here; of greater concern is how widely they are adopted and the ease with which knowledge bases can be used by different computer programs – an aspect of 'interoperability'. History shows that where technical development coalesces around common standards for data encoding and data exchange – i.e. where there is more interoperability – technical innovation is more effective (and cost-effective). But history also shows that business and publishing models can struggle to migrate to online environments and there will be challenges in establishing models that support the development and publishing of high-quality computer-based biological identification resources.

4.6.1 Separating software and the knowledge-bases

One benefit of software architectures that make a clear distinction between knowledge-bases and programs that create and/or use them has already been mentioned in relation to generating printed keys with different layouts, e.g. yoked and bracketed, from the same underlying knowledge base. Furthermore, knowledge bases encoded in standard formats can be used across different software platforms, making them more widely accessible and useful.

But another benefit of encouraging a broad community of biological identification resource developers to support a standard format for encoding knowledge bases is that it encourages innovation. This may seem counter-intuitive, but many great innovations result from finding new ways to visualise, analyse or otherwise explore *existing* data. Single software products rarely, if ever, have the range of functionality or breadth of vision to leverage all the possible benefits from an information-rich knowledge base. Each product has its own strengths and by representing a knowledge-base in a different way, for example through a different graphics interface, each can synthesise different information from the same knowledge base.

The incentive for developers to innovate new ways of extracting and representing information from knowledge bases is greater if there is a large pool of them – encoded to common standards – on which new products can operate. And from the point of view of the creator of knowledge bases – the keys – there can be a greater incentive to create them if the products of their labour are available through a wider range of tools and therefore to a larger audience.

A comparison of some of the more accessible multi-access biological identification tools currently available (e.g. Xper^2, Lucid and Delta Intkey) shows that the basic graphical user interfaces are remarkably similar and none are very sophisticated or engaging in their use of graphics, demonstrating that there is always room for new innovation. A new product, or an enhancement to an existing product, offering a different graphical experience for a user or, for example, facilities to compare groups of images or taxonomic descriptions more flexibly, would access information from existing knowledge bases which is not currently exploited. And the impact of a new innovation of this kind is directly related to the number of knowledge bases it can exploit.

4.6.2 The Descriptive Language for Taxonomy (DELTA) format

The DELTA data format underpins the influential suite of programs for computer-based biological identification developed by Mike Dallwitz and his colleagues at CSIRO since the 1970s like Key and Intkey (e.g. Dallwitz, 1974, 1992; Dallwitz et al., 2007; Dallwitz, 2011; Dallwitz et al., 2013). Intkey, in particular, has been a popular program and as a result there are many keys available in the DELTA format. DELTA has also been used, to an extent, by the developers of other software like Navikey and Pankey (table 1) .

In 1991, DELTA was adopted as a standard by the Taxonomic Database Working Group (TDWG) – now also operating under the banner of Biodiversity Information Standards – which is a working group of the International Union of Biological Sciences (IUBS) (TDWG, 2007). The format is open and published on the TDWG website. However because the DELTA format was so closely tied to a particular suite of software, the TDWG found that there were barriers to implementing suggested improvements to the format. In addition, new developments – such as Lucid – continued to develop new knowledge base encoding systems to meet their own needs. In the late 1990s the TDWG started to consult on the development of an independent standard which they called Structured Data Description (SDD).

4.6.3 Structured Data Description (SDD)

TDWG (2007) states: “*The goal of the SDD standard is to allow capture, transport, caching and archiving of [taxonomic] descriptive data [...], using a platform- and application-independent, international standard. Such a standard is crucial to enabling lossless porting of data between existing and future software platforms including identification, data-mining and analysis tools, and federated databases*.”

SDD has been designed as a ‘superset’ capable of meeting the data encoding requirements of a diverse suite of applications and functions. SDD is, of course, capable of encoding knowledge-bases in the form of taxon-character matrices that are fundamental to implementing multi-access keys, but it is also capable of directly, and fully, encoding a knowledge-base describing a dichotomous key that is not supported by a taxon-character matrix. It can therefore support software that works with multi-access keys (Ung et al., 2010; Walter and Winterton, 2007) and software that is designed to print, or interactively run, single-access keys (Delgadocalvoflores et al., 2006; Burguiere, 2013).

SDD is an implementation of the extremely widely used XML (Extensible Markup Language) standard and, as a result, programmers of many different platforms and programming languages – virtually all of which support XML – can work with it very easily. In an ideal world, all biological identification programs would be ‘native’ users of SDD – in other words they would all read from and write to it directly. However, products under active development often need to go beyond what is possible using a standard format like SDD. A standard like SDD is continually evolving and in many cases will eventually accommodate such new requirements, but the process of changing a consensus-built standard is, by nature, quite slow. So, in reality, most programs are built on their own data encoding formats which can be changed more quickly. But the benefits of common standards are such that many programs support *conversion* to and from their own formats and standard formats like SDD – so facilitating the interchange of knowledge bases between different programs. This appears to be the developing situation with SDD, with many of the major developments – particularly the more recent ones – supporting import and/or export of data in SDD format (see table 1).

4.6.4 Development models

One reason for the lack of progress in computer-based biological identification over the last 40 years is that many developments and innovations have proceeded in isolation without reference to each other. Common standards for knowledge-base encoding and data exchange will help to overcome this problem, but faster progress will be made if a general spirit of collaboration and openness can be fostered. There are some encouraging signs that things have started to move in this direction, particularly in Europe.

