Systematics and Biodiversity **7** (1): 5–20
doi:10.1017/S147720000800282X Printed in the United Kingdom
Issued 22 February 2009

N. M. Franz[1*] & R. K. Peet[2]
[1]Department of Biology, PO Box 9012, University of Puerto Rico, Mayagüez, PR 00681, USA
[2]CB 3280, Coker Hall, University of North Carolina, Chapel Hill, NC 27599-3280, USA

submitted March 2008
accepted June 2008

Perspectives

Towards a language for mapping relationships among taxonomic concepts

Abstract Taxonomic concepts (*sensu* Berendsohn) embody the underlying meanings of scientific names as stated in a particular publication, thus offering a new way to resolve semantic ambiguities that result from multiple revisions of a taxonomic name. This paper presents a comprehensive and powerful language for representing the relationships among taxonomic concepts. The language features terms and symbols for concept relationships within a single taxonomic hierarchy, or between two related but independently published hierarchies. Taxonomic concepts pertaining to a single hierarchy are characterised by parent/child relationships, whereas those pertaining to two independent hierarchies may have the following basic relationships: congruence, inclusion (non-symmetrical, relative to the side of comparison), overlap, and exclusion. The relationships are asserted by specialists who have the option to add or subtract concepts on one or both sides of a relationship equation in order to reconcile differences between non-congruent taxonomic perspectives. The terms 'and', 'or' and 'not' are available, respectively, to connect multiple simultaneously or alternatively valid relationship assessments, or to explicitly negate the validity of a relationship. The language also permits the decomposition of a relationship according to the intensional (property referencing) and ostensive (member pointing) aspects of the compared taxonomic concepts. Adopting the concept relationship language will facilitate a more precise documentation of similarities and differences in multiple succeeding taxonomic perspectives, thereby preparing the stage for an ontology-based integration of taxonomic and related biological information.

Key words biodiversity databases, data integration, Linnaean nomenclature, ontology, taxonomic concepts, Semantic Web

Introduction

The field of biology is being transformed by an increasing demand to share and integrate information (Atkins *et al*., 2003; Jones *et al.,* 2006; Ludäscher *et al*., 2006; Michener *et al*., 2007). In order to facilitate such integration it is necessary to maintain precise and resolvable labels for a wide range of biological data. Taxonomic names play a critical role in this process, because they are standard identifiers for biological taxa and thus for any observations linked to these taxa. However, taxonomic names are not capable of supporting all existing or envisioned needs for information resolution. To provide an example, large-scale phylogenetic or ecological analyses must often process name-referenced information stored in various repositories, including traditional publications, desktop or online databases, and information stored in museum collections. The names used in these sources reflect a significant degree of regional, temporal and individual taxonomic bias. Yet only some of the differences in taxonomic perspective are deducible from the names themselves. This is so because the definition and validity of a Linnaean name depends primarily upon its relationship to individual type specimens instead of a full-blown taxonomic view (see Farber, 1976; Stevens, 1984). The same name may acquire different meanings through time as it is redefined in successive taxonomic treatments. On the other hand, several different names (i.e. nomenclatural synonyms) can have the same underlying meaning. Therefore, as time progresses, Linnaean names and taxonomic perspectives tend to accumulate many-to-many relationships that are difficult to trace (Beach *et al*., 1993; Berendsohn, 1995; Koperski *et al*., 2000; Garrity & Lyons, 2003; Geoffroy & Berendsohn, 2003; Berendsohn & Geoffroy, 2007). Perpetuating name/meaning divergence is a major impediment to the integration of biological information.

The taxonomic concept approach has emerged as a solution for overcoming the semantic ambiguities that result from multiple revisions of a taxonomic name (e.g. Berendsohn, 1995; Kennedy *et al*., 2005; Franz *et al*., 2008). A *taxonomic*

*Corresponding author. Email: franz@uprm.edu

concept is the underlying meaning, or referential extension, of a scientific name as stated by a particular author in a particular publication. It represents the author's full-blown view of how the name reaches out to observed or unobserved objects in nature (beyond statements about type specimens). It is a direct reflection of what has been written, illustrated, and deposited by a taxonomist, regardless of his or her theoretical orientation.[1] Taxonomic concepts are labelled using the abbreviation 'sec.' for the Latin *secundum*, or 'according to' (Berendsohn, 1995). The 'sec.' is preceded by the full Linnaean name and followed by the specific author and publication, as in *Andropogon virginicus* L. sec. Radford *et al.* (1968), an earlier concept, versus *Andropogon virginicus* L. sec. Weakley (2006), which is a later and narrower concept. The consistent practice of handling a taxonomic name only in connection with a specific source makes it possible to trace the evolution of its multiple meanings through time.

This paper introduces a set of terms and symbols used specifically for mapping relationships among taxonomic concepts. Elaborating on efforts by previous authors (Koperski & Sauer, 1999; Gradstein *et al.,* 2001; Geoffroy & Berendsohn, 2003), the new language is more powerful than conventional nomenclatural relationships such as synonymy. The herein presented terms and examples should concern anyone interested in managing and integrating name-referenced biological data, particularly taxonomic experts and biodiversity researchers. Because concept relationships permit a more accurate representation of taxonomic perspectives in a computerised working environment, they represent an important step in preparing the field of taxonomy for the Semantic Web (see Berners-Lee *et al*., 2001; Atkins *et al*., 2003; Gangemi & Mika, 2003; Page, 2005, 2006).

Vision for information resolution using taxonomic concepts

How should one envisage the process of semantic resolution via taxonomic concepts in a global information network? The future management of biodiversity data will be driven by an extensive metadata approach (Michener & Brunt, 2000; Page, 2006; Pyle *et al*., 2008). Reliable, long-term data integration will depend upon the availability of precisely articulated concepts and their concept relationships on one side, and the linking of biodiversity information to such concepts on the other side. An increasing pool of taxonomic concepts will be supplied through *publication* in many of the same outlets presently providing information on taxonomic names (cf. Scoble, 2004), e.g. peer-reviewed publications, authoritative taxonomic services, on-line revisions (cf. Godfray, 2002), and other networked archives for taxonomic information. To facilitate recognition and transport across platforms, taxonomic concepts will receive globally unique digital identifiers (e.g. Paskin, 2005; Page, 2006, 2008). Software portals to a growing taxonomic concept archive will permit experts to visualise multiple related taxonomies and map relationships among their respective concepts.

Researchers who provide biodiversity data (e.g. DNA sequences, specimen distributions, life history information) to the network will have options to label all submitted information as accurately as possible. Ideally this includes labelling primary observations not only with a taxonomic name ('*A. virginicus* L.') but also creating a link to one or more established taxonomic concepts ('*A. virginicus* L. sec. Weakley, 2006'). The process of linking an observation to a published concept is called *identification.* Users should have some flexibility to adjust the degree of taxonomic resolution to the needs and limitations of a particular dataset, e.g. by selecting multiple taxonomic concepts if a complex of closely related species could not be differentiated more finely.

If primary observations are identified to taxonomic concepts, and these concepts are mapped to each other via concept relationships, then the long-term integration of biological data can be achieved using the concept relationship terms and symbols. In other words, the primary observations are resolvable as long as the originally identified concepts remain relatable to those published in succeeding taxonomic perspectives.

A review of the necessary steps towards the above vision is beyond the scope of this paper. However, we point out that several key elements are already implemented at a smaller scale, and others are being developed. For instance, metadata standards for on-line data submission are in use in various biological disciplines (e.g. Michener *et al*., 1997; Kennedy *et al.*, 2006; Leebens-Mack *et al.*, 2006). Several major providers of taxonomic information have adopted the concept approach (see Berendsohn *et al.*, 2003; Garrity & Lyons, 2003; Franz *et al.*, 2008). A protocol for transitioning from a name-based system to taxonomic concepts is also available (Kennedy *et al*., 2005). Several applications have been developed to visualise multiple taxonomic perspectives and map relationships among their constituent concepts (e.g. Graham *et al.*, 2002; Parr *et al.*, 2004; Craig & Kennedy, 2008). Nevertheless, the existing products tend to have a still limited user community, needing integration into widely used applications and information storage environments (see Jones *et al.*, 2006).

