



This work is licensed under a Creative Commons Attribution License (CC BY 4.0).

Collection management

An Xper3 reference guide for taxonomists: a collaborative system for identification keys and descriptive data

Adeline KERNER^{*1}, Elie Mario SALIBA², Sylvain BOUQUIN³,
Rémy PORTIER⁴ & Régine VIGNES-LEBBE⁵

¹UMR 7207 – Centre de Recherche en Paléontologie de Paris (CR2P), CNRS, Sorbonne Université, Muséum national d’Histoire naturelle, Paris, France.

^{2,3,5}UMR 7205 – Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d’Histoire naturelle, CNRS, Sorbonne Université, EPHE, UA, Paris, France.

⁴UAR 3468 – Bases de données Biodiversité, Écologie, Environnement et Sociétés (BBEES), CNRS, Muséum national d’Histoire naturelle, Paris, France.

*Corresponding author: kerner@mnhn.fr

² Email: elie.saliba@mnhn.fr

³ Email: sylvain.bouquin@sorbonne-universite.fr

⁴ Email: remy.portier@mnhn.fr

⁵ Email: regine.vignes_lebbe@sorbonne-universite.fr

Abstract. The Xper3 software enables collaborative management of descriptive data on taxa. This paper focuses on the essential information about the knowledge model that underlies the software, its features and user interface which has been evolving since 2014. The best practices for representing taxonomic descriptions, managing polymorphism and missing data in a consistent way are also presented. Its ease of use and the versatility of its data model mean that the software can be used to create and distribute identification keys for all types of taxa.

Keywords. Xper3, SDD format, identification key, taxonomic descriptions, collaborative online taxonomy.

Kerner A., Saliba E.M., Bouquin S., Portier R. & Vignes-Lebbe R. 2025. An Xper3 reference guide for taxonomists: a collaborative system for identification keys and descriptive data. *European Journal of Taxonomy* 987: 281–302. <https://doi.org/10.5852/ejt.2025.987.2875>

Introduction

Managing data and making scientific knowledge available is a long-standing concern. It has led to an increase in the use of databases and software in all fields, particularly in life sciences. The establishment of an international group, the Taxonomic Data Working Group, now called Biodiversity Information Standards, to propose standards and data formats for taxonomy dates back to the 1980s (<https://www.tdwg.org/>). The creation of the Global Biodiversity Information Facility (<https://www.gbif.org/>) in 2001 and its current expansion to more than 60 countries, is also evidence of this reality.

Morpho-anatomical data is one of the types of data that has been increasingly computerised in the 1980s and 1990s and is the basis for the application of identification algorithms. The Delta system (Dallwitz 1980) was one of the forerunners in this field, as well as the work of R. Pankhurst (1970), R. Vignes and J. Lebbe (Lebbe *et al.* 1989) and the LucID team (Norton *et al.* 2000). Molecular data have aided the delimitation of species and the identification of specimens with the development of barcoding (Hebert & Gregory 2005) and metabarcoding (Taberlet *et al.* 2012). Images and sounds have also become more widely available thanks to new acquisition and analysis methods (Aide *et al.* 2013).

The International Congress “Tools for Identifying Biodiversity: Progress and Problems” (Nimis & Vignes-Lebbe 2010), which took place in Paris, was an opportunity to review the diversity of needs, methods and systems available. More than a decade later, given the boom in artificial intelligence, automatic statistical systems would probably occupy a much larger place in such a congress. However, while their efficiency is constantly increasing, their explanatory and pedagogical powers remain weak (Picek *et al.* 2022; Bartlett *et al.* 2022). Proposing a way to manage, analyse, and disseminate phenotypic data in a structured form, intelligible by humans, and processable by algorithms remains crucial for taxonomic identification and its accompanying automated methods (Hagedorn 2007).

The sheer quantity of the data available in digital format, and the development of high-performance querying techniques (Agathokleous *et al.* 2023) could have led to the belief that the need to store structured data would be largely in decline, but this is not the case. The requirements of traceability, reproducibility, and openness of scientific data (Wilkinson *et al.* 2016; Open Science 2023) led an ever-growing community to be involved in its computerisation.

The Xper3 web application that is presented in this article is dedicated to managing taxonomic description data in a structured form (Vignes Lebbe *et al.* 2016). It is designed to assist in the taxonomic review and the identification of specimens. Its online availability recently reached the 10th anniversary and an ever-increasing community has adopted Xper3, proof of its appeal and ease of use. Xper3 was developed and is maintained in Paris by a part of the LIS (Laboratory of Informatics and Systematics) team.

The multiplication of data papers focused on content and identification keys* made with Xper3 (Dubuisson *et al.* 2018; Jouveau *et al.* 2018; Bodin *et al.* 2019; Del Rio *et al.* 2020; Lombard *et al.* 2021; Goharimanesh *et al.* 2021; Kerner 2022; Palomares-Rius *et al.* 2022; Schneider *et al.* 2023; Hilgenhof *et al.* 2023) as well as the occasional requests from their authors to the LIS team suggest the need for a reference article on the software and its features. This is what will be provided in this article, with examples of uses in different contexts.

Material and methods

The software architecture, the data model and the main mathematical functions on which Xper3 is based are described below. Xper3 users are split into 2 categories: knowledge base* creators that provide content to generate identification keys, and the identification key end users. From now on, they will be referred to as content creators and content users, respectively.

Technical landscape

Xper3 is a web application accessible via a browser with no specific installation required. The data storage layer is based on a relational database management system. For data exchange, Xper3 uses CSV or XML files. For the XML schema, Xper3 follows the SDD (Structured Descriptive Data) international standard (Structure of Descriptive Data Subgroup 2005) defined by the TDWG community (<https://www.tdwg.org/>).

Table 1. Xper3 technical architecture.

Data storage	Relational database Maria DB
Object/Relational Mapping	Hibernate framework
Frontend interface	HTML, Ajax, jQuery and Spring framework
Exchange formats	CSV, SDD (XML)
Identification services	REST-type services
Backend code	Java

Xper3 includes a web collaborative editor with several functions or tools, to import, export, compare and check the data. The main services are related to computer-aided identification with two software components: Mkey* for interactive identification and Ikey* for simple-access identification key generation (Burguiere *et al.* 2013). These two services, coded in Java, do not store any data but take as input an Xper3 database in SDD format. They can be included in your own website as explained in the API documentation (<https://xper3.fr/en/documentation-2/documentation-in-pdf-format/>) (for an example, see <https://dynafor.toulouse.inra.fr/carabagri/index.php?objet=determination>). More details on the technical landscape of Xper3 are provided in Table 1.

Formal background and data model

Xper3 is based on the formalisation of taxonomic descriptions proposed by Lebbe (1991). The software provides a generic knowledge model, which is a kind of metamodel, depicting how to decompose a description*. Each knowledge base* implements the metamodel in the form of a descriptive model centred on the taxonomic group of interest.

Generic knowledge model (metamodel)

The terms used in Xper3 are listed in the glossary of Table 2. Each term included in the glossary appears with * in the text. Figure 1 is a conceptual schema of the Xper3 knowledge model.