The KeyToNature programme which ran between 2007 and 2010 (funded by the EU eContentplus programme) aimed to develop a range of new, paper-free identification tools, for use within schools and universities across Europe and one of the key objectives was to foster interoperability. Interoperability and collaborative development are themes which reoccurred through many of the projects that contributed to KeyToNature (Nimis and Lebbe, 2010). Some of the developments were Open Source – meaning that the source code is freely available – including Biowikifarm Wiki Keys (Hagedorn et al., 2010a) and Open Key Developer (van Spronsen et al., 2010) – and many others were designed to complement existing developments and platforms. The objectives of another major EU funded programme – ViBRANT – running between 2010 and 2013 include "*the provision of analytical services for users to build identification keys*" within a wider open and integrated environment for biodiversity research (Roberts and Smith, 2010).

Taxonomy and systematics is a collaborative venture and the manifold opportunities offered by eTaxonomy to increase both the degree and effectiveness of collaboration are bringing about a real resurgence of the field. One only has to look at list of contributors and collaborators on important facets of the eTaxonomy environment, such as Encycolpedia of Life (EOL, 2013) and Catalogue of Life (Catalogue of Life, 2013), to appreciate how important interoperability and openness are. New developments and innovations cannot exist in a vacuum in this new world – the degree to which they interoperate with, and add value to, existing initiatives is a key feature on which their success hinges. Developments in computer-based biological identification are not excepted from these influences – interoperability is key. New developments have most chance of success if they add value to the way in which biological recorders current work – not demand that they throw out everything they have and start again.

But seismic changes such as those happening in the field of taxonomy and systematics never occur painlessly. The fact that new eTaxonomy paradigms have not always been universally welcomed, reflects, in part, the inertia of existing systems that have evolved over a century or more. These well-established systems include mechanisms that recognise and reward professional (and to some extent amateur) taxonomists and systematists for their work, for example by the publication of taxonomic revision monographs in peer-reviewed journals. To be successful, the eTaxonomy paradigm needs to incorporate acceptable mechanisms for recognising, acknowledging and rewarding people's work (Mayo et al., 2008).

A tension of this sort is evident between the paper and electronic biological identification resource publishing models. It costs – in time or money or both – to develop, test and publish high-quality biological identification resources, whether they are paper- or computer-based. Traditional paper-based publishing models have evolved to cover these costs (and even make a profit), as well as recognise the contributions of authors. Developing successful publishing models for computer-based biological identification that enable authors and publishers to cover costs of developing high-quality resources, and be recognised for this work, will be a challenge (Hominick and Schalk, 2010). On the face of it, interoperability and development models such as Open Source may seem to work counter to this objective, but the development of e-publishing in general (including books, music and other resources) shows that it is possible to find ways to accommodate successful publishing models in an open environment.

In the UK there is a lack of leadership in terms of promoting the collaborative environment required to really advance the field of computer-based biological identification. In particular there doesn’t seem to be any significant support coming from the UK for the SDD standard for knowledge-base encoding or interchange. Arguably the most significant UK development in the field – iSpot Bayesian keys – doesn’t currently support SDD.

4.7 Platforms for computer-based biological identification

Over recent years, an important development for computer-based biological identification has been the proliferation and diversification of hardware platforms on which resources can be delivered to users – particularly mobile devices. A perceived barrier to the development and use of computer-based biological identification resources in the past has been that, to be effective, many of them need to be used in the field. Now that portable devices like smart-phones and tablet computers have become so highly functional and ubiquitous, this can no longer be considered a serious barrier (Williams et al., 2011).

It is worth remembering that this barrier cannot, on its own, have been responsible for the slow development of computer-based biological identification since many of its applications are not field-based. And despite the current tendency to focus on biological identification resources that can be delivered to people in the field (Araya, 2013) – which is partly technology-led (Snaddon et al., 2012) and partly a response to the desire to nurture more ‘citizen science’ (Roy et al., 2012) – there remains great untapped potential for computer-based biological identification which is best delivered back in the lab after fieldwork has been completed.

Consider invertebrates which are so diverse, with so many cryptic and critical taxa, that for the majority of groups it is necessary to collect specimens and identify them at a desk indoors where specialist equipment like a microscope can be brought to bear. Accessing a computer and a microscope from a single desk has been possible from the early days of home-based computing and has been particularly easy since laptop computers became widely available a couple of decades ago. And yet despite the well-known advantages of computer-based multi-access keys for biological identification, especially when compared to complex dichotomous keys or specialist texts such as are used for invertebrate identification (boxes 1-3), many new identification resources for invertebrates continue to be published as paper-based single-access keys. The technology to deliver these kinds of resources to biological recorders more effectively on desktop/laptop computers adjacent to their microscopes is relatively well established (table 1).

How users may wish to use a particular biological identification resource is a major consideration for developers and publishers. For example if the resource needs to be used primarily in the field by occasional users, then there’s not much point in developing a resource that only runs on operating systems which are primarily for desktop/laptops such as Windows or Mac OS. For a product like that, Android, iOS or web-based delivery may be more appropriate in order to deliver it on a smart-phone or tablet. On the other hand, a resource which will only have real utility away from the field could be developed and published in formats suitable for a desktop/laptop platform.

An advantage of promoting a community-wide common standard for knowledge-based encoding and/or interchange – such as SDD – is that resources can be developed without undue reference to the platforms on which they may be delivered, potentially circumventing some of the operating system interoperability barriers which inevitably exist in a commercial environment.

Another way of circumventing these barriers is to produce biological identification resources – whether developed to SDD standards or otherwise – that can be delivered over the internet since practically every operating system supports a web browser. This has been the strategy of the iSpot Bayesian keys initiative (Open University, 2013a). Any platform with a web-browser and an internet connection can access iSpot keys. The small screen of many smart-phones can make using web-pages designed for bigger screens a problem, but many websites (iSpot Bayesian keys included) overcome this by providing versions of web-pages written specifically for delivery on smaller displays. A disadvantage, for some field situations, is that an internet connection is required.

5 Other tools for computer-based biological identification

Although this paper has focussed on computer-based *keys* for biological identification, computer-based resources for biological identification encompass much more than keys. A brief overview is given here.