Understanding the context of concept mapping

The use of taxonomic concepts implies that Linnaean names are handled exclusively in connection with particular sources in which their meanings are explained. Individuals and institutions using taxonomic concepts must in each instance understand their role as communicators in order to minimise the proliferation of vague and potentially redundant taxonomic concepts ('concept inflation' *sensu* Berendsohn, 1995). The *authoring* of taxonomic concepts – i.e. the creation of new name/source connections – is a task that should be reserved for experts publishing significant new results or summaries on the systematics of a particular lineage. Ideally these products will be accompanied by new empirical information, voucher specimens or genetic sequences, textual circumscriptions, illustrations, and additional clarifying evidence. In all other contexts users should resort to *citing* taxonomic concepts

Category	Explanation
1. Original concept	Appears in the earliest reference where a new name/type association was established, or a new name was created (for higher-level taxa). Example: *Andropogon virginicus* L. sec. Linnaeus (1753).
2. Revisional concept	Appears in a comprehensive revision of an existing taxonomic name and lineage, e.g. a monograph, a species webpage, or a field manual with descriptions and illustrations. Example: *Andropogon virginicus* L. sec. Campbell (1983).
3. Relational concept	Appears in a comprehensive hierarchy of concepts (typically without descriptions and illustrations); therefore the meaning may be inferred through parent/child relationships, or by examining the entire list of mutually exclusive concepts. Example: *Andropogon virginicus* L. sec. ITIS (2006). [an on-line taxonomic service]
4. Informal concept	Appears in a taxonomic treatment yet is poorly specified (mentioned 'in passing'), or merely implied, being outside of the discussion of the focal names and taxa. Example: *Andropogon hallii* Hackel sec. Weakley (2006). [a Midwestern species]
5. Nominal concept	Appears outside of the context of a particular source (without a sec.); indirectly links to all other concepts in which the name occurs. Example: *Andropogon virginicus* L.

Table 1 Categories of taxonomic concepts.

that were published elsewhere. A useful categorisation of taxonomic concepts is presented in Table 1.

Concept relationships are established among two or more unique taxonomic concepts. They are called *parent/child* or *vertical* relationships if the related concepts pertain to the same published source. Parent/child relationships are readily deduced from the hierarchical structure of a single classificatory system or phylogeny. On the other hand, *sibling* or *horizontal* relationships exist between concepts in independently published systems. These relationships may be published along with a series of new taxonomic concepts (see Weakley, 2006), or separately, and thus refer to concepts published in two preceding sources ('third party relationships'). The mapping of horizontal concept relationships is a non-trivial task that requires an intimate understanding of the included systematic perspectives. It is possible that different assessors propose different relationships among the same set of taxonomic concepts. Therefore, concept relationships have specific and explicit authors and publication times, just like taxonomic concepts (see Kennedy *et al*., 2005).

Parent/child concept relationships[2]

The mapping of parent/child concept relationships is possible for any taxonomic system with a consistent hierarchical structure, including (e.g.) traditional systematic catalogues or highly resolved phylogenies. Parent/child relationships can apply to concepts with Linnaean ranks as well as informally named lineages (Fig. 1; Table 2). The terms 'is a parent of' and 'is a child of' are reserved exclusively for a single-hierarchy context; they are not semantically equivalent to the terms 'includes' and 'is included in' detailed below. For the purpose of assembling a hierarchical structure, one should consider every child concept as an integral part of its most immediate parent concept, but not vice-versa. Possible exceptions to this rule are taxonomic concepts for hybrid taxa with multiple parent concepts. The nestedness of the structure implies that all same-level concepts are exclusive of each other – a property which their children inherit.

Horizontal concept relationships

The mapping of horizontal concept relationships involves two or more related yet independently published taxonomies. A relationship assessment should reconcile all relevant taxonomic information connected with each concept, including the types and additional listed specimens, subsumed child concepts, character attributes, and the larger systematic context. Five basic terms and symbols, derived from set theory, are used for this purpose (Fig. 2; Table 3; see also Thau & Ludäscher, 2008). These are most effective when the meaning of each is considered exclusive of the other (Koperski *et al*., 2000). 'Overlaps with' means that each concept has at least one non-shared element in addition to their co-extensional parts, and thus differs from both 'includes' and the inverse 'is included in'. The term 'excludes' means that not even the concepts' respective type specimens yield a match. This is a rare situation within the Linnaean system and indicative of an error, such as homonyms based on unrelated taxa. Whenever possible an explanation should accompany the assessment, particularly in the case on non-congruence.

Koperski *et al.* (2000) pioneered the use of horizontal relationships for concepts of German mosses (Fig. 3; Table 4). At the generic level, these authors mainly used the number of constituent Central European species as a criterion for comparison. Accordingly, the concept *Amblystegium* Schimp. sec. Koperski *et al.* (2000) includes eight species, whereas *Amblystegium* Schimp. sec. Frahm & Frey (1992) includes three species, with the remaining species placed into three additional genera. In other words, the (2000) concept of *Amblystegium* Schimp has a wider circumscription than its (1992) predecessor, which is reflected in the 'includes' (>) relationship. In another comparison, *Amblystegium* Schimp. sec. Koperski *et al.* (2000) shares

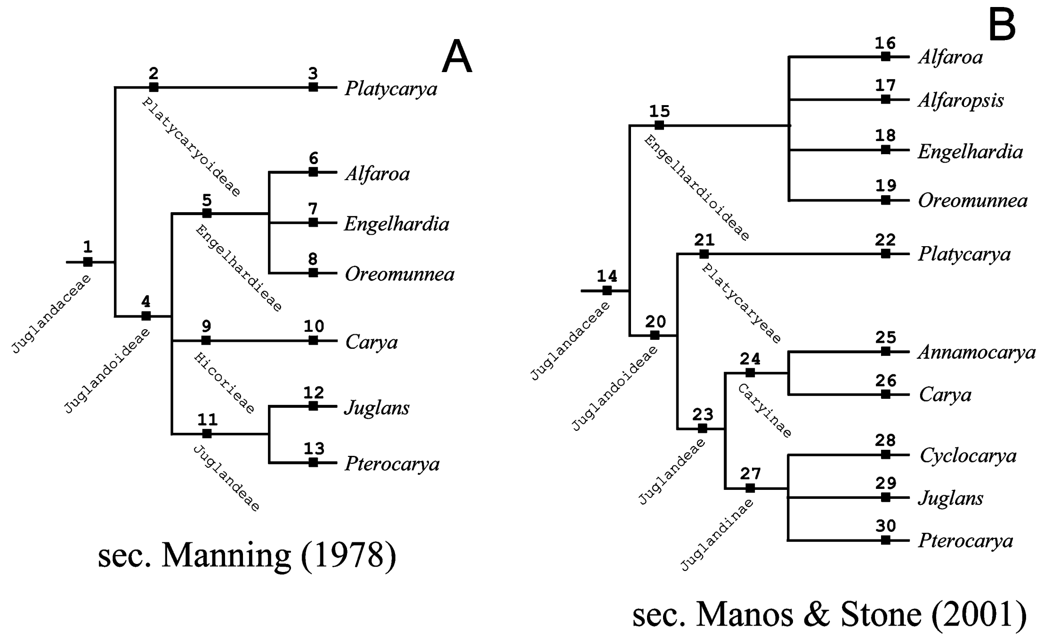

Figure 1 Higher-level phylogeny of walnuts according to (A) Manning (1978) and (B) Manos and Stone (2001). The numbers for the respective concepts are the same as in Tables 2 and 6.