The metamodel distinguishes the items*, which are the units to be described (most often specimens or taxa), and a descriptive model* (including descriptors* and descriptor states*) to structure the terminology of each knowledge base*. The list of the items and their descriptions expressed according to the descriptive model*, constitute the data of an Xper3 knowledge base.

The software is primarily designed to manage descriptions of taxa. However, it can be used to manage data of different natures: specimens, species, genera, diseases, artistic works, etc.

In taxonomy the term ‘character’ is classically used to designate characteristics of specimens or taxa but that sometimes leads to confusion between the notion of character and the notion of character state. In Xper3 the terms descriptor and states were chosen to name these two concepts.

In the descriptive model, each descriptor has a type (categorical or numerical), a set of possible values or states (its domain) and conditions of applicability* which are expressed as a rule. For example, the descriptor “number of leaflets” has an applicability condition which is the rule “the number of leaflets is applicable only if the descriptor ‘type of leaves’ has the value ‘compound leaves’”.

Xper3 also allows special categorical descriptors* named calculated descriptors*: in that case, each state corresponds to a logical combination of states or values of other descriptors (see the ‘Descriptive model’ tab in the Results section).

Table 2. Glossary of the terms used in Xper3.

Term	Definition
applicability or inapplicability	Status of a descriptor in a description. A descriptor is applicable in a description if the description does not contain an eliminatory state as an attribute value for the parent descriptor.
attribute	Data associating an item and a descriptor. In the case of a categorical descriptor, an attribute can be one or a set of states if expressing polymorphism is needed for the item. In the case of a numerical descriptor, the attribute is a statistical distribution of values expressed by its minimum and maximum or by its mean and standard deviation. Syn. element description.
calculated descriptor	A descriptor whose values are automatically calculated from other data using logical (Boolean) operators.
categorical descriptor	A descriptor taking as possible values qualitative states, or numeric intervals. Examples: “Blade shape”; “Number of antenna articles” with states “less than five”; “five to ten”; “more than ten”. A descriptor that takes qualitative states or numerical intervals as possible values. Examples: “blade shape”; “number of antennae articles” with states “less than five”; “five to ten”; “more than ten”.
checkbase	A tool to check the consistency of the descriptions and to compare them in pairs.
description	Set of attributes relating to the same item.
descriptive model	Set of descriptors with their relationships (hierarchy) and their list of states or metrics. The descriptive model defines the common terminology for all the descriptions in a knowledge base.
descriptor	Element used to describe items, essentially a character for taxonomists, or a trait for ecologists.
discriminant power	Ability of a descriptor to distinguish between items.
elimination state	State used in a rule to define the condition of inapplicability of a descriptor.
extension	Volume of the representation space corresponding to the description of an item. It is therefore the set of single descriptions (without variability, one value per descriptor) compatible with the item description.
group	Grouping, under a chosen heading, a set of objects of the same nature; a set of descriptors, or a set of items.
Ikey+	Web service for generating a single-access identification key.
item	Object which can be described in a knowledge base. For taxonomists, it usually is a taxon.
key	A practical means to identify a specimen. We distinguish the interactive key (see Mkey+) and single access key (see IKey+).
knowledge base	Dataset in an Xper3 database.
Mkey+	Web service for multi-access interactive identification key.
numerical descriptor	A descriptor expressing a continuous numerical value, with a measurement unit, e.g., antenna length in mm.
parent descriptor	A descriptor used to define the rule of inapplicability of another descriptor (child descriptor).
state (categorical state)	A possible value of categorical descriptors, e.g., for a descriptor named “Colour of the eye”, its states could be “Blue”, “Black”, etc.
unknown value	Special value to express ignorance. In case of an unknown value, Xper3 considers all the states or all the numerical values as equally possible.

An item description* is a set (or tuple) of descriptive elements (called attributes* in J. Lebbe’s formalism), one for each descriptor. An attribute* is a set of states for a categorical descriptor, a statistical distribution for a numerical descriptor* or one of the special values to express ignorance or inapplicability*. This knowledge representation allows polymorphism to be taken into account (e.g., intra-specific variation in shape, colour, size, etc.).

Taxon-centred model

The creation of an Xper3 application requires the instantiation of the metamodel. The descriptors, states and dependencies constitute an ontology (or reference vocabulary) in the context of the knowledge base. Each concept (item, descriptor, state) has a title, a unique identifier, a definition and associated resources that can include external links and illustrations (Vignes Lebbe *et al.* 2016). An example is provided below.

The description D of item B (Table 3) is thus the following tuple:

D(item B) = ((wings colour, {red, blue}), (antenna shape, {elongated}), (number of cerci, {2}), (surface of wings, {scales, hairs}))

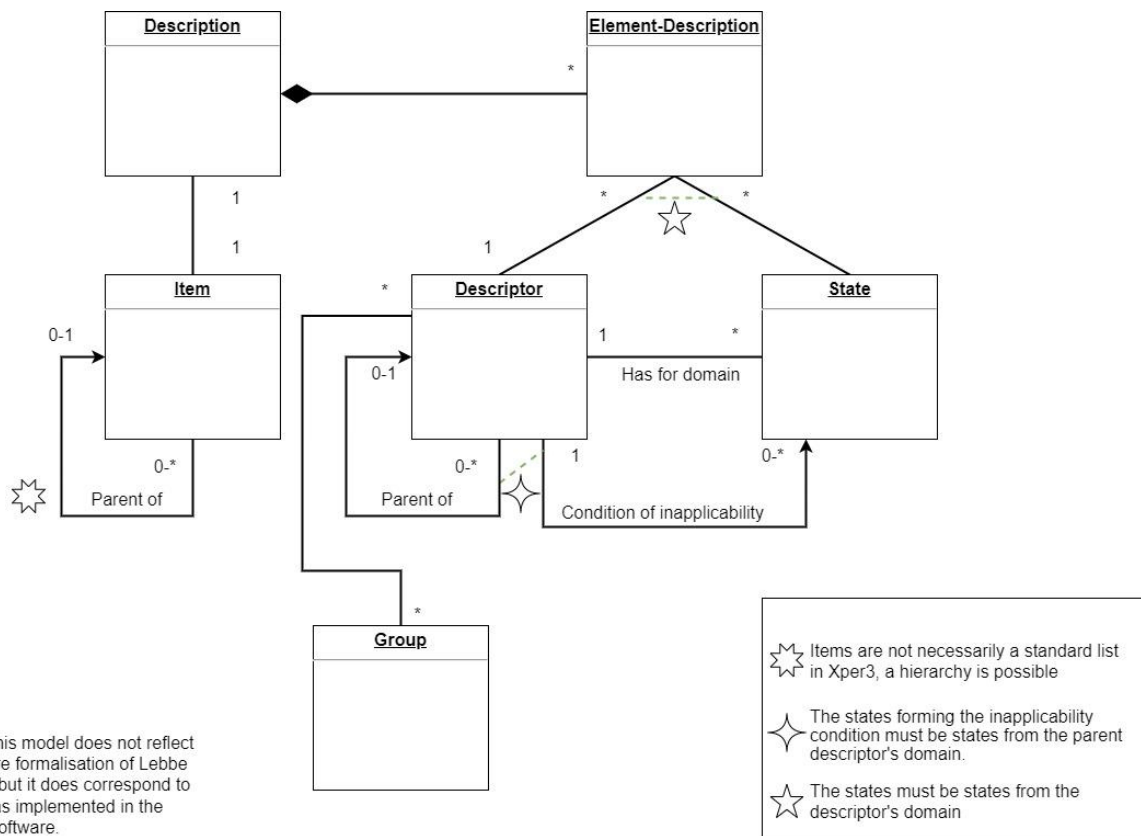


Fig. 1. UML schema of the Xper3 generic knowledge model. The main concepts are items, descriptors and states. The group* allows the creation of sets of descriptors, i.e., to group all the descriptors linked to the same entity (flower, leaf, fruit, head, thorax, abdomen, wings, etc.). Element-description is equivalent to an attribute, it means a set of states or a numerical interval. The Description of an item is a set of Element-Description.