'Electronic Field Guides' distributed on CDs, and latterly DVDs, have been around for a couple of decades. A great advantage of these media over traditional printed field guides include fewer space restrictions (so more images and/or text can be included) and the ability to accommodate a greater variety of media such as video and sound – which can have such great utility in biological identification. However despite these advantages, as with computer-based keys, computer-based field guides have not done as well as anticipated (Morrison, 2011; Hominick and Schalk, 2010). Hominick and Schalk (2010) predicted that "*for the foreseeable future, books will outsell e-products even for the same title: books have the clear advantage in portability, comfort of handling, familiarity, shelf life and identifiable prestige on the shelf.*" However, the same authors also note "*the trend in scientific publishing is towards smaller print runs, pitched to a known market and then straight to Publish on Demand. E-products are ideal for this trend, as they can be produced easily at little cost in small numbers, and they can also be updated easily*." And, even since the author's comments in 2010, the paper book's advantages of 'portability', 'comfort of handling' and 'familiarity' are fast disappearing under a wave of popular technological innovation around portable devices like smart-phones, tablets and eBook readers.

In the last decade, rapid technical advances in digital photography – that show no signs of slowing – have resulted in a number of innovations. Some naturalists have created large galleries of good quality images which they share through web-based platforms like Flickr (Flickr, 2013) and Ipernity (ipernity, 2013) and other naturalists can use these as for reference. For example Steven Falk has published some comprehensive galleries of invertebrates, including hoverflies, on his Flickr site (Falk, 2013) and Ian Smith has published images of marine invertebrates on Flickr which are annotated to show diagnostic identification features (Smith, 2013). Many others share images on forums like Facebook (Facebook, 2013), Yahoo Groups (Yahoo, 2013) and Google Groups (Google, 2013), often using their contacts to suggest identifications for the organisms they have photographed. The iSpot website (iSpot, 2013), takes this a step further with a rich architecture designed specifically around the concept of people sharing photographs of wildlife and identifying them for each other.

It's never been easier to publish content-rich websites, blogs etc and it is a rare national recording scheme or society which does not now have a strong web presence and their websites often include a number of resources for biological identification. For example the website of the Bees, Wasps and Ants Recording Society (BWARS, 2013) includes an excellent series of species accounts which include high quality images, distribution maps and information on identification. The standard of the content on such sites is improving all the time and often incorporates external information such as NBN distribution maps which are delivered 'live' to the website through NBN Gateway web services (the BWARS website is an example of this). Many individual naturalists also often publish useful web-based identification resources, like the widely-known resource list of Richard Comment published on his blog (Comment, 2013).

General levels of ICT skills are improving throughout the population and wherever groups of naturalists join together in self-support groups, there is usually one or more people who can support that group with some ICT expertise. As a result, even small local groups often create biological identification resources tailored to the interests of their group members, and often their own geographical area, using standard technology like word processors and spreadsheets. Good examples include the highly regarded document 'Pugs of Lancashire and Cumbria' produced for moth recorders in Lancashire and Cumbria by Lancashire Moth Group member Brian Hancock (Hancock, 2012). Another example is the Excel spreadsheet produced by Nigel Cain-Honeysett of the Shropshire Spider Group which is, essentially, a multi-access key for groups of Linyphiid spiders (money spiders) drawn from a knowledge base published in The Spiders of Great Britain and Ireland (Roberts, 1985) which is rather cumbersome to use in its original paper format.

The rapidly developing tablet and smart-phone 'apps' market – so vividly described by Williams et al. (2011) in relation to the field of chemistry and drug discovery – will blur the line between biological identification and other aspects of the field of ecology. It is already happening; for example the iRecord Ladybirds app released in 2013 is both a ladybird ID tool and recording device – utilising a smart-phone's camera and GPS – all rolled into one. Snaddon et al. (2012) paints a picture of rapid technological change driving the way we do science and interact with nature. An environment of innovation that encourages 'mash-ups' between different technologies and knowledge bases is one that favours common standards for data exchange and this likely scenario is another powerful argument for promoting such standards in field of computer-based biological identification.

6 Computer-based tools for mentoring, developing and networking biological recorders

Social Media is: "*a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content*" (Kaplan and Haenlein, 2010). Web 2.0 is an umbrella term for the technologies which moved the web away from 'static' content (fixed HTML pages) to more 'dynamic' content tailored in response to user's actions. It was the evolution of Web 2.0 technologies that put internet users in the driving seat as content creators – rather than only passive consumers – ultimately giving rise to Social Media where people can creatively exchange information and ideas (User Generated

Content). The Social Media applications currently at the forefront of this revolution include Facebook (Facebook, 2013), Twitter (Twitter, 2013), Wikipedia (Wikipedia, 2013) and YouTube (YouTube, 2013), but there are many more facilities and tools besides these.

On-line communities of naturalists and biological recorders existed before the advent of Web 2.0; for example the popular Yahoo Group forums (Yahoo, 2013) – based around the pre-Web 2.0 technologies of bulletin boards and email-list servers – have hosted many discussion groups for naturalists, particularly centred on local field groups and societies. But there is every indication that the level of participation in online forums by naturalists and biological recorders is increasing dramatically with the advent of Social Media. This is a welcome opportunity at a time when the traditional model of the field naturalist groups and societies is perceived to be in decline, especially when judged by the demographics of their memberships (Scottish Natural Heritage, 2010; Boxshall and Self, 2011). If used imaginatively, and in an integrated way, Social Media could help to encourage new members of these groups – especially younger members.