1st Name	1st Source	Relationship	2nd Name	2nd Source
Juglandaceae (1)	sec. Manning (1978)	is a parent of	Platycaryoideae (2)	sec. Manning (1978)
Juglandaceae (1)	sec. Manning (1978)	is a parent of	Juglandoideae (4)	sec. Manning (1978)
Platycaryoideae (2)	sec. Manning (1978)	is a parent of	*Platycarya* (3)	sec. Manning (1978)
Juglandoideae (4)	sec. Manning (1978)	is a parent of	Engelhardieae (5)	sec. Manning (1978)
Juglandoideae (4)	sec. Manning (1978)	is a parent of	Hicorieae (9)	sec. Manning (1978)
Juglandoideae (4)	sec. Manning (1978)	is a parent of	Juglandeae (11)	sec. Manning (1978)
Engelhardieae (5)	sec. Manning (1978)	is a parent of	*Alfaroa* (6)	sec. Manning (1978)
Engelhardieae (5)	sec. Manning (1978)	is a parent of	*Engelhardia* (7)	sec. Manning (1978)
Engelhardieae (5)	sec. Manning (1978)	is a parent of	*Oreomunnea* (8)	sec. Manning (1978)
Hicorieae (9)	sec. Manning (1978)	is a parent of	*Carya* (10)	sec. Manning (1978)
...	...	...	...	...
Juglandaceae (14)	sec. Manos & Stone (2001)	is a parent of	Engelhardioideae (15)	sec. Manos & Stone (2001)
Juglandaceae (14)	sec. Manos & Stone (2001)	is a parent of	Juglandoideae (20)	sec. Manos & Stone (2001)
Engelhardioideae (15)	sec. Manos & Stone (2001)	is a parent of	*Alfaroa* (16)	sec. Manos & Stone (2001)
Engelhardioideae (15)	sec. Manos & Stone (2001)	is a parent of	*Alfaropsis* (17)	sec. Manos & Stone (2001)
Engelhardioideae (15)	sec. Manos & Stone (2001)	is a parent of	*Engelhardia* (18)	sec. Manos & Stone (2001)
Engelhardioideae (15)	sec. Manos & Stone (2001)	is a parent of	*Oreomunnea* (19)	sec. Manos & Stone (2001)
Juglandoideae (20)	sec. Manos & Stone (2001)	is a parent of	Platycaryeae (21)	sec. Manos & Stone (2001)
Juglandoideae (20)	sec. Manos & Stone (2001)	is a parent of	Juglandeae (23)	sec. Manos & Stone (2001)
Platycaryeae (21)	sec. Manos & Stone (2001)	is a parent of	*Platycarya* (22)	sec. Manos & Stone (2001)
Juglandeae (23)	sec. Manos & Stone (2001)	is a parent of	Caryinae (24)	sec. Manos & Stone (2001)
...	...	...	...	...

Table 2 Examples of parent/child relationships for concepts representing the higher-level phylogeny of walnuts according to Manning (1978) versus Manos and Stone (2001; see Fig. 1).

one species-level concept with *Campylium* (Sull. Mitt.) sec. Frahm & Frey (1992). Consideration of the other non-shared elements results in the relationship 'overlaps with' (><). Only the two concepts for *Leptodictyum* (Schimp.) Warnst. are taxonomically congruent in this example, providing semantically accurate resolution of information via the Linnaean name alone.

Reconciling taxonomic perspectives through addition or subtraction of concepts

Comparisons of concepts stemming from multiple taxonomies are likely to produce numerous non-congruent relationships

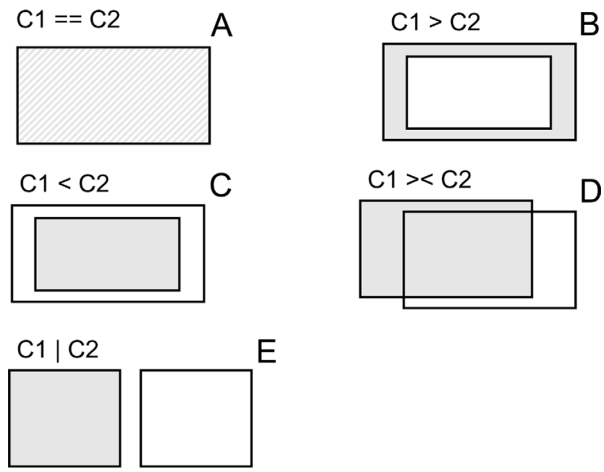

Figure 2 Schematic representation (Venn diagrams) of the five basic kinds of horizontal concept relationships. The referential extension of concept C1 is indicated by the shaded rectangle, whereas that of concept C2 is indicated by the white rectangle. (A) C1 is congruent with C2 (complete overlap); (B) C1 includes C2; (C) C1 is included in C2; (D) C1 overlaps with C2; and (E) C1 excludes C2.

(Franz *et al*., 2008). When 'lumping' or 'splitting' within a shared higher-level taxonomic entity is involved, the differences are frequently reconcilable through addition (+) or subtraction (−) of lower-level concepts on one or both sides of the relationship equation (Table 3). Application of these operators resolves a large number of taxonomic incongruencies that are neither captured in the names nor in the existing synonymy relationships (see also Koperski *et al*., 2000).

The mapping of concepts for Eastern United States species of *Andropogon* L. published in four different treatments reveals largely incongruent nomenclatural and taxonomic views (Fig. 4). Nevertheless it is possible to identify congruence at certain levels (Table 5). For example, by adding together two species concepts sec. Weakley (2006) – i.e. *Andropogon glomeratus* (with two subspecies concepts) and *Andropogon tenuispatheus* – one obtains a taxonomic entity that is congruent with *Andropogon virginicus* var. *abbreviatus* sec. Godfrey & Wooten (1979). Such assessments of 'hidden congruence' convey an important message, i.e. that two taxonomic treatments are partially compatible in spite of obvious differences in nomenclature and taxonomic subdivisions.

Combining multiple relationships, expressing uncertainty and negating relationships

The term 'AND' is used to connect multiple *simultaneously* valid relationship assessments among two sets of concepts. Combinations of horizontal concept relationships with existing nomenclatural relationships are primary cases for applying the term. For example, *Andropogon virginicus* var. *abbreviatus* sec. Godfrey & Wooten (1979) includes (>) *Andropogon glomeratus* var. *glomeratus* sec. Weakley (2006), AND the former is a heterotypic synonym of the latter. Combined statements of nomenclatural terms and concept relationships are more informative than using either system in isolation. A listing of available nomenclatural adjectives and relationships is provided in Appendix 1.

Relationship	Symbol	Explanation
is a parent of[1]		One concept subsumes another concept within the same hierarchy.
is a child of[1]		One concept is subsumed under another concept within the same hierarchy.
is congruent with	==	Two concepts have a congruent taxonomic meaning or referential extension.
Includes	>	The meaning of one concept is more inclusive than that of another concept.
is included in	<	The meaning of one concept is less inclusive than that of another concept.
overlaps with	><	Two concepts are partially congruent yet in addition each concept includes a taxonomic entity that is excluded from the other concept.
Excludes	|	Two concepts have non-overlapping taxonomic meanings.
Plus	+	Used to add one concept to another concept.
Minus	–	Used to subtract one concept from another concept.
And	**AND**	Used to aggregate multiple simultaneously valid concept relationships into a whole.
Or	**OR**	Used to aggregate multiple alternatively valid concept relationships into a whole, with the intent of expressing ambiguity in the assessment.
Not	**NOT**	Used to negate a particular concept relationship.
Intensional	**(INT)**	Used to indicate the intensional character of a concept relationship which was assessed focusing only on the described attributes of the related concepts.
Ostensive	**(OST)**	Used to indicate the ostensive character of a concept relationship which was assessed focusing only on the demonstrated members of the related concepts.

Table 3 Summary of terms and symbols used for mapping concept relationships.

[1]Applies to vertical concept relationships. All other terms and symbols apply to horizontal concept relationships.

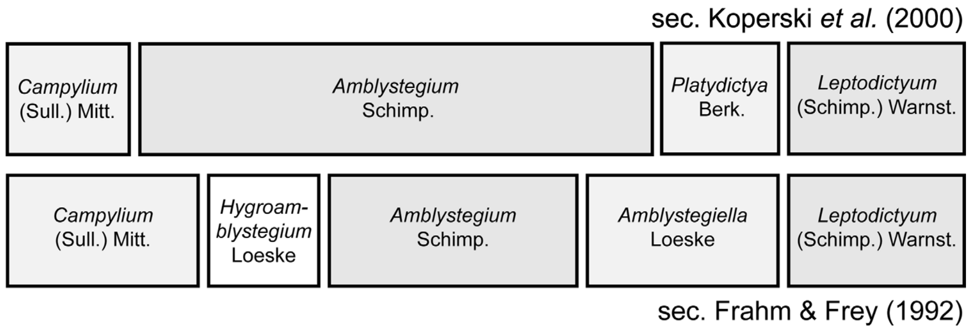

Figure 3 Representation of horizontal relationships among two sets of concepts for select German moss genera according to Frahm and Frey (1992) and Koperski *et al.* (2000; adopted from the latter authors). The width and position of each concept correspond to the identity and number of lower-level elements. See also Table 4.