Table 3. Small data set. Each cell (at the cross of an item and a descriptor) is an attribute. The complete description of an item is the conjunction or tuple of its attributes.

Descriptor	Species A	Species B	Species C	Species D	Species E
Wings colour	red	red; blue	blue	red	blue
Antenna shape	<i>unknown</i>	elongated	in clubs	elongated	in clubs
Number of cerci	2	2	3	2	3
Surface of wings	scales	scales; hairs	hairs	scales	hairs

Metadata

The metadata of an Xper3 database includes all the information defining its perimeter and authorship (taxonomic, geographic, development stage limits, authors, license,...) and other information that might be useful in the context of a data paper.

Main mathematical functions

Aside from the publication of identification keys, one of the main goals of Xper3 is to help taxonomic revision by providing an efficient way to compare descriptions. All these purposes call for functions to compare both attributes and descriptions.

Comparison of two attributes: $\text{comp}()$

Let $d_k(i)$ be the attribute for the item i and descriptor k .

Let i and j be two items.

The comparison function $\text{comp}()$ used by Xper3 is a Boolean measure of incompatibility: two items i and j are incompatible for a descriptor k if they have no possible state in common for k . If the items are incompatible for k , $\text{comp}_k(i, j)$ will return 1; otherwise, $\text{comp}_k(i, j)$ will return 0.

In other terms, if k is a categorical descriptor, $\text{comp}_k(i, j) = 1$ if the intersection between the list of states for i and the list of possible states for j is empty. If k is a numerical descriptor, it means that the intervals for i and j are disjoint.

$$d_k(i) \cap d_k(j) = \emptyset \Leftrightarrow \text{comp}_k(i, j) = 1$$

$$d_k(i) \cap d_k(j) \neq \emptyset \Leftrightarrow \text{comp}_k(i, j) = 0$$

Remark: $\text{comp}_k(i, j)$ is computable only if the descriptor k is applicable for the two items.

Discriminant power of a descriptor: $\text{PD}()$

The discriminant power* of a descriptor k , $\text{PD}(k)$ is the result of the sum of the $\text{comp}_k()$ measures for all pairs of items described by this descriptor.

$$\text{PD}(k) = \sum_{x,y} \text{comp}_k(i_x, i_y)$$

Note: in the case of a parent descriptor*, its discriminant power is the maximum of the discriminant powers of its child(ren) descriptor(s).

Extension or coverage of a description: $\text{ext}()$

Let i and j be two items, and D_i be the description of i for a set of descriptors.

We define the extension* of a description $\text{ext}(D_i)$ as the set of tuples with only one value for each descriptor. It is equivalent to all the possible instances (variations) of the item i . If the item is a species, the extension of its description is the set of all the possible specimen descriptions that are compatible with the species description (cf. Table 3). In the case of categorical descriptors, this extension is the Cartesian product of attributes. In the case of N numerical descriptors, the extension is a hypervolume in the N dimensions.

Comparison of two item descriptions: $\text{dist}()$

The function to compare two item descriptions $\text{dist}(i, j)$ depends on the extension of each one, $\text{ext}(D_i)$ and $\text{ext}(D_j)$. In Xper3, this function takes into account dependencies (applicability of descriptors) and distinguishes between four results (Checkbase function*, Fig. 2):

Equal items: items with the same descriptions, and thus the same extensions.

$$\text{dist}(i, j) = \text{equal} \Leftrightarrow \text{ext}(D_i) = \text{ext}(D_j)$$

Distinct items: items with no overlapping extensions.

$$\text{dist}(i, j) = \text{distinct} \Leftrightarrow \text{ext}(D_i) \cap \text{ext}(D_j) = \emptyset$$

Overlapping items: items with overlapping extensions, but without the inclusion of one extension in the other.

$$\text{dist}(i, j) = \text{overlap} \Leftrightarrow \text{ext}(D_i) \cap \text{ext}(D_j) \neq \emptyset \text{ AND } \text{ext}(D_i) \cap \text{ext}(D_j) \notin \{\text{ext}(D_i), \text{ext}(D_j)\}$$

Including items: one extension covers completely the extension of the second item.

$$\text{dist}(i, j) = \text{included} \Leftrightarrow \text{ext}(D_i) \cap \text{ext}(D_j) \notin \{\text{ext}(D_i), \text{ext}(D_j)\}$$

The extension of the species A description (Table 3) includes the following set of tuples (the unknown value* is interpreted as all states of the descriptor domain being potentially possible):

$$\text{ext}(D_A) = \{((\text{wings colour}, \{\text{red}\}), (\text{antenna shape}, \{\text{elongated}\}), (\text{number of cerci}, \{2\}), (\text{surface of wings}, \{\text{scales}\})), ((\text{wings colour}, \{\text{red}\}), (\text{antenna shape}, \{\text{in clubs}\}), (\text{number of cerci}, \{2\}), (\text{surface of wings}, \{\text{scales}\}))\}$$

Checkbase

The screenshot shows the Checkbase interface. On the left, there are two dropdown menus: 'Items checking' and 'Descriptions checking'. The 'Items checking' dropdown is open, showing three options: 'Items with same descriptions (1)', 'Items sometimes discriminated and included (2)', and 'Items sometimes discriminated but not included (1)'. On the right, there are three panels displaying results:

- Items with same descriptions (1)**: Species E as same description than Species C
- Items sometimes discriminated and included (2)**: Species D has his description included in Species B; Species D has his description included in Species A
- Items sometimes discriminated but not included (1)**: Species A can be describe as Species B

Fig. 2. Items checking from Checkbase applied to Table 3 dataset.

During the identification process, a specimen is compatible with a taxon (or item) if the description of the specimen is included in the extension of the description of the taxon. If a specimen is compatible with only one item, then it is identified with the corresponding taxon.

Merging function to generalise descriptions: merge()

Let i and j be two items and D_i and D_j be their descriptions.

$\text{merge}(i, j)$ creates a new description D_{ij} for the set $\{i, j\}$ where:

$$\text{merge}(i, j) = D_{ij} \text{ with } \forall k, d_k(ij) = d_k(i) \cup d_k(j)$$

It must be kept in mind that this merging function may produce an overly broad result, resulting in a larger extension for the merged description compared to the union of the extensions of the original descriptions. This merged description can then overlap with the description of other items (e.g., merging Species C and D of Table 3 will result in an item description that includes all the other species' descriptions of the table).