The FSC Biodiversity Fellowship programme (FSC, 2013) – a programme of training events in ID and survey skills aimed at some of the under-recorded taxonomic groups – has been supported by a closed Facebook Group where individuals who registered as ‘biodiversity fellows’ could interact with each other and with the FSC. Of some 400+ registered biodiversity fellows, 140 joined the Facebook Group, which is a considerable proportion (although we should forget that something like 260 didn’t). As well as providing the FSC with a fast way of reaching people with relevant news (e.g. when new subsidised course places became available), the group quickly developed into a self-help forum where biodiversity fellows could give each other the benefit of their considerable collective experience. It demonstrates the potential value of such groups, but also gives rise to many questions like ‘what differentiates this group from other similar groups?’ (e.g. the Facebook group of the National Forum for Biological Recorders) and ‘how is it best to communicate with those biodiversity fellows – the majority – who did not join the Facebook group?’

The need to offer effective mentoring of new and improving biological recorders is widely acknowledged (Boxshall and Self, 2011; Roy et al., 2012; Scottish Natural Heritage, 2010). There’s a need to encourage more biological recorders but, at the same time, there are fewer and fewer experts around to mentor them. Social networking and Virtual Learning Environments (VLEs) could help to address this mismatch by providing more opportunities for expert and new/improving biological recorders to interact – even if they are geographically remote from each other. VLEs include the open-source Moodle (Martín-Blas and Serrano-Fernández, 2009; Dougiamas and Taylor, 2003) and commercial products like WebCT. In 2013 the BSBI piloted an innovative training project that includes a VLE at its heart (Sue Townsend, personal communication, August 2013). This pilot project marketed as a ‘plant identification course for beginners in serious botany’, is implemented as a distance learning package and distribution and marking of assignments was accomplished using Moodle. In addition, trainees were also assigned to tutors living in the same general geographic area as them. Initial indications are that the mix of personal mentoring and the VLE worked well.

The Scratchpads project run by the Natural History Museum (Smith et al., 2009, 2011) is a Virtual Research Environment (VRE) relying heavily on Web 2.0 technologies and incorporating some features found in social networking sites. Scratchpads are websites put together flexibly from a large collection of building blocks designed to facilitate eTaxonomy. Smith et al. (2009) describes the environment as: "*a data-publishing framework for groups of people to create their own social networks supporting natural history science*". Scratchpads were adopted as the architectural framework for the EU-funded ViBRANT project (Roberts and Smith, 2010) and there is a considerable amount of innovative work going on around the platform as a result of that project. For example Burguiere (2013) reported on a development called IKey+ which will produce a plugin for the Scratchpad platform to allow the integration of single-access interactive keys.

Because of the focus around biodiversity and taxonomy and the incorporation of tools which facilitate social networking, Scratchpads have potential to act as platforms through which mentoring and development of biological recording skills could happen. To date, much of the focus has been on creating VRE's for professional taxonomists and systematists, but the platform is open and available for use by collectors of primary biodiversity data, including biological recorders. A number of UK recording schemes have implemented websites on the Scratchpads platform, for example the Tachinid Recording Scheme (Raper and Smith, 2013). The Tomorrow's Biodiversity Project has a scratchpad website that will be used to explore the facilities offered by the platform and, quite likely, as a forum to engage with collaborators and people/organisations that we consult with (Burkmar, 2013).

The iSpot project is probably the most ambitious and feature-rich online social forum developed specifically for biological recorders (iSpot, 2013). It was originally established as part of the OPAL project (Imperial College London, 2013), but is now supported by the Open Science Laboratory (Open University, 2013b). At its core, iSpot is a website where any registered user (registration is free) can upload images of wildlife and seek help from other users of the site to identify the species they have photographed (McAndrew et al., 2010; Dodd, 2013; Silvertown, 2009). To be effective, iSpot relies on a large interacting community of users and now has over 20,000 registered users and 160,000 observations of 7,000 species – 80% of which get identified within 24 hours of being posted on the site (Snaddon et al., 2012; Dodd, 2013). Note that iSpot is aimed primarily at encouraging new biological recorders and helping with identifications from photographs (at any level), but because some much information can be added with a photograph, there is the potential to generate primary biodiversity data – biological records – from much of the information entered, and the iSpot team are developing relationships with some schemes and societies to make that happen.

A major feature of iSpot which has contributed to its success is its reputation system (Dodd, 2013) described as 'state of the art' by Rotman et al. (2012). Certain experts who use the system are pre-assigned reputation levels, but all users gain credit for identifications they make and gradually build a publicly visible reputation level for specific taxonomic groups. Reputation is gained by a user when another user agrees with an identification they have made. The reputational value of an agreement depends on the reputation of the user agreeing with the identification. Any user of iSpot can make an identification of another person's photograph or agree with an identification. The upshot is that easy identifications, of the type required by complete novices, can be made by any iSpot user with sufficient experience and means that experts can confine the time they spend on iSpot to helping with more difficult identifications.

Other iSpot features that enhance the community and mentoring aspects of the site include forums (basic 'traditional' forums where users can ask questions of each other or otherwise interact) and 'badging'. The badging facility allows users to associate themselves with particular groups and organisations such as recording schemes and societies. When a user is badged to a particular organisation, a small icon representing that organisation is displayed next to their name. A user can have more than one badge. When a user requests to be badged to a particular organisation, iSpot administrators seek approval from that organisation.

iSpot Bayesian keys have been described elsewhere in this document, but one feature considered by the iSpot team, and relevant to the discussion here, is to provide a facility for any iSpot user to create and upload Bayesian keys to the iSpot website in association with a user-moderated system for rating these keys, somewhat akin to the current reputational system for identification expertise (iSpot team, personal communication, 16th July 2013). It is the application of a rating system that makes this such an interesting proposition. If any user can upload a key, then it is easy to imagine hundreds – or even thousands – of keys becoming available. This proliferation of keys, with no simple way of assessing their effectiveness, might serve more to confuse than help potential users. However, with the addition of a rating system, one can imagine a system where good keys are highlighted to potential users, climbing the 'charts' as they gain a better reputation. In addition, the ratings earned by keys could provide reputational kudos for their creators and be a significant incentive to create or improve keys. It is an interesting and innovative potential application of computer-based social networking for biological recording.