The term 'OR', on the other hand, is used to connect multiple *alternatively* valid relationships among two sets of concepts. This becomes necessary whenever one or more of the compared concepts are poorly specified and therefore do not permit an unambiguous relationship assessment. Koperski *et al.* (2000) labelled such cases of uncertainty with a question mark ('?') juxtaposed to the most likely symbol. However, the herein proposed solution is both more precise and more flexible. For example, using the term 'OR' allows experts to signal that one concept relationship is congruent with or more inclusive than another concept (== OR >; analogous to the mathematical term '≥'), or the reverse (== OR <; analogous to '≤ '); that two concepts overlap in an unspecified way (== OR > OR < OR ><; analogous to '∼'); or that they are not congruent (> OR < OR >< OR |; analogous to '≠'). Each of these combinations conveys a sense of ambiguity, yet without rendering the relationship assessment meaningless. They clearly restrict the range of valid relationship terms.[3]

Lastly, the term 'NOT' is available to explicitly *negate* a horizontal concept relationship, as in: *Andropogon virginicus* L. sec. Weakley (2006) is not congruent with (NOT ==) *Andropogon virginicus* L. sec. Radford *et al.* (1968).

Distinguishing intensional from ostensive components of concept definitions

Taxonomic concepts often have two components that jointly specify their meanings, i.e. an intensional and an ostensive component. In the philosophy of language (e.g. Devitt & Sterelny, 1999), an *intensional definition* is one that represents the meaning of a term by listing all the *properties* ('necessary and sufficient conditions') required for something to fall under the definition. In systematics, intensional definitions may list putative synapomorphies or other diagnostic features of a lineage, e.g. the triploid endosperm of angiosperms or the abdominal spinnerets of spiders. An *ostensive definition,* in turn, is one that represents the meaning of a term by listing one or more (though not necessarily all) *members* that are part of the set and serve as pointed-out examples. The selection of a type specimen or type species is part of such an ostensive definition. Koperski *et al.* (2000) used ostensive definitions to compare concept for German mosses, adding up the constituent species-level concepts to establish the width of a genus-level concept.

Intensional and ostensive components serve complementary roles in specifying the referential extension of a taxonomic concept. In practice their respective meanings and functions are rather different. Intensional definitions have an important predictive aspect, referring to objects at any subordinate taxonomic scale that remain unmentioned or unobserved at the time of publication. In order to work they require a close matching of intersubjective interpretations of character information; otherwise they can lead to misunderstandings. Ostensive definitions tend to be more precise and reproducible, but may be less conducive to understanding a taxon's limits – beyond the set of demonstrated objects. As taxonomic concepts progress from lower to higher levels, systematists tend to emphasise intensional over ostensive definitions, the latter component being represented by a relatively small number of analysed exemplars.

Horizontal concept relationships are greatly enriched by separating the intensional and ostensive components of each related concept. Operationally, this is achieved by inserting the terms '(INT)' and '(OST)' after each separate relationship assessment and connecting the two statements with an 'AND' (see above). For example, *Bryum* Hedw. sec. Koperski *et al.* (2000) is intensionally congruent with (== INT) *Bryum* Hedw. sec. Frahm & Frey (1992) AND the former ostensively includes (> OST) the latter, since Koperski *et al.* (2000) include six species-level concepts that Frahm and Frey (1992) were unaware of in their earlier publication. The combined intensional/ostensive assessment implies that the two author teams agree on what should 'belong' into *Bryum* Hedw., even though their respective lists of species-level members are non-congruent. Had Frahm and Frey (1992) examined the six additional species, they would have recognised them as pertaining to their concept of *Bryum* Hedw., as is reflected in the '== (INT)' relationship assessment. This sort of information is extremely valuable given that many systematic treatments are restricted to a particular region or use an exemplar approach to infer taxonomic and phylogenetic relationships. Especially at higher levels, the only type of match one might expect to occur regularly among concepts is intensional congruence. The separation of intensional and ostensive concept components can thus capture a critical element of agreement among

1st Name	1st Source	Relationship	2nd Name	2nd Source
Campylium (Sull.) Mitt.	sec. Koperski *et al.* (2000)	is included in (<)	*Campylium* (Sull.) Mitt.	sec. Frahm & Frey (1992)
Amblystegium Schimp.	sec. Koperski *et al.* (2000)	overlaps with (><)	*Campylium* (Sull.) Mitt.	sec. Frahm & Frey (1992)
Amblystegium Schimp.	sec. Koperski *et al.* (2000)	includes (>)	*Hygroamblystegium* Loeske	sec. Frahm & Frey (1992)
Amblystegium Schimp.	sec. Koperski *et al.* (2000)	includes (>)	*Amblystegium* Schimp.	sec. Frahm & Frey (1992)
Amblystegium Schimp.	sec. Koperski *et al.* (2000)	overlaps with (><)	*Amblystegiella* Loeske	sec. Frahm & Frey (1992)
Platydictya Berk.	sec. Koperski *et al.* (2000)	is included in (<)	*Amblystegiella* Loeske	sec. Frahm & Frey (1992)
Platydictya Berk.	sec. Koperski *et al.* (2000)	excludes (|)	*Amblystegium* Schimp.	sec. Frahm & Frey (1992)
Leptodictyum (Schimp.) Warnst.	sec. Koperski *et al.* (2000)	is congruent with (==)	*Leptodictyum* (Schimp.) Warnst.	sec. Frahm & Frey (1992)
. . .	. . .	. . .	. . .	. . .

Table 4 Examples of horizontal relationships for concepts representing moss genera according to Koperski *et al.* (2000) versus Frahm and Frey (1992; see Fig. 3).

systematists in spite of their pointing at incongruent sets of taxonomic entities.

A series of abstract examples illustrates the utility of combined intensional/ostensive relationships (Fig. 5). At Time 1, Author A publishes a taxonomic concept C5, defined by a perceived property p(C5) and also pointing at two included member concepts C1 and C2 (Fig. 5A). However, the property diagnosis is formulated broadly so as to include two additional entities C3 and C4 which Author A does not mention. At another Time 2, Author B publishes a revisional concept C5′ with a more restrictive intensional diagnosis p(C5′) that refers only to the corresponding revisional concepts C1′ and C2′, at the exclusion of C3′ and C4′ (Fig. 5B). From the perspective of Author B, this definitional refinement is mirrored in the combined relationship C5′ sec. Author B (Time 2) < (INT) AND == (OST) C5 sec. Author A (Time 1).

In another, more complicated scenario (Fig. 5C), Author C publishes a concept C5′ which (for the purpose of realism) precedes the work of Author A. This concept also points at the included member concepts C1′ and C2′. However, in the assessment of another systematist (such as Author A), Author C's property p(C5′) actually refers to concepts C2′ and C6 – the latter unexamined by Author C – whereas concept C1′ is erroneously subsumed under property p(C5′). Therefore, in Author A's assessment, C5 sec. Author A (Time 1) >< (INT) AND == (OST) C5′ sec. Author C (Time 3).

In a third case, Author D publishes a concept C5′ at Time 4 (Fig. 5D). This concept's intensional component p(C5′) is congruent with Author A's property p(C5),[4] but points at one entity (C3′) not mentioned by Author A while failing to mention C1′. This is reflected in the relationship C5′ sec. Author D (Time 4) == (INT) AND >< (OST) C5 sec. Author A (Time 1).

All three examples share the characteristic of having only one of the two components labelled as congruent. Such cases of partial congruence are especially informative because they represent both elements of continuity and change in succeeding taxonomic revisions.

A worked example

The higher-level taxonomy of the walnut family has been subject to numerous revisions over the past 50 years, exemplified by the alternative classification schemes of Manning (1978; Fig. 1A) and Manos and Stone (2001; Fig. 1B). The similarities and differences in these two perspectives are assessed here in a near-comprehensive set of horizontal concept relationships (Table 6). The goal of such a mapping is to connect every concept published in one system with at least one concept, typically the closest match, in the other system. In practice this can be achieved by starting at the root of the more recent classification and working gradually towards the tips, establishing horizontal relationships at each step. In a second pass, one should ensure that all concepts in the less recent classification are accounted for, adding relationships whenever necessary. Additions to the first pass are needed particularly if the earlier classification reaches to a lower taxonomic level than its successor. In a third pass, one may further reconcile taxonomic

sec. Hitchcock & C. (1950)	sec. RAD (1968)	sec. Godfrey & W. (1979)	sec. Weakley (2006)
A. capillipes	*A. virginicus*	*A. capillipes*	*A. capillipes* var. *capillipes*
A. capillipes	*A. virginicus*	*A. capillipes*	*A. capillipes* var. *dealbatus*
A. virginicus var. *virginicus*	*A. virginicus*	*A. virginicus* var. *virginicus*	*A. virginicus* var. *virginicus*
A. virginicus var. *virginicus*	*A. virginicus*	*A. virginicus* var. *virginicus*	*A. virginicus* var. *virginicus*
A. virginicus var. *virginicus*	*A. virginicus*	*A. virginicus* var. *virginicus*	*A. virginicus* var. *decipiens*
A. virginicus var. *glaucopsis*	*A. virginicus*	*A. glaucopsis*	*A. glaucopsis*
A. virginicus var. *hirsutior*	*A. virginicus*	*A. virginicus* var. *abbreviatus*	*A. glomeratus* var. *hirsutior*
A. glomeratus	*A. virginicus*	*A. virginicus* var. *abbreviatus*	*A. glomeratus* var. *glomeratus*
A. glomeratus	*A. virginicus*	*A. virginicus* var. *abbreviatus*	*A. tenuispatheus*