Results

The Xper3 platform went online in October 2013, the first demo was done during the TDWG 2013 international conference, and then presented during the UNESCO 2014 International Conference on Botany (Vignes Lebbe *et al.* 2016; Kerner *et al.* 2021). Content creators are invited to sign up for a personal account at <https://xper3.fr>. This allows the creation and editing of knowledge bases, which can be potentially shared for collaborative work, and published online as identification keys for external users.

The system's main functionalities and specific features are described below, and some recommendations for creating a knowledge base are given. We will focus on the points that raised questions from content creators. More details can be found in the online documentation (<https://xper3.fr/en/documentation-2/>).

A general overview of the editing interface

The editing interface has been developed with particular care, to ensure ease of use. It consists of a menu and 6 tabs: 'Items', 'Descriptive model', 'Description', 'Identification', 'Tools' and 'Properties & management'.

The 'Items', 'Descriptive model', and 'Description' tabs manage the content creation; the 'Identification' and 'Tools' tabs handle the content's use and analysis, and the 'Properties & management' tab deals with metadata.

All elements of the knowledge model can be documented using text and images. Xper3 does not store images, only URLs, captions and automatically generated thumbnails. As a result, all images must be accessible online, and content creators need only supply URLs and captions.

With the mouse right-click, the content creators have access to the operations available for each Xper3 object (items, descriptors, states, etc.). The basic operations are adding, deleting, editing, and copying.

On the one hand, the storage in a relational database automatically saves the data. On the other hand, the Undo option is unavailable.

Creating a knowledge base step-by-step

Creating content needs at least four main steps followed by additional steps to check the knowledge base and publish keys.

(1) Scoping of the database: defining the taxonomic and geographical coverage and, if necessary, stratigraphic delimitation, developmental stage, etc. This metadata is edited in the ‘Properties & management’ tab. The scope of the database is reflected in the list of items edited in the ‘Items’ tab.

(2) Defining a descriptive model: defining the descriptors, states and logical dependencies in the ‘Descriptive Model’ tab. The labels chosen for each element of this model standardise the terminology. The states for each categorical descriptor must be an exhaustive list, in order to cover all the observable diversity. They must also be mutually exclusive to ensure correct identification. For a colour descriptor, a “pale blue” state would be impossible to distinguish from “blue”, so there should not be a “blue” and a “pale blue” state for the same descriptor. In addition, if an organ can present several colours at the same time, then this needs to be treated as a specific state, e.g., by creating “blue and red” in addition to “blue only” and “red only”.

(3) Editing the description of the items: it is accessible via the ‘Description’ tab. For each taxon, the description must cover the full diversity of the specimens by selecting at least one state per applicable descriptor. In the case of missing information, Xper3 offers to store this as the special value ‘unknown’. During identification, an item with an empty descriptor is not treated in the same way as one with the value ‘unknown’.

(4) Illustrating and documenting all the objects of the model. It is possible to add one or more images, via URL, for each item, descriptor and state. You can add captions. Thumbnails will be created to display the images. It is also possible to add textual details for items, descriptors and states. HTML tags can be included in these detail sections, especially to add hyperlinks.

(5) Verifying and testing the content. From the ‘Tools’ tab, various analysis functions allow checking the data quality of a database using the `comp()` and `dist()` functions. ‘Checkbase’ tests the consistency and completeness of descriptions, and searches for duplicate or overlapping descriptions. It is also possible to compare items or to compare groups of items; the result appears as a matrix of items X descriptors with different colours according to the value of `comp()`.

(6) Testing identification. The interactive identification service and the algorithm for generating unique access keys can be accessed via the ‘Identification’ tab.

(7) The main menu (top left button) allows the export of the data and the online publication of the identification.

Details of the tabs

The ‘Items’ tab

The ‘Items’ tab allows the content creator to enter the list of items and edit the different information related to each one (Fig. 3). The list can be flat or hierarchical to manage taxa of different ranks. Merging items (see `merge()` function above) makes it possible to infer the description of a group (e.g., the description of a genus from the descriptions of included species). The copy function duplicates an item and its description, making it possible to create an entry identical to another, when most of the characters are shared, and then modify only the attributes that differ between these two items. Both functions can facilitate the update of data in case of a modification of the classification (merging in case of lumping, copying followed by editing in case of splitting).

In the case of sexual dimorphism or geographical variants, it is recommended to represent a taxon by several morphotypes (e.g., for a bird species: one item for an adult male, another for an adult female, and a third for a juvenile).

The ‘Descriptive model’ tab

The ‘Descriptive model’ tab follows the same principle as the ‘Items’ tab for the descriptors and their states (Fig. 4). In addition to the basic knowledge management operations (editing, deleting, modifying), duplication and merging operations are also available for descriptors. Merging descriptors allows the modification of an initial descriptive model without losing the already present description information. Copying a descriptor creates a new descriptor with the same list of states and identical associated resources. The item descriptions are also replicated for the descriptor and its copy.

The domain of a categorical descriptor is a list of different states that can be added or edited in the ‘States’ tab on the right. Numerical descriptors are used for continuous values, with a unit of measurement. Metameric characters should be used with categorical descriptors, as they are not associated with a unit of measurement, by dividing the range into distinct intervals (e.g., number of segments in an antenna).

The inapplicability rule between descriptors can be edited in the ‘Dependencies’ tab by selecting the parent descriptor’s state(s) that invalidates the child descriptor (Fig. 5). Descriptors can also be arranged into groups by using the group view. These groups do not represent a hierarchical relationship between descriptors and are simply a convenient way of sorting descriptors. For example, managing all the descriptors about the wings or the head can be made easier by creating groups.