7 Summary of main points

Two features of this review stand out above everything else: firstly the range of techniques and resources for biological identification is increasing dramatically and, secondly, to make a meaningful contribution to the rapidly changing field of taxonomy – eTaxonomy – interoperability of new tools and resources is key. Taxonomy is a collaborative venture and developments which don't facilitate collaboration are wasteful. The outputs of biological recorders (their records) and the tools they use for biological identification are not exempt from this maxim.

The degree to which the elements of eTaxonomy are interoperable and interrelated is reflected in extent to which they harvest information from, or provide information to, other elements. The Encyclopedia of Life (EOL, 2013), exemplifies this with over 250 'content partners' including other major elements of the eTaxonomy framework like the Catalogue of Life (Catalogue of Life, 2013). Many of these partners provide content to EOL dynamically. For example, Flickr users can make selected images available for harvesting by EOL by adding certain 'machine tags' (e.g. the taxon name) to their images and sharing them with the Flickr EOL group. EOL can then harvest these images automatically.

EOL itself makes its content discoverable and usable by other tools and facilities via its Application Programming Interface (API). So, for example, any third party could create a website or program and dynamically populate it with information or images from EOL through use of its APIs. APIs, like support for data interchange standards, are a feature of interoperable tools. For example it is through APIs (implemented as 'web services') that the NBN Gateway enables third-party websites (e.g. those of schemes and societies) to dynamically embed 'live' distribution maps on their pages.

The following are key points to consider when developing new electronic tools or resources for use by, or to support the development, mentoring and networking of, biological recorders. Tools and resources should:

1. Capitalise on existing developments, including other tools, standards and content available through APIs.
2. Bring something new to the party!
3. Be interoperable and 'outward looking' themselves.
4. Facilitate the evolution of user's working methods rather than demanding a revolution.
5. Recognise the contribution of content providers/users.
6. Operate within a financially sustainable business environment.

Biological identification and, as a consequence, biological recording is on the cusp of a period of major and rapid change on the tails of a technological transformation already underway in the wider field of taxonomy & systematics. Accelerating advances in mobile computing and electronic publishing are helping to drive this transformation. These changes are fundamentally important to the delivery of the Tomorrow's Biodiversity project and are of general importance for the delivery of biodiversity education and resources within the FSC.

8 References

Agranat, I. (2013) "Bat species identification from zero crossing and full spectrum echolocation calls using Hidden Markov Models, Fisher scores, unsupervised clustering and balanced winnow pairwise classifiers." In **Proceedings of Meetings on Acoustics**. **2 June 2013**. Acoustical Society of America

Araya, Y. (2013) "There is an app for that" - The next level of ecological mobile technology. **BES Bulletin**, 44 (3): 36–38

Bach, B., Miller, L.A., Journal, S., et al. (1977) A Portable Ultrasonic Detection System for Recording Bat Cries in the Field. **Journal of Mammalogy**, 58 (2): 226–229

Becerra, J.M. and Valdecasas, A.G. (2004) Landmark superimposition for taxonomic identification. **Biological Journal of the Linnean Society**, 81 (2): 267–274

Bertolani, R., Rebecchi, L. and Cesari, M. (2010) "A model study for tardigrade identification." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 333–339

Boxshall, G. and Self, D. (2011) **UK Taxonomy & Systematics Review - 2010**

BRC (2013) **iRecord - manage and share your wildlife records** [online]. Available from: http://www.brc.ac.uk/irecord/ [Accessed 3 July 2013]

BTO (2012) **Recording birds on the go: BirdTrack goes Mobile** [online]. Available from: http://www.bto.org/news-events/news/2012-04/recording-birds-go-birdtrack-goes-mobile [Accessed 3 July 2013]

BTO (2013) **BirdTrack BTO/RSPB/BWI/SOC/WOS** [online]. Available from: http://blx1.bto.org/birdtrack/main/data-home.jsp [Accessed 3 July 2013]

Burguiere, T. (2013) IKey + : A New Single-Access Key Generation Web Service. **Systematic Biology**, 62 (1): 157–161

Burkmar, R. (2013) **Tomorrow's Biodiversity Project Scratchpad** [online]. Available from: http://tombio.myspecies.info/ [Accessed 14 August 2013]

BWARS (2013) **Welcome to BWARS | BWARS** [online]. Available from: http://www.bwars.com/ [Accessed 9 August 2013]

Casiraghi, M., Labra, M., Ferri, E., et al. (2010) "DNA barcoding: theoretical aspects and practical applications." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 269–273

Catalogue of Life (2013) **CLHome | Catalogue of Life** [online]. Available from: http://www.catalogueoflife.org/ [Accessed 5 August 2013]

CBOL Plant Working Group and Janzen, D.H. (2009) A DNA Barcode for Land Plants. **Proceedings of the National Academy of Sciences of the United States of America**, 106 (31): pp. 12794–12797

CEH (2013) **A news blog from the CEH News Team: A revolution in ladybird recording** [online]. Available from: http://cehsciencenews.blogspot.co.uk/2013/05/a-revolution-in-ladybird-recording.html [Accessed 3 July 2013]

Cheshire rECOrd (2013) **RODIS Home** [online]. Available from: http://www.rodis.co.uk/ [Accessed 3 July 2013]

Chesmore, D. and Yorke, L. (1997) "Proceedings of inaugural meeting held held at The University of Wales Cardiff 2-3 July 1997." In **Proceedings of inaugural meeting held held at The University of Wales Cardiff 2-3 July 1997**. **Cardiff**. **1997**. The BioNET-international group for computer-aided taxonomy (BIGCAT).