Figure 4 Representation of horizontal relationships among species- and subspecies-level concepts for the grass genus *Andropogon* L. according to Hitchcock and Chase (1950), Radford *et al.* (1968), Godfrey and Wooten (1979), and Weakley (2006). Each column contains a coherent taxonomic perspective and each row represents a congruent taxonomic concept, irrespective of the names used to label the individual cells. The various shadings indicate taxonomic concepts whose circumscriptions are shared among multiple authors. See also Table 5.

differences by using concept addition and subtraction and by comparing intensional and ostensive concept components. A concept-by-concept matrix can assist in checking for omissions. Once this process is completed, a computer algorithm (see Thau & Ludäscher, 2008) may infer many additional and implicit relationships based on the hierarchical structure of both taxonomic perspectives and their horizontal relationships. Many of these implicit relationships will be of the type 'exclusion' between members of different sections in the respective hierarchies.

Manning's (1978) and Manos and Stone's (2001) higher-level classifications of walnuts display some form of incongruence in 12 of 20 assessed concept relationships (Table 6). For the purpose of demonstration, the ostensive/intensional distinction was used only for comparisons above the lowest taxonomic level (i.e. above the generic level). Manning (1978) provides a comprehensive list of relational concepts (see Table 1) for nearly 60 walnut species which are not explicitly matched here.

Starting at the root of each scheme, the two corresponding concepts for Juglandaceae (i.e. concepts 14 and 1; Fig. 1) are congruent in the sense that the authors are in overall agreement about the family's diagnostic features (== INT; see Table 6). The ostensive component of this relationship illustrates a phenomenon commonly associated with higher-level definitions based on pointing at perceived taxa. For one, it may be difficult to decide how far down the hierarchy one should look in order to find matching members. Both Manning (1978) and Manos and Stone (2001) include a subfamily-level concept 'Juglandoideae' as an immediate child of the root concept (as in: >< OST); however, at the next lower level these two entities have incongruent elements (as in: | OST). Such ambiguity may be expressed using the 'OR' term (Table 6).

The semantic utility and complementarity of ostensive relationships is usually maximised if one focuses narrowly on membership differences at the closest subordinate level. This strategy is suited to expose even minor incongruences in structural arrangement – at the level where they occur. For example, the Caryinae sec. Manos and Stone (2001) are intensionally congruent with the Hicorieae sec. Manning (1978), although ostensively the former point at two genus-level concepts instead of one (Fig. 1). Manning (1978) treated *Annamocarya* sec. Manos & Stone (2001) as a section (named *Rhamphocarya*) of his concept *Carya.* This and two additional cases of rank shifts are reconciled by adding up Manos and Stone's (2001) narrower genus-level concepts (Table 6). Overall, the relationship assessments uncover elements of partial congruence, and thus facilitate semantic integration, between all but two concept pairs (Juglandoideae and Juglandeae).

Strengths and limitations

The proposed language for mapping concept relationships improves upon previous efforts (Koperski *et al.*, 2000) both in terms of semantic flexibility and precision. The use of parent/child relationships allows representation of an infinite number of nested and reticulate taxonomic hierarchies within a single taxonomic service environment. Expert-authored horizontal relationships may include traditional as well as modern phylogenetic concepts ranging from the level of variety to the highest taxonomic ranks. Areas of uncertainty are made explicit and at the same time the realm of likely relationships is limited using the term 'OR'. The application of concept addition and subtraction and the specification of intensional and ostensive

1st Name	1st Source	Relationship	2nd Name	2nd Source
Andropogon virginicus var. *virginicus* + *Andropogon virginicus* var. *decipiens*	sec. Weakley (2006)	is congruent with (==)	*Andropogon virginicus* var. *virginicus*	sec. Godfrey & Wooten (1979)
Andropogon glomeratus var. *hirsutior* + *Andropogon glomeratus* var. *glomeratus* + *Andropogon tenuispatheus*	sec. Weakley (2006)	is congruent with (==)	*Andropogon virginicus* var. *abbreviatus*	sec. Godfrey & Wooten (1979)
A. virginicus	sec. Weakley (2006)	is congruent with (==)	*Andropogon virginicus* – *Andropogon virginicus* var. *abbreviatus*	sec. Godfrey & Wooten (1979)
A. virginicus	sec. Weakley (2006)	is congruent with (==)	*Andropogon virginicus* – *Andropogon virginicus* var. *glaucopsis* – *Andropogon virginicus* var. *hirsutior*	sec. Hitchcock & Chase (1950)
Andropogon glomeratus – *Andropogon glomeratus* var. *hirsutior* + *Andropogon tenuispatheus*	sec. Weakley (2006)	is congruent with (==)	*Andropogon glomeratus*	sec. Hitchcock & Chase (1950)
Andropogon virginicus var. *abbreviatus*	sec. Godfrey & Wooten (1979)	is congruent with (==)	*Andropogon virginicus* var. *hirsutior* + *Andropogon glomeratus*	sec. Hitchcock & Chase (1950)
...	...	...	...	...

Table 5 Examples of using the addition and subtraction operators to reconcile concepts representing *Andropogon* species according to four distinct taxonomies (see Fig. 4).