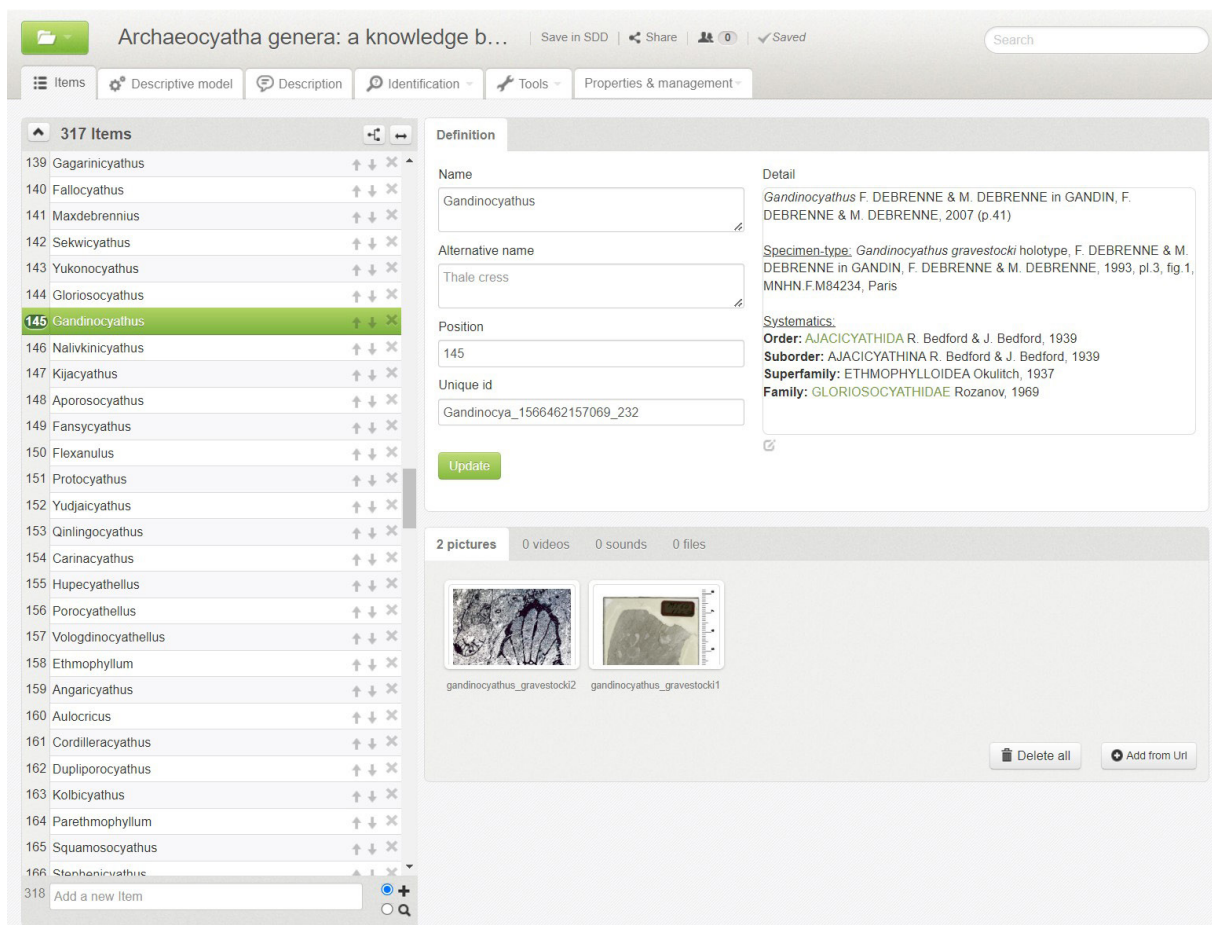


Fig. 3. The ‘Items’ tab. The left side displays a list of items, and the right side describes the various data for the items selected in the list. On the right side of the screen, the ‘detail’ field allows the user to add supplementary information. HTML formatting is allowed, and active links can be added.

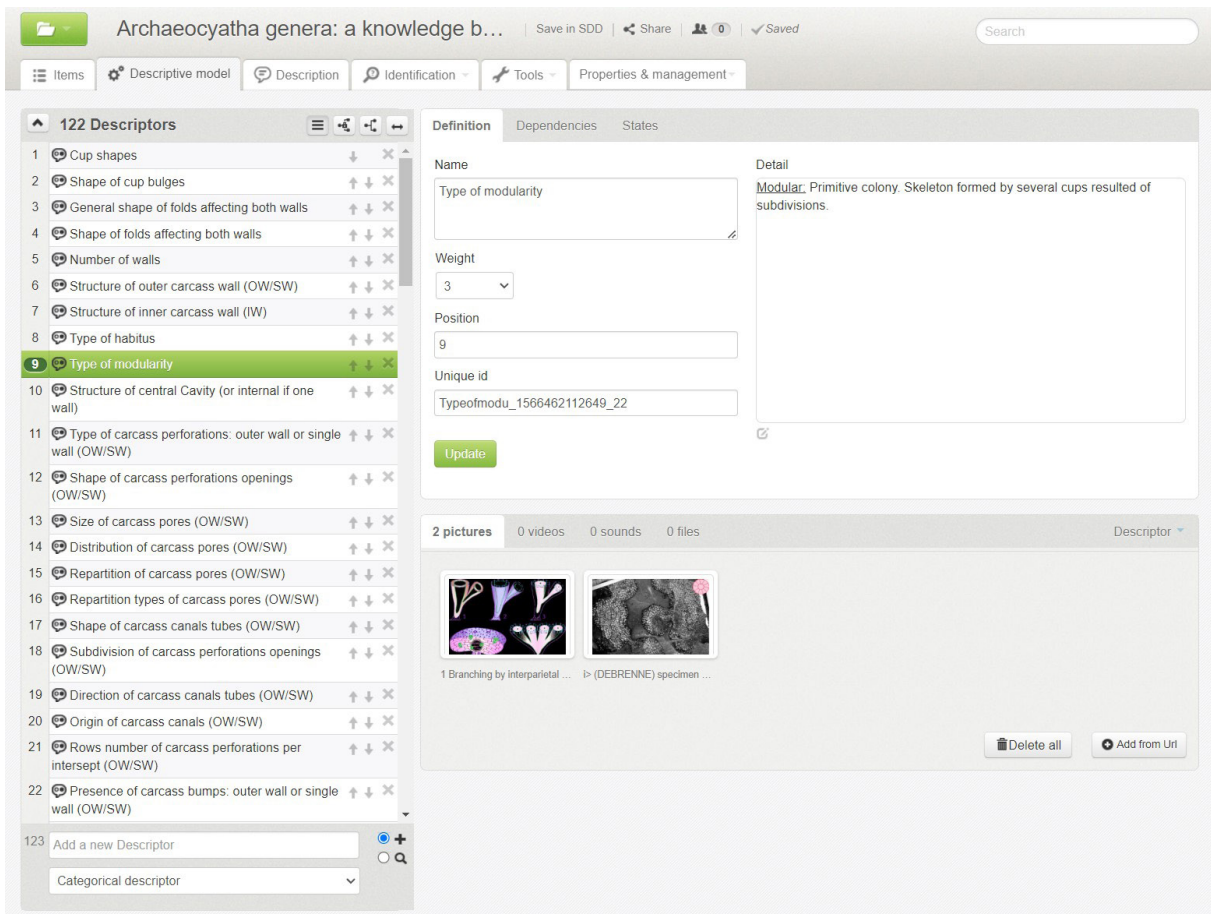


Fig. 4. ‘Descriptive model’ tab. The left side displays the list of descriptors, and the right side describes the various data for the descriptor selected in the list. On the right side of the screen, the ‘detail’ field allows the user to add supplementary information. HTML formatting is allowed, and active links can be added.