Commont, R. (2013) **Insect rambles: Web ID resources** [online]. Available from: http://insectrambles.blogspot.co.uk/2012/12/web-id-resources.html [Accessed 9 August 2013]

Cubelio, S.S., Bineesh, K.K., Raj, K., et al. (2010) "Biological identifications through mitochondrial and nuclear molecular markers: the case of commercially important crabs from Indian EEZ." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. p. 345

Dallwitz, M. (1992) A Comparison of Matrix-Based Taxonomic Identification Systems with Rule-Based Systems Xiong, F.L. (ed.). **Proceedings of IFAC workshop on expert systems in agriculture**. 66 (2) pp. 215–218

Dallwitz, M.J. (1974) A flexible computer program for generating identification keys. **Systematic Biology**, 23 (1): 50–57

Dallwitz, M.J. (2011) **Comparison of Interactive Identification Programs** [online]. Available from: ftp://delta-intkey.com/www/comparison.pdf [Accessed 15 August 2013]

Dallwitz, M.J., Paine, T.A. and Zurcher, E.J. (2007) **Interactive Identification Using the Internet** [online]. Available from: ftp://delta-intkey.com/www/netid.pdf [Accessed 15 August 2013]

Dallwitz, M.J., Paine, T.A. and Zurcher, E.J. (2013) **Principles of Interactive Keys** [online]. Available from: http://www.delta-intkey.com/www/interactivekeys.pdf [Accessed 15 August 2013]

Daou, A., Johnson, F., Wu, W., et al. (2012) A computational tool for automated large-scale analysis and measurement of bird-song syntax. **Journal of Neuroscience Methods**

Delgadocalvoflores, M., Fajardocontreras, W., Gibajagalindo, E., et al. (2006) XKey: A tool for the generation of identification keys. **Expert Systems with Applications**, 30 (2): 337–351

Dmitriev, D.A. (2006) 3I , a new program for creating Internet-accessible interactive keys and taxonomic databases and its application for taxonomy of Cicadina (Homoptera). **Russian Entomological Journal**, 15 (2): 263–268

Dodd, M. (2013) iSpot - helping everyone to identify and record wildlife. **АКТУАЛЬНЫЕ ПРОБЛЕМЫ ЭКОЛОГИИ И ЗДОРОВЬЯ ЧЕЛОВЕКА**, pp. 62–65

Dougiamas, M. and Taylor, P.C. (2003) “Moodle: Using Learning Communities to Create an Open Source Course Management System.” In Lassner, D. and McNaught, C. (eds.). **Proceedings of world conference on educational multimedia hypermedia and telecommunications**. **2003**. AACE. pp. 171–178

Ebach, M.C. and Holdrege, C. (2005) More Taxonomy , Not DNA Barcoding. **Bioscience**, 55 (10): 822–823

Edwards, M. and Morse, D.R. (1995) The potential for computer-aided identification in biodiversity research. **Trends in Ecology & Evolution**, 10 (4): 153–158

EOL (2013) **Encyclopedia of Life - Animals - Plants - Pictures & Information** [online]. Available from: http://eol.org/ [Accessed 5 August 2013]

Facebook (2013) **Facebook** [online]. Available from: https://www.facebook.com/ [Accessed 9 August 2013]

Falk, S. (2013) **Steven Falk’s collections on Flickr** [online]. Available from: http://www.flickr.com/photos/63075200@N07/collections/ [Accessed 9 August 2013]

Farr, D.F. (2006) On-Line Keys: More than Just Paper on the Web. **Taxon**, 55 (3): 589–596

Flickr (2013) **Welcome to Flickr - Photo Sharing** [online]. Available from: http://www.flickr.com/ [Accessed 9 August 2013]

FSC (2013) **Biodiversity Fellows - FSC** [online]. Available from: http://www.field-studies-council.org/supporting-you/fsc-projects/current-projects/biodiversity-fellows.aspx [Accessed 12 August 2013]

Galimberti, A., Martinoli, A., Russo, D., et al. (2010) “Molecular Identification of Italian Mouse-eared Bats (genus Myotis).” In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 289–294

Ganchev, T. and Potamitis, I. (2007) Automatic acoustic identification of singing insects. **Bioacoustics**, 16 (3): 281–328

Gaston, K.J. and O’Neill, M.A. (2004) Automated species identification: why not? **Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences**, 359 (1444): 655–667

GBIF (2013) **Gbif.org: Home Page** [online]. Available from: http://www.gbif.org/ [Accessed 5 August 2013]

Geng, H., Russell, J., Shin, B.-S., et al. (2012) “A flexible method for localisation and classification of footprints of small species.” In **Advances in Image and Video Technology**. Springer. pp. 274–286

Google (2013) **Google Groups** [online]. Available from: https://groups.google.com/ [Accessed 9 August 2013]

Griffing, L.R. (2011) Who invented the dichotomous key? Richard Waller’s watercolors of the herbs of Britain. **American journal of botany**, 98 (12): 1911–23

Grube, M. and Muggia, L. (2010) “Identifying algal symbionts in lichen symbioses.” In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 295–299

Hagedorn, G., Press, B., Hetzne, S., et al. (2010a) “A MediaWiki implementation of single-access keys.” In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 77–82

Hagedorn, G., Rambold, G. and Martellos, S. (2010b) “Types of identification keys.” In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 59–64

Hall, A. V. (1970) A computer-based system for forming identification keys. **Taxon**, 19 (1): 12–18

Hancock, B. (2012) **Pugs of Lancashire and Cumbria**

Hebert, P.D.N., Cywinska, A., Ball, S.L., et al. (2003) Biological identifications through DNA barcodes. **Proceedings. Biological sciences / The Royal Society**, 270 (1512): 313–21