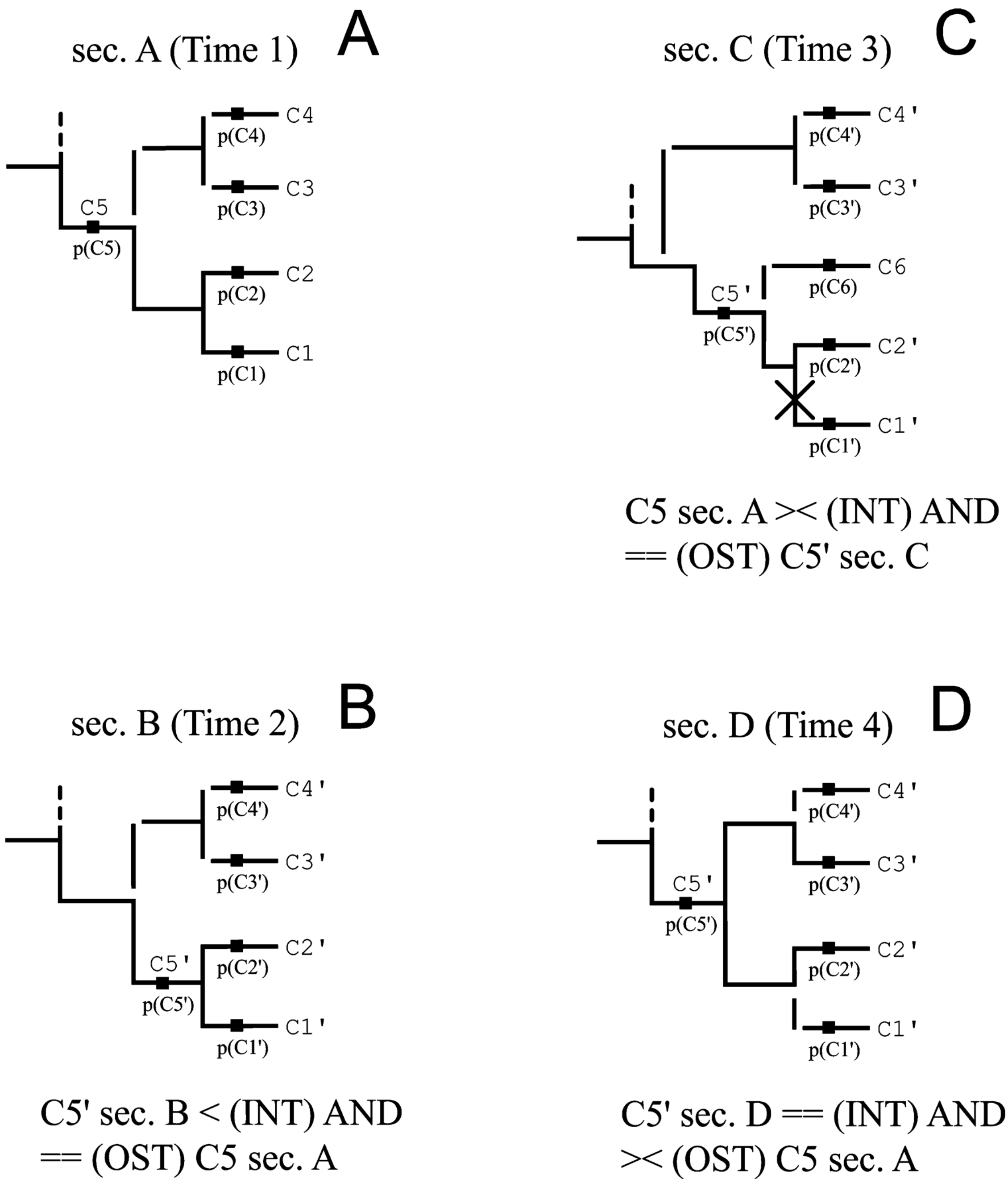

Figure 5 Three examples of partial congruence among concepts represented in hypothetical revisions of a taxonomic entity, with (A) as the reference classification in each comparison. The concepts are represented as parts of phylogenetic trees, with lower-level elements and properties. Contiguous lines indicate that the connected member concepts (e.g. C1) are explicitly included in the circumscription of a higher-level concept (e.g. C5). Interrupted lines, in turn, indicate that a lower-level concept (e.g. C3) is not mentioned explicitly, and subsumed under a higher-level (e.g. C5) concept only by virtue of that concept's intensional component (e.g. p[C5]). The 'X' in (C) indicates that concept C1′ is erroneously subsumed under concept C5′ and property p(C5′). See text for further details.

concept components have an important explanatory role. They provide users with a sense of referential continuity and change in taxonomic perspective that cannot be expressed with names and synonymy relationships alone (see Appendix 1). The added precision allows experts to use concept relationships in order to *measure* the semantic reliability of taxonomic names over time (Geoffroy & Berendsohn, 2003; Weakley, 2006; Franz *et al.,* 2008). Adopting this language does not require subscription to a particular school of systematic inference or classification, and furthermore does not undermine the validity and utility of the Linnaean system.

On the other hand, terms and symbols are unable to wholly replace visual or textual representations of taxonomic concepts and their relationships. The five basic symbols are especially limited in the context of answering such questions as: 'how different are two non-congruent concepts?', 'how reliable is the assessment?' and 'is the difference likely to be significant?' Answers to these questions are critical for users who

1st Name	1st Source	Relationship	2nd Name	2nd Source
Juglandaceae (14)	sec. Manos & Stone (2001)	== (INT) AND >< (OST) OR | (OST)	Juglandaceae (1)	sec. Manning (1978)
Engelhardioideae (15)	sec. Manos & Stone (2001)	== (INT) AND > (OST)	Engelhardieae (5)	sec. Manning (1978)
Alfaroa (16)	sec. Manos & Stone (2001)	==	*Alfaroa* (6)	sec. Manning (1978)
Alfaropsis (17)	sec. Manos & Stone (2001)	<	*Engelhardia* (7)	sec. Manning (1978)
Engelhardia (18)	sec. Manos & Stone (2001)	<	*Engelhardia* (7)	sec. Manning (1978)
Alfaropsis (17) + *Engelhardia* (18)	sec. Manos & Stone (2001)	==	*Engelhardia* (7)	sec. Manning (1978)
Oreomunnea (19)	sec. Manos & Stone (2001)	==	*Oreomunnea* (8)	sec. Manning (1978)
Juglandoideae (20)	sec. Manos & Stone (2001)	>< (INT) AND >< (OST)	Juglandoideae (4)	sec. Manning (1978)
Platycaryeae (21)	sec. Manos & Stone (2001)	== (INT) AND == (OST)	Platycaryoideae (2)	sec. Manning (1978)
Platycarya (22)	sec. Manos & Stone (2001)	==	*Platycarya* (3)	sec. Manning (1978)
Juglandeae (23)	sec. Manos & Stone (2001)	> (INT) AND >< (OST) OR | (OST)	Juglandeae (11)	sec. Manning (1978)
Caryineae (24)	sec. Manos & Stone (2001)	== (INT) AND == (OST) OR > (OST)	Hicorieae (9)	sec. Manning (1978)
Annamocarya (25)	sec. Manos & Stone (2001)	<	*Carya* (10)	sec. Manning (1978)
Carya (26)	sec. Manos & Stone (2001)	<	*Carya* (10)	sec. Manning (1978)
Annamocarya (25) + *Carya* (26)	sec. Manos & Stone (2001)	==	*Carya* (10)	sec. Manning (1978)
Juglandinae (27)	sec. Manos & Stone (2001)	== (INT) AND > (OST)	Juglandeae (11)	sec. Manning (1978)
Cyclocarya (28)	sec. Manos & Stone (2001)	<	*Pterocarya* (13)	sec. Manning (1978)
Juglans (29)	sec. Manos & Stone (2001)	==	*Juglans* (12)	sec. Manning (1978)
Pterocarya (30)	sec. Manos & Stone (2001)	<	*Pterocarya* (13)	sec. Manning (1978)
Cyclocarya (28) + *Pterocarya* (30)	sec. Manos & Stone (2001)	==	*Pterocarya* (13)	sec. Manning (1978)

Table 6 Examples of horizontal relationships for higher-level concepts in the walnut family according to Manning (1978) versus Manos and Stone (2001; Fig. 1).

integrate information across multiple classificatory schemes; however, they can only be offered through additional textual explanations or probabilistic statements.

Conclusions

As biological data become more diverse and networked, the need to identify congruent versus incongruent taxonomic entities will match, or even exceed, the mandate to use 'valid' or 'correct' taxonomic labels (Kennedy *et al*., 2005; Berendsohn & Geoffroy, 2007; Franz *et al*., 2008). Concept relationships will therefore be essential to integrating any sort of past, present and future biological information linked to taxonomic names. The proposed terms will facilitate dynamic and perfectly archived updates of databases representing traditional and modern phylogenetic taxonomies side by side. They will provide experts with a more accurate means to communicate their latest insights into nature's relationships, and thereby help reduce an increasing and undesirable disjunction between phylogenies and classifications (Franz, 2005). Future applications should lead to further refinements of the concept relationship language, and ultimately to the development of a standard ontology adopted by major digital repositories for systematic information.

Acknowledgements

We thank the members of the SEEK Taxonomy and Classification Working Group, particularly Xianhua Liu and Jessie Kennedy, as well as Alan Weakley, for critical input shaping the content of this paper. This material is based upon work supported through a Postdoctoral Fellowship by the Andrew W. Mellon Foundation (NMF) and through Award 0225676 made by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Endnotes

[1]In this sense a taxonomic concept is not the same as a species concept. Species concepts are theories about what species are, how they arise and how to recognise them (see Wheeler & Meier, 2000).

[2]The symbols used throughout this paper are readily mapped onto existing alternatives (e.g. Koperski *et al*., 2000; Thau & Ludäscher, 2008) and are reproducible with standard word processing software packages.

[3]As pointed out by one team of referees, uncertainty also occurs in the context of transmitting natural history information along a concatenation of concept relationships (see Berendsohn *et al*., 2003).

[4]If concept C5 has multiple co-extensional synapomorphies or diagnostic features, then p(C5) and p(C5′) can be intensionally congruent without referring to the same property.

References

ATKINS, D. E., DROEGEMEIER, K. K., FELDMAN, S. I., GARCIA-MOLINA, H., KLEIN, M. L., MESSERSCHMITT, D. G., MESSINA, P., OSTRIKER, J. P. & WRIGHT, M. H. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Available at http://www.communitytechnology.org/nsf_ci_report/ Accessed VII-20–2008.

BEACH, J. H., PRAMANIK, S. & BEAMAN, J. H. 1993. Hierarchic taxonomic databases. In: FORTUNER, R., Eds., *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*, pp. 241–256. John Hopkins University Press, Baltimore, MA.

BERENDSOHN, W. G. 1995. The concept of "potential taxa" in databases. *Taxon* **44**, 207–212.

BERENDSOHN, W. G. & GEOFFROY, M. 2007. Networking taxonomic concepts – uniting without "unitary-ism". In: CURRY, G. and HUMPHRIES, C., Eds., *Biodiversity Databases: From Cottage Industry to Industrial Networks*. Systematics Association Special Volume (73). Taylor & Francis, Boca Raton, FL, pp. 13–22.