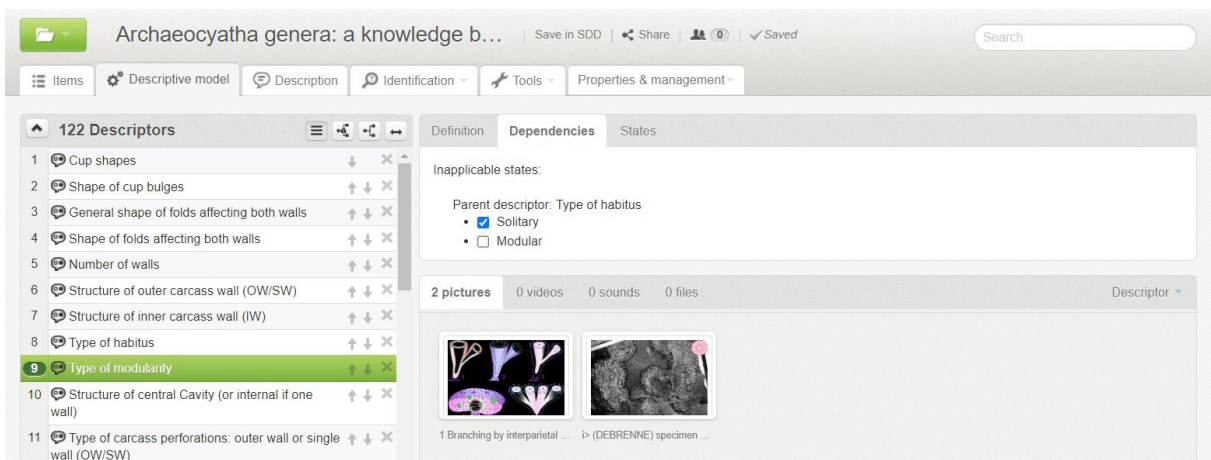


Fig. 5. The ‘Dependencies’ tab is used to define conditions that render the selected descriptor inapplicable. Some animals can either be solitary or live in colonies.

By default, descriptors weigh 3 (out of 5). A higher weight means that the descriptor has priority when it comes to identification.

If the “Type of habitus” (parent descriptor) is “solitary”, then the “Type of modularity”, which is dependent on having a colonial lifestyle is inapplicable for this item. There can only be one parent descriptor in a dependency rule.

Calculated descriptors are automatically computed from other descriptors by using logical (Boolean) operators (Fig. 6). The list of available conditions for generating calculated descriptors is AND, OR, SELECTED, NOT SELECTED. Combinations of conditions are possible (Kerner & Vignes Lebbe 2016, 2019), but without parentheses; conditions are therefore evaluated in the order of the list on the interface and not according to the order of prevalence of the AND/OR operators.

The ‘Description’ tab

The ‘Description’ tab allows each item to be assigned states according to the descriptive model, either individually or by batch. One or more items (Fig. 7) can be selected in the left column. The descriptor of interest is in the middle column and the corresponding states are in the rightmost column. In the case of a categorical descriptor, one or more states can be selected to describe a given item or set of items. For numerical descriptors, the interval must be defined by at least one of the following pairs of values: (1) Min (minimum observed value) and Max (maximum observed value), (2) Mean and SD (standard deviation), assuming a normal distribution, or (3) Umeth Lower (lower value of the usual distribution) and Umeth Upper (higher value of the usual distribution). During identification, the range used is primarily [Min, Max] if these values are filled, followed by [Umeth Lower, Umeth Upper], then [Mean - 2SD, Mean + 2SD].

Nothing needs to be done for calculated descriptors. Values are logically derived from other information.

The ‘Unknown values’ check box must be ticked if no data are available.

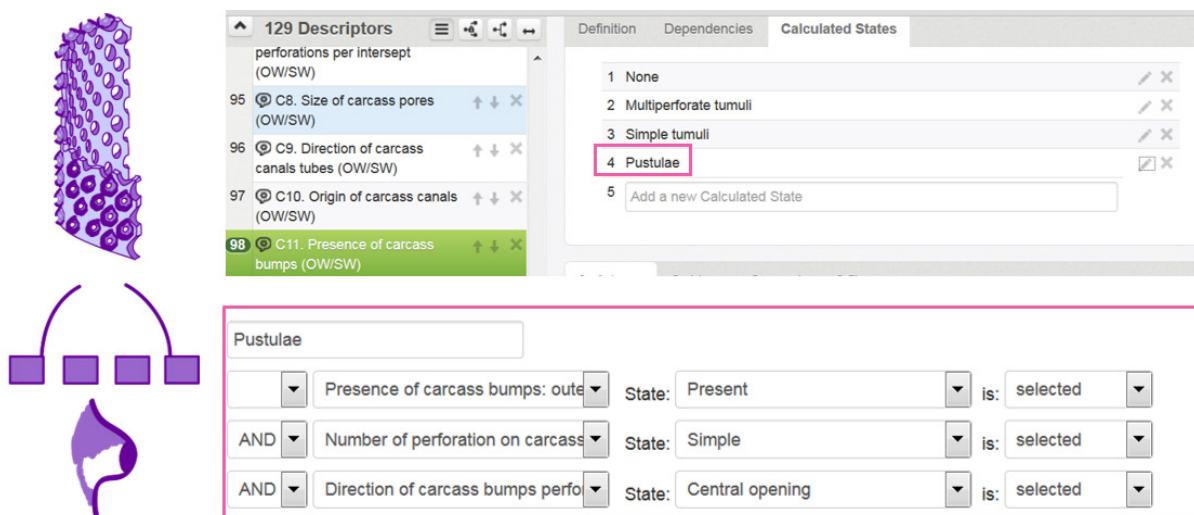


Fig. 6. Example of the interface to define the states of the calculated descriptor “Presence of carcass bumps (Outer Wall/Single Wall)”. The state “Pustulae” (in purple) is true in a description if the following condition is valid: “Presence of carcass bumps” is “present” AND “The number of perforations...” is “simple” AND the “Direction of carcass bumps...” is “central opening”.

Use and analysis of the content

The ‘Tools’ tab

The ‘Tools’ tab provides access to the analysis section.

‘View Description Matrix’ offers a global view and another way to modify the data by editing each cell of the matrix.

‘Compare Items’ and ‘Compare Groups’ allow the user to compare two or more selected items. The result is displayed as a matrix with different colours depending on whether a descriptor differentiates between the selected items (see `comp()` function). You can filter the matrix according to three values: (1) discrimination (at least one pair of selected items can be distinguished by this descriptor); (2) partial discrimination (at least one pair of selected items can be distinguished by this descriptor, i.e., there is at least one possible common value for the two items); (3) no discrimination (no pair of selected items can be distinguished by this descriptor, i.e., all descriptions for this given descriptor are the same). In the case of groups of items, the comparison is the same, but only takes into account the states that are shared within each group.

‘Checkbase’ is an automatic search tool that detects inconsistencies and checks whether items are distinguishable.

‘Xper1 tools’ refers to some tools developed for the first version of Xper, including: Extracting a possible minimal list of descriptors sufficient to discriminate all items (‘Minimal descriptors set’);

Displaying a diagnosis or minimal description of an item that distinguishes it from all others (‘Diagnosis’);

Suggesting how to merge states of a descriptor without losing discriminative power (‘Merging states’);

Generating the distance matrix between the items (‘Item distance matrix’);

Generating the distance matrix between descriptors (‘Descriptor distance matrix’).

Please note that these tools, originally developed for Xper1, have to respect the limitations of the content of this version. The tools only consider qualitative (categorical) descriptors and the labels of the different concepts (items, descriptors, states) must not exceed 255 characters.