Hominick, B. and Schalk, P. (2010) “Markets for biodiversity information products: real or imaginary?” In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 429–435

House of Lords Science & Technology Committee (2008) **Systematics and Taxonomy review 2008**

Hubbard, C.E. and Sampson, J. (1968) **Grasses: A Guide to Their Structure, Indentification, Uses, and Distribution in the British Isles**. Penguin books

IdentifyLife (2013) **Home | IdentifyLife** [online]. Available from: http://www2.identifylife.org/ [Accessed 5 August 2013]

Imperial College London (2013) **OPAL | Explore your local environment** [online]. Available from: http://www.opalexplorenature.org/ [Accessed 14 August 2013]

ipernity (2013) **ipernity: Welcome to ipernity - Photo and video sharing** [online]. Available from: http://www.ipernity.com/ [Accessed 9 August 2013]

iSpot (2013) **iSpot - your place to share nature** [online]. Available from: http://www.ispot.org.uk/ [Accessed 9 July 2013]

Ji, Y., Ashton, L., Pedley, S.M., et al. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. **Ecology letters**

JNCC (2013) **Recorder 6** [online]. Available from: http://jncc.defra.gov.uk/page-4592 [Accessed 3 July 2013]

Jones-Walters, L.M. (1989) Keys to the families of British spiders. **Field studies**, 7 (2): 365–443

Kaplan, A.M. and Haenlein, M. (2010) Users of the world, unite! The challenges and opportunities of Social Media. **Business Horizons**, 53 (1): 59–68

Lawrence, A. and Hawthorne, W. (2006) **Plant Identification: Creating User-Friendly Field Guides for Biodiversity Management**. Routledge

Legg, C.J. (1992) Random-access guide to Sedges of the British Isles using a microcomputer. **Field Studies**, 8: 31–57

Leggett, R. and Kirchoff, B.K. (2011) Image use in field guides and identification keys: review and recommendations. **AoB plants**

LIS (2013a) **Laboratoire Informatique et Systématique** [online]. Available from: http://lis.snv.jussieu.fr/lis/?q=en [Accessed 1 August 2013]

LIS (2013b) **Xper2 knowledge base** [online]. Available from: http://lis-upmc.snv.jussieu.fr/xper2/infosXper2Bases/en/index.php [Accessed 1 August 2013]

Living Record (2013) **LivingRecord.net** [online]. Available from: http://www.livingrecord.net/extras/Leaflet01.pdf [Accessed 3 July 2013]

MapMate (2013) **MapMate - UK's favourite biological recording software.** [online]. Available from: http://www.mapmate.co.uk/ [Accessed 3 July 2013]

Martín-Blas, T. and Serrano-Fernández, A. (2009) The role of new technologies in the learning process: Moodle as a teaching tool in Physics. **Computers & Education**, 52 (1): 35–44

Mayo, M. and Watson, A.T. (2007) Automatic species identification of live moths. **Knowledge-Based Systems**, 20 (2): 195–202

Mayo, S.J., Allkin, R., Baker, W., et al. (2008) Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis. **Kew Bulletin**, 63 (1): 1–16

McAndrew, P., Scanlon, E. and Clow, D. (2010) An open future for higher education. **Educause Quarterly**, 33 (1)

Millennium Ecosystem Assessment (2005) **Ecosystems and Human Well-being: Biodiversity Synthesis**. Washington DC: World Resources Institute

Morrison, D. (2011) Why is taxonomy still presented to the world as books? **Australasian Systematic Botany Society Newsletter**, 149 (December): 17–22

Morrison, D. (2012) Tools for Identifying Biodiversity: Progress and Problems. **Systematic Biology**, 61 (4): 710–712

Morse, L.E. (1971) Specimen identification and key construction with time-sharing computers. **Taxon**, pp. 269–282

Morse, L.E., Beaman, J.H. and Shetler, S.G. (1968) A computer system for editing diagnostic keys for Flora North America. **Taxon**, pp. 479–483

Natural Capital Committee (2013) **The State of Natural Capital**

NBN (2013a) **NBN data Exchange format - National Biodiversity Network** [online]. Available from: http://nbn.org.uk/Share-Data/Providing-Data/NBN-Data-Exchange-format.aspx [Accessed 4 July 2013]

NBN (2013b) **Online Recording** [online]. Available from: http://www.nbn.org.uk/Tools-Resources/Recording-Resources/Online-Recording-Toolkit.aspx [Accessed 4 July 2013]

Nikolaou, N., Sampaziotis, P., Aplikioti, M., et al. (2010) "VeSTIS: A Versatile Semi- Automatic Taxon Identification System from Digital Images." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 231–236

Nimis, P.L. and Lebbe, R.V. (2010) **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress, Paris, September 20-22, 2010, Muséum National D'histoire Naturelle - Grand Amphithéâtre**. Nimis, P.L. and Lebbe, R.V. (eds.). ETU - Edizioni Universitá

O'Neill, M.A., Gauld, I.D., Gaston, K.J., et al. (1997) "DAISY: an automated invertebrate identification system using holistic vision techniques." In Chesmore, D. and Yorke, L. (eds.). **Proceedings of inaugural meeting held held at The University of Wales Cardiff 2-3 July 1997**. **Cardiff**. **1997**. The BioNET-international group for computer-aided taxonomy (BIGCAT). pp. 13–22

Open University (2013a) **iSpot keys for identifying wildlife** [online]. Available from: http://www.ispot.org.uk/keys [Accessed 1 August 2013]

Open University (2013b) **The OpenScience Laboratory** [online]. Available from: http://www.open.ac.uk/researchprojects/open-science/ [Accessed 14 August 2013]

Pankhurst, R.J. (1970) A computer program for generating diagnostic keys. **The Computer Journal**, 13 (2): 145–151