BERENDSOHN, W. G., DÖRING, M., GEOFFROY, M., GLÜCK, K., GÜNTSCH, A., HAHN, A., KUSBER, W.-H., LI, J.-J., RÖPERT, D. & SPECHT, F. 2003. The Berlin taxonomic information model. *Schriftenreihe für Vegetationskunde* **39**, 15–42.

BERNERS-LEE, T., HENDLER, J. & LASSILA, O. 2001. The semantic web. *Scientific American, May 2001*, **284**, 34–43.

CAMPBELL, C. S. 1983. Systematics of the *Andropogon virginicus* complex (Gramineae). *Journal of the Arnold Arboretum* **64**, 171–254.

CRAIG, P. & KENNEDY, J. 2008. Concept Relationship Editor: a visual interface to support the assertion of synonymy relationships between taxonomic classifications. *Visualization and Data Analysis*. (in press)

DEVITT, M. & STERELNY, K. 1999. *Language and Reality: An Introduction to the Philosophy of Language. Second Edition*. MIT Press, Cambridge, MA.

FARBER, P. L. 1976. The type concept in zoology in the first half of the nineteenth century. *Journal of the History of Biology* **9**, 93–119.

FRAHM, J. P. & FREY, W. 1992. *Moosflora, 3. Auflage*. Ulmer Verlag, Stuttgart.

FRANZ, N. M. 2005. On the lack of good scientific reasons for the growing phylogeny/classification gap. *Cladistics* **21**, 495–500.

FRANZ, N. M., PEET, R. K. & WEAKLEY, A. S. 2008. On the use of taxonomic concepts in support of biodiversity research and taxonomy. In: WHEELER, Q. D., Ed., *The New Taxonomy*. Systematics Association Special Volume (74). Taylor & Francis, Boca Raton, FL, pp. 63–86.

GANGEMI, A. & MIKA, P. 2003. Understanding the semantic web through descriptions and situations. *Lecture Notes in Computer Science* **2888**, 689–706.

GARRITY, G. M. & LYONS, C. 2003. Future-proofing biological nomenclature. *OMICS* **7**, 31–33.

GEOFFROY, M. & BERENDSOHN, W. G. 2003. The concept problem in taxonomy: importance, components, approaches. *Schriftenreihe für Vegetationskunde* **39**, 5–14.

GODFRAY, H. C. J. 2002. Challenges for taxonomy. *Nature* **417**, 17–19.

GODFREY, R. K. & WOOTEN, J. W. 1979. *Aquatic and Wetland Plants of Southeastern United States, Monocotyledons.* University of Georgia Press, Athens.

GRADSTEIN, S. R., SAUER, M., KOPERSKI, M., BRAUN, W. & LUDWIG, G. 2001. TaxLink, a program for computer-assisted documentation of different circumscriptions of biological taxa. *Taxon* **50**(4), 1075–1084.

GRAHAM, M., WATSON, M. F. & KENNEDY, J. B. 2002. Novel visualisation techniques for working with multiple, overlapping classification hierarchies. *Taxon* **51**(2), 351–358.

HAWKSWORTH, D. L., Comp. and Ed. 1994. *A Draft Glossary of Terms Used in Bionomenclature*. IUBS Monograph 9. International Union of Biological Sciences, Paris.

HITCHCOCK, A. S. & CHASE, A. 1950. *Manual of the Grasses of the United States. Second Edition*. United States Department of

Agriculture Miscellaneous Publication (200). Reprinted 1971 by Dover Publications, New York.

ITIS 2006. *Integrated Taxonomic Information System*. Available at http://www.itis.gov/ Accessed VII-20–2008.

JONES, M. B., SCHILDHAUER, M. P., REICHMAN, J. & BOWERS, S. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution and Systematics* **37**, 519–544.

KENNEDY, J., KUKLA, R. & PATERSON, T. 2005. Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. In: LUDÄSCHER, B. & RASCHID, L., Eds., *Data Integration in the Life Sciences: Proceedings of the Second International Workshop, San Diego, CA, USA, July 20–22*, pp. 80–95. DILS 2005, LNBI 3615.

KENNEDY, J., HYAM, R., KUKLA, R. & PATERSON, T. 2006. A standard data model representation for taxonomic information. *OMICS* **102**, 220–230.

KOPERSKI, M. & SAUER, M. 1999. Das Projekt "Referenzliste der Moose Deutschlands" – Dokumentation unterschiedlicher taxonomischer Auffassungen mit Hilfe des Datenbankprogrammes TAXLINK. *Stuttgarter Beiträge zur Naturkunde, Serie A* **590**, 1–10.

KOPERSKI, M., SAUER, M., BRAUN, W. & GRADSTEIN, S. R. 2000. Referenzliste der Moose Deutschlands. *Schriftenreihe für Vegetationskunde* **34**, 1–519.

LEEBENS-MACK, J., VISION, T., BRENNER, E., BOWERS, J. E., CANNON, S., CLEMENT, M. J., CUNNINGHAM, C. W., DEPAMPHILIS, C., DESALLE, R., DOYLE, J. J., EISEN, J. A., GU, X., HARSHMAN, J., JANSEN, R. K., KELLOPP, E. A., KOONIN, E. V., MISHLER, B. D., PHILIPPE, H., PIRES, J. C., QIU, Y.-L., RHEE, S. Y., SJÖLANDER, K., SOLTIS, D. W., SOLTIS, P. S., STEVENSON, D. W., WALL, K., WARNOW, T. & ZMASEK, C. 2006. Taking the first steps towards a standard for reporting on phylogenies: minimal information about a phylogenetic analysis (MIAPA). *OMICS* **10**, 231–237.

LINNAEUS, C. 1753. *Species Plantarum. Vol. 2*. Stockholm.

LUDÄSCHER, B., LIN, K., BOWERS, S., JAEGER-FRANK, E., BRODARIC, B. & BARU, C. 2006. Managing scientific data: from data integration to scientific workflows. *GSA Today Special Paper, Geoinformatics: Data to Knowledge*, **397**, 109–129.

MANNING, W. E. 1978. The classification within the Juglandaceae. *Annals of the Missouri Botanical Garden* **65**, 1058–1087.

MANOS, P. S. & STONE, D. E. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. *Annals of the Missouri Botanical Garden* **88**, 231–269.

MICHENER, W. K. & BRUNT, J. W., Eds. 2000. *Ecological Data: Design, Management, and Processing*. Blackwell Science, Malden, MA.

MICHENER, W. K., BRUNT, J. W., HELLY, J., KIRCHNER, T. B. & STAFFORD, S. G. 1997. Non-geospatial metadata for the ecological sciences. *Ecological Applications* **7**, 330–342.

MICHENER, W. K., BEACH, J. H., JONES, M. B., LUDÄSCHER, B., PENNINGTON, D. D., PEREIRA, R. S., RAJASEKAR, A. & SCHILDHAUER, M. 2007. A knowledge environment for the biodiversity and ecological sciences. *Journal of Intelligent Information Systems* **29**, 111–126.

PAGE, R. D. M. 2005. Phyloinformatics: towards a phylogenetic database. In: WANG, J. T. L., ZAKI, M. J., TOIVONEN, H. T. T. & SHASHA, D. E., Eds., *Data Mining in Bioinformatics, Advanced Information and Knowledge Processing. Volume XI*. Springer Verlag, Berlin, pp. 219–241.

PAGE, R. D. M. 2006. Taxonomic names, metadata, and the Semantic Web. *Biodiversity Informatics* **3**, 1–15.

PAGE, R. D. M. 2008. LSID Tester, a tool for testing Life Science Identifier resolution services. *Source Code for Biology and Medicine* 2008 **3**, 2.

PARR, C. S., LEE, B., CAMPBELL, D. & BEDERSON, B. 2004. Tree visualizations for taxonomies and phylogenies. *Bioinformatics* **20**, 2997–3004.

PASKIN, N. 2005. Digital object identifiers for scientific data. *Data Science Journal* **4**, 12–20.

PYLE, R. L., EARLE, J. L. & GREENE, B. D. 2008. Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. *Zootaxa* **1671**, 3–31.

RADFORD, A. E., AHLES, H. E. & BELL, C. R. 1968. *Manual of the Vascular Flora of the Carolinas*. University of North Carolina Press, Chapel Hill, NC.

SCOBLE, M. J. 2004. Unitary or unified taxonomy? *Philosophical Transactions of the Royal Society of London, Series B* **359**, 699–711.