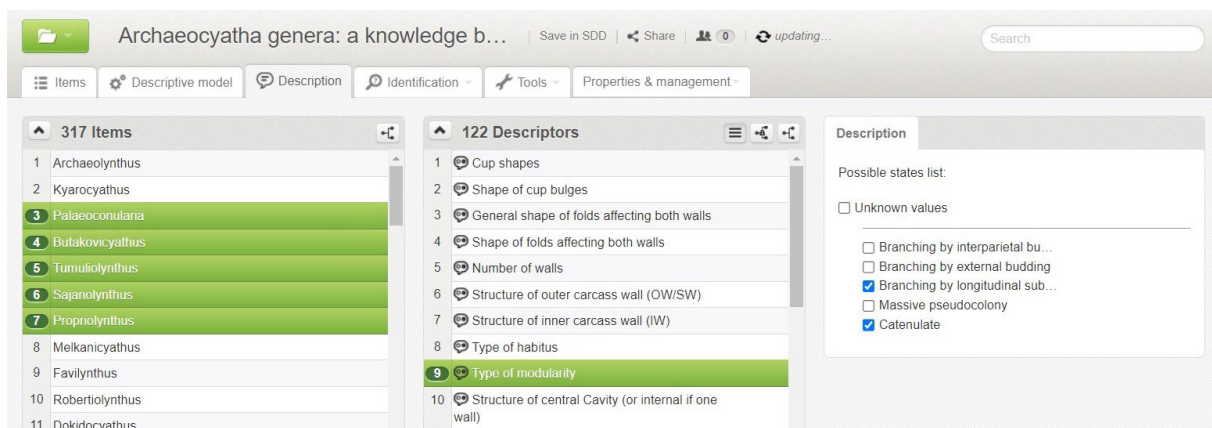


Fig. 7. Several items are selected to enter the same description simultaneously.

The ‘Identification’ tab

The ‘Identification’ tab provides access to the generation of simple access keys and the interactive key (Fig. 8) (Gérard & Vignes Lebbe 2010; Hagedorn *et al.* 2010).

The creation of simple access keys, via the Ikey+* web service (Burguiere *et al.* 2013), is based on an algorithm that optimises the topology of the key based on global criteria. By default, the first criterion minimises the average number of questions (Lebbe & Vignes 1991; Darmoni *et al.* 1993) and is then mitigated by the relative weight of each descriptor chosen by the content creator.

Interactive identification in Xper3, via the Mkey+* web service (Saucède *et al.* 2021), is carried out by successive elimination. During the description of the object to be identified, the system does not add compatible items but eliminates incompatible ones. In other words, at the beginning of the process, all items are considered as potential outcomes of the identification, and then, as descriptor states are selected, the incompatible taxa are definitively eliminated. The procedure is thus convergent: at a step $t+1$, the list of compatible taxa is always a subset of the list at the previous step t . At each step, it is possible to view the complete description of each item and to know why one was eliminated. During the identification process, the PD() function, for discriminant power, is used to sort the descriptor list according to the remaining taxa (Fig. 8).

The screenshot displays the Xper3 interactive key interface. On the left, there are two main sections for descriptor selection:

- Presence of vertical intervallar structures:** Includes a diagram of a cell wall with septa and taeniae. Description: "Vertical intervallar structures: Longitudinal and often radial structures inside the intervallum (septa, taeniae...)." Definition: "Intervallum: Space enclosed between the walls."
- Presence of carcass bumps: outer wall or single wall (OW/SW):** Includes four diagrams (1a, 1b, 1c, 2) showing different wall structures. Description: "Bumps: Outer wall structure in shape of a subspherical/hemispherical dome. Similar: Tumuli, pustule." Definition: "Carcass: Part of wall on intervallum side." Further definitions: "Outer wall: External part of the skeleton, composed by a porous carcass, simple or coated by a additional sheath." "Single wall: External part of the skeleton for one-walled cup, composed by a porous carcass, simple or coated by a additional sheath."

Below these are two radio button options:

- Present (34):** Includes a diagram of a pustule. Description: "Pustula (plur. pustulae): Wall structure covering an individual pore and pierced by a single central orifice. NB: Pustulae resemble simple tumuli." Definition: "Tumulus (plur. tumuli): Outer wall structure in shape of a subspherical dome covering an individual and downwardly projected wall opening."
- None (279):** Includes a diagram of a wall structure. Description: "No subspherical/hemispherical structure."

On the right side, a list of 312 remaining taxa is shown, sorted by discriminant power. The list includes: Archaeolythus, Kyarocyathus, Palaeoconularia, Butakovicyathus, Tumuliolythus, Sajanolynthus, Propriolythus, Melkanciyathus, Favilythus, Robertiolythus, Dokidocyathus, Dokidocyathella, Incurvocyathus, Cordobicyathus, Kidrjasocyathus, Kaltatocyathus, Papillocyathus, Subtilocyathus, Batschycicyathus, Zhuravlevaocyathus, Kymbecyathus, and Ajacicyathus.

Fig. 8. Xper3 interactive key (Mkey+ web service). On the left, descriptors are sorted in descending discriminant power for the remaining taxa (on the right).

An empty attribute is treated as if nothing is possible, and an attribute coded as unknown is treated as if all states are possible. In practice, when state A is selected, Xper3 will consider compatible those elements that have state A only, state A and one or more other states (polymorphism), or have ‘unknown value’ selected. All the other items will be eliminated, including the undescribed items for that descriptor, hence the importance of completing the knowledge base before using it as an identification tool.

Discussion

Over the 10 years of Xper3’s existence, the number of user accounts and knowledge bases has grown steadily (Figs 9–10). The review published three years after Xper3 was made available, already showed a variety of uses with a large geographical and taxonomic scope (Pinel *et al.* 2017). Xper3 continues to be used for different taxonomic groups, e.g., Diptera (Santos Neves *et al.* 2024), fossil Rodentia (Linchamps *et al.* 2023), Nematoda (Palomares-Rius *et al.* 2022), Echinoidea (Saucède *et al.* 2021), Orchidaceae (Pignal *et al.* 2023), Hymenophyllaceae (Saïd *et al.* 2017). The associated context also varies from taxonomic revision, citizen science (e.g., <http://spipoll.identificationkey.fr>) or knowledge sharing (e.g., Kerner 2022) to pedagogical use (e.g., <https://webiodiv.dsi.upmc.fr/basexper.php>) and expertise. A comprehensive list of publications using Xper3 can be accessed via the ‘Bibliography’ link on the Xper3 website (<https://xper3.fr/en/bibliography/>).

The user-friendly interface and the ability to illustrate and add supplementary text to specimens, descriptors and conditions are highly valued. A group of naturalists can easily collaborate on a regional fauna or flora, or on the revision of a taxonomic group. A teacher can have students work in groups on the same knowledge base. A specialist can complete their knowledge base step by step and then publish their key online with a simple click.