Pankhurst, R.J. (1978) Biological identification. **Edward Arnold, London**

Pankhurst, R.J. (1988) An interactive program for the construction of identification keys. **Taxon**, pp. 747–755

Pankhurst, R.J. (1991) **Practical taxonomic computing.** Cambridge University Press

Raper, C. and Smith, M. (2013) **Tachinid Recording Scheme | ... it's a bristle thing** [online]. Available from: http://tachinidae.myspecies.info/ [Accessed 12 August 2013]

Reaka-Kudla, M.L., Wilson, D.E. and Wilson, E.O. (1996) **Biodiversity II: Understanding and Protecting Our Biological Resources (Google eBook)**. Joseph Henry Press

Record Wildlife (2013) **Record Wildlife - Wildlife App for Iphone/Android** [online]. Available from: http://www.recordwildlife.co.uk/ [Accessed 3 July 2013]

Roberts, D. and Smith, V. (2010) "ViBRANT—Virtual Biodiversity Research and Access Network for Taxonomy." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. p. 53

Roberts, M.J. (1985) **The spiders of Great Britain and Ireland. Volume 2. Linyphiidae and check list.** Harley Booksl

Rosewell, J. and Edwards, M. (2009) Bayesian keys : biological identification on mobile devices. **Open Research Online**

Rotman, D., Hansen, D., Parr, C.S., et al. (2012) Supporting Content Curation Communities : The Case of the Encyclopedia of Life. **Journal of the American Society for Information Science and Technology**, 63 (6): 1092–1107

Roy, H.E., Pocock, M.J.O., Preston, C.D., et al. (2012) **Understanding Citizen Science and Environmental Monitoring**

Sbordoni, V. (2010) "Strength and Limitations of DNA Barcode under the Multidimensional Species Perspective." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 275–280

Scottish Natural Heritage (2010) **Involving people in biological recording**

Shen, Y.-Y., Chen, X. and Murphy, R.W. (2013) Assessing DNA barcoding as a tool for species identification and data quality control. **PloS one**, 8 (2): 1–5

Silvertown, J. (2009) A new dawn for citizen science. **Trends in ecology & evolution**, 24 (9): 467–71

Smith, I. (2013) **Morddyn's photosets on Flickr** [online]. Available from: http://www.flickr.com/photos/56388191@N08/sets/ [Accessed 12 August 2013]

Smith, V.S., Rycroft, S.D., Brake, I., et al. (2011) Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. **ZooKeys**, 70 (150): 53–70

Smith, V.S., Rycroft, S.D., Harman, K.T., et al. (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. **BMC bioinformatics**, 10 Suppl 1: S6

Snaddon, J., Petrokofsky, G., Jepson, P., et al. (2012) Biodiversity technologies : tools as change agents. **Biology Letters**, 9

Van Spronsen, E., Martellos, S., Seijts, D., et al. (2010) "Modifiable digital identification keys." In Nimis, P.L. and Lebbe, R.V. (eds.). **Tools for Identifying Biodiversity: Progress and Problems. Proceedings of the International Congress**. **Paris**. **2010**. EUT Edizioni Universitá di Trieste. pp. 127–131

Stace, C.A. (1999) **Field flora of the British Isles**. Cambridge University Press

Stevenson, R.D., Haber, W.A. and Morris, R.A. (2003) Electronic Field Guides and User Communities in the Eco-informatics Revolution. **Conservation Ecology**, 7 (1): 3

TDWG (2007) **Introduction and Primer to the SDD Standard** [online]. Available from: http://wiki.tdwg.org/twiki/static/index.htm [Accessed 2 August 2013]

Tilling, S. (1984) Keys to biological identification: their role and construction. **Journal of Biological Education**, 18 (4): 293–304

Tilling, S.M. (1987) Education and taxonomy: the role of the Field Studies Council and AIDGAP. **Biological Journal of the Linnean Society**, 32 (1): 87–96

Twitter (2013) **Twitter** [online]. Available from: https://twitter.com/ [Accessed 12 August 2013]

UK National Ecosystem Assessment (2011) **UK National Ecosystem Assessment: Synthesis of the Key Findings**. Cambridge: UNEP-WCMC

Ung, V., Dubus, G., Zaragüeta-Bagils, R., et al. (2010) Xper2: introducing e-taxonomy. **Bioinformatics (Oxford, England)**, 26 (5): 703–4

Walter, D.E. and Winterton, S. (2007) Keys and the crisis in taxonomy: extinction or reinvention? **Annual review of entomology**, 52: 193–208

Wikipedia (2013) **Wikipedia, the free encyclopedia** [online]. Available from: http://en.wikipedia.org/wiki/Main_Page [Accessed 12 August 2013]

Williams, A.J., Ekins, S., Clark, A.M., et al. (2011) Mobile apps for chemistry in the world of drug discovery. **Drug discovery today**, 16 (21-22): 928–39

Wilson, E.O. (2003) The encyclopedia of life. **Trends in ecology & evolution**, 18 (2): 77–80

Wright, J.F., Morse, D.R. and Tardivel, G.M. (1995) **An investigation into the use of hypertext as a user interface to taxonomic keys**., 11 (1): 19–27

Yahoo (2013) **Yahoo! Groups - Join or create groups, clubs, forums & communities** [online]. Available from: http://uk.groups.yahoo.com/ [Accessed 9 August 2013]

YouTube (2013) **YouTube** [online]. Available from: http://www.youtube.com/ [Accessed 12 August 2013]

Zauner, H. (2009) Evolving e-taxonomy. **BMC evolutionary biology**, 9: 141

Zhang, L.Z.L., Kong, J.K.J., Zeng, X.Z.X., et al. (2008) “Plant Species Identification Based on Neural Network.” In **Fourth International Conference on Natural Computation**. **2008**. pp. 90–94