STEVENS, P. F. 1984. Metaphors and typology in the development of botanical systematics 1690–1960, or the art of putting new wine in old bottles. *Taxon* **33**, 169–211.

THAU, D. & LUDÄSCHER, B. 2008. Reasoning about taxonomies in first-order logic. *Ecological Informatics* **3** (in press).

WEAKLEY, A. S. 2006. *Flora of the Carolinas, Virginia, and Georgia. Working Draft of 17 January 2006*. University of North Carolina, Chapel Hill, NC.

WHEELER, Q. D. & MEIER, R. 2000. *Species Concepts and Phylogenetic Theory: A Debate*. Columbia University Press, New York, NY.

Context	Relationship	Explanation	Inverse relationship
Name change	is a new name for	name 1 replaces the previously established name 2 while retaining the same type (*nomen novum*)	has a new name
	is an alternative name for	name 1 is an alternative usage for name 2 (*nomen alternativum*)	has an alternative name
	is a new combination for	name 1 is newly combined with another higher-level name to yield name 2 (*combination nova*)	has a recombination
	is a corrected name for	name 1 is a corrected version (emendation) of name 2 (*nomen correctum*)	has a corrected name
	is a replacement name for	name 1 is a replacement for the now invalid name 2 (*nomen substitutum*)	has a replacement name
Ambiregnal	is an ambiregnal name for	name 1 and name 2 are linked to the same type yet are placed in two different kingdoms and thus fall under two Codes	has an ambiregnal name
Vernacular	is a vernacular name for	name 1 is a common non-Latinised usage for name 2	has a vernacular name
Validation	is a validation of	name 1 is the valid version for name 2, although name 1 was published later	has a validation
	is a later validation of	name 1 is the valid version for name 2, although name 1 was published later and is explicitly recognised as of junior origin	has a later validation
Contradiction	is a contradiction of	name 1 is in conflict with name 2	has a contradiction
Misapplication	is a misapplied name for	name 1 is misapplied in the sense that it excludes the type of name 2	has a misapplied name
	is the same name with misapplied authorship for	name 1 is the same as name 2 except for the authorship attribution which is misapplied	is the same name with misapplied authorship for
Spelling	is a subsequent spelling for	name 1 is a later spelling of name 2	has a subsequent spelling
	is a correct subsequent spelling for	name 1 is a correct later spelling of name 2 (*orthografia correcta*)	has a correct subsequent spelling
	is an incorrect subsequent spelling for	name 1 is an incorrect later spelling of name 2	has an incorrect subsequent spelling
	is an altered spelling of	name 1 is an altered spelling of name 2 (*orthografia mutata*)	has an altered spelling
Conservation	is conserved against	name 1 is retained as valid over name 2 (*nomen conservandum*)	has a conserved name
	is proposed for conservation against	name 1 is proposed to be retained as valid over name 2 (*nomen conservandum propositum*)	has a proposed conserved name
	is a conserved spelling against	name 1's spelling is retained as valid over name 2's spelling (*orthografia conservada*)	has a conserved spelling
	is a proposed conserved spelling against	name 1's spelling is proposed to be retained as valid over name 2's spelling (*orthografia conservada proposita*)	has a proposed conserved spelling
	is a conserved type against	name 1 is retained as the valid type for name 2 (*typus conservandus*)	has a conserved type

Rejection	is a rejected name for	name 1 is rejected in favour of name 2 (*nomen rejiciendum*)	has a rejected name
	is a proposed rejected name for	name 1 is proposed to be rejected in favour of name 2 (*nomen rejiciendum propositum*)	has a proposed rejected name
	is a rejected spelling for	name 1's spelling is rejected in favour of name 2's spelling (*orthografia rejicienda*)	has a rejected spelling
	is a proposed rejected spelling for	name 1's spelling is proposed to be rejected in favour of name 2's spelling (*orthografia rejicienda proposita*)	has a proposed rejected spelling
	is a rejected type for	name 1 is rejected as the type for name 2 (*typus rejiciendus*)	has a rejected type
Typification	is type of	name 1 is selected as the type name 2	has a type
	is an epitypification of	name 1 is selected as the epitype for name 2	has an epitypification
	is a lectotypification of	name 1 is selected as the lectotype for name 2	has an epitypification
	is a neotypification of	name 1 is selected as the neotype for name 2	has a neotypification
Basionymy	is a basionym for	name 1 represents the proper stem for name 2 via the Rule of Priority	has a basionym
Protonymy	is a protonym for	name 1 was effectively published yet name 2, with the same string and published later, is valid	has a protonym
Homonymy	is a homonym of	name 1 has the same string as name 2, yet the respective taxa are not the same	has a homonym
	is a primary homonym of (zool.)	name 1 has the same string as name 2 and referred to the same genus-level name when initially published, yet the respective taxa are not the same	has a primary homonym
	is a secondary homonym of (zool.)	name 1 has the same string as name 2 and referred to another genus-level name when initially published, yet the respective taxa are not the same	has a secondary homonym
	is a senior homonym of (bact., zool.)	name 1 is the earliest published homonym of name 2	has a senior homonym
	is a junior homonym of (bact., zool.)	name 1 is one of the later published homonyms of name 2	has a junior homonym
	is an earlier homonym of (bot.)	name 1 has the same string as the subsequently published name 2, yet the respective taxa are not the same	has an earlier homonym
	is later homonym of (bot.)	name 1 has the same string as the previously published name 2, yet the respective taxa are not the same	has a later homonym
	is treated as later homonym of (bot.)	name 1 has the same string as the previously published name 2 and is treated explicitly as of junior origin, yet the respective taxa are not the same	has a homonym treated as later
	is an isonym of	name 1 is a later publication of name 2 rooted in the same type	has an isonym

Appendix 1 List of terms commonly used to express nomenclatural relationships.[1]

Context	Relationship	Explanation	Inverse relationship
Synonymy	is a synonym for	name 1 is a synonym of name 2	has a synonym
	is a holonym for	name 1 is a synonym of name 2 representing a similar range of variation	has a homonymy
	is a hyponym for	name 1 is a synonym of name 2 representing a smaller range of variation	has a hyponym
	is a hypernym for	name 1 is a synonym of name 2 representing a greater range of variation	has a hypernym
	is a homotypic synonym for	name 1 is a synonym of name 2 referring to the same type	has a homotypic synonym
	is a heterotypic synonym for	name 1 is a synonym of name 2 referring to a different type	has a heterotypic synonym
	is a nomenclatural synonym for (bot.)	name 1 is a synonym of name 2 referring to the same type	has nomenclatural synonym
	is a taxonomic synonym for (bot.)	name 1 is a synonym of name 2 referring to a different type	has a taxonomic
	is an objective synonym for (zool.)	name 1 is a synonym of name 2 referring to the same type	has an objective synonym
	is a subjective synonym for (zool.)	name 1 is a synonym of name 2 referring to a different type	has a subjective synonym
	is a partial synonym for	name 1 is a synonym of name 2 including part of name 2's entire taxonomic range	has a partial synonym
	is partial and homotypic synonym for	name 1 is a synonym of name 2 including the same type and part of name 2's entire taxonomic range	has a partial and homotypic synonym
	is partial and heterotypic synonym for	name 1 is a synonym of name 2 including a different type and part of name 2's entire taxonomic range	has a partial and heterotypic synonym
	is a pro parte synonym for	name 1 is a synonym of name 2 including part of name 2's entire taxonomic range	has a pro parte synonym
	is pro parte and homotypic synonym for	name 1 is a synonym of name 2 including the same type and part of name 2's entire taxonomic range	has a pro parte and homotypic synonym of
	is pro parte and heterotypic synonym for	name 1 is a synonym of name 2 including a different type and part of name 2's entire taxonomic range	has a pro parte and heterotypic synonym of
	is a replaced synonym for	name 1 has been replaced by name 2	has a replaced synonym
Life cycle	is an anamorph of (bot.)	name 1 refers to the asexual or mitotic reproductive stage in a pleomorphic life cycle in which name 2 refers to the teleomorph or meiotic reproductive stage	has an anamorph
	is a teleomorph of (bot.)	name 1 refers to the teleomorph or meiotic reproductive stage in a pleomorphic life cycle in which name 2 refers to the asexual or mitotic reproductive stage	has a teleomorph
Negation	not	used to negate any of the aforementioned nomenclatural relationships	not applicable

Appendix 1 Continued.

[1]Not exhaustive; taken mainly from Hawksworth (1994) and Berendsohn *et al.* (2003). Precise definitions of these and other potentially useful relationships are available in the respective *Codes*.