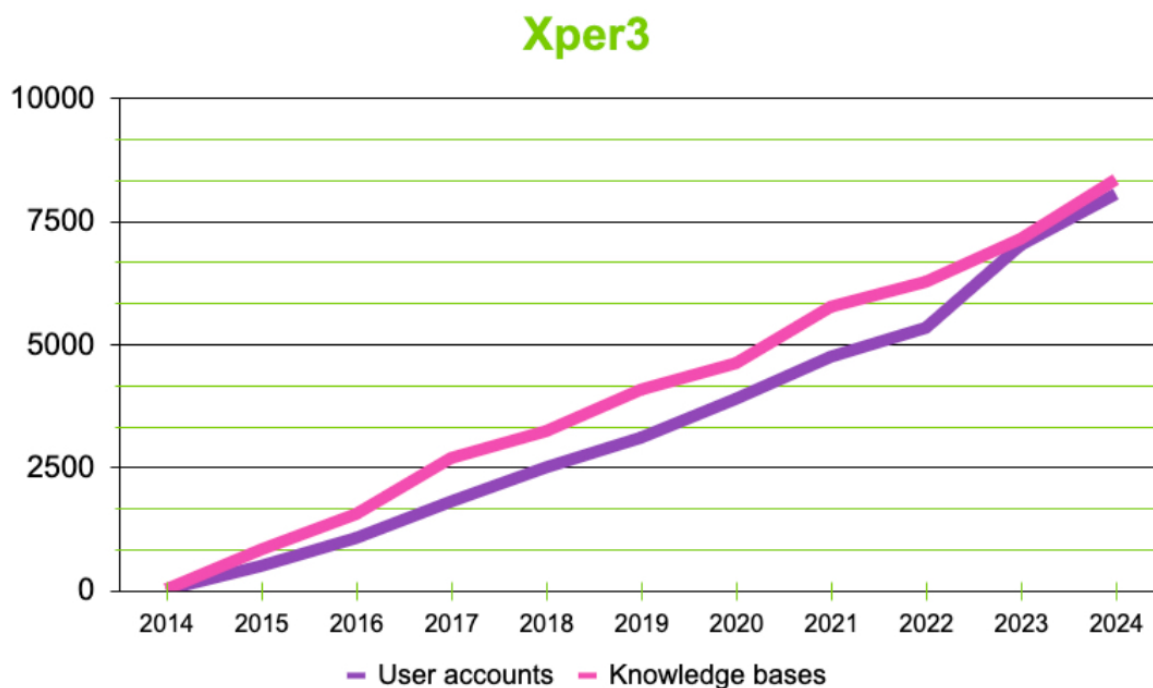


Fig. 9. Growth in the number of Xper3 content creators and databases. The majority of content creators are from academic institutions.

Special features and limits

The ability to include calculated descriptors is an original feature of the software. However, the use of calculated descriptors is still rare (Kerner & Vignes Lebbe 2016, 2019). Calculated descriptors can be used to reformulate morphological descriptors (adapted for identification purposes) into homologous characters (adapted for phylogeny). Similarly, calculated descriptors can be used to reconstruct literature terminology from a set of more restricted descriptors. Many scientific terms used in taxonomy are a synthesis of different concepts. In botany, a spike is a “racemose inflorescence with alternate and sessile flowers along a common unbranched axis” (Beentje 2010). Thus ‘spike’ contains multiple information (racemose, alternate, sessile, common unbranched axis), each of which can be a state of another given descriptor. It is then possible to reconstruct the notion of a spike using calculated descriptors. Similarly, calculated descriptors can be used to generate descriptors adapted to different contexts and content users (alternative characters in DELTA are limited by the fact that they require the same distributions of their states (Dallwitz *et al.* 2018)).

In terms of knowledge representation, the formal model of Xper3 is effective but limited. Polymorphism is represented as a set of possible values but without the nuances of an associated probability or typicality distribution. Such nuances would be useful for identification, as shown by Darmoni *et al.* (1993) in the context of clinical algorithms for medical applications. In taxonomy, however, such probabilities are rarely explicitly available in descriptions, but relative frequencies (as denoted by ‘mostly’, ‘rarely’, ‘usually’, etc.) are more common. Finally, the ability to create hierarchies of items or groups of descriptors is operational in data processing, but its full use in identification and content analysis is still under development.

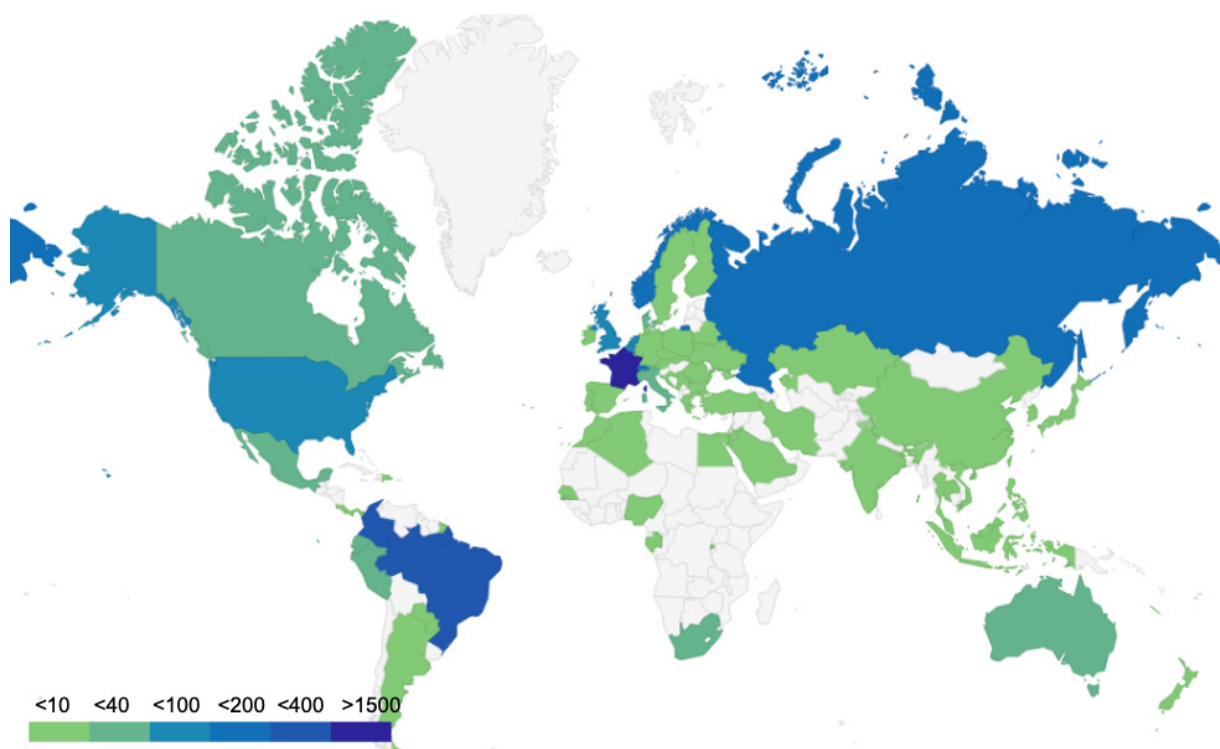


Fig. 10. The geographical origin of the content creators covers more than 70 countries. This data is derived from the email addresses (.fr, .cz, .br...), but covers only 44% of the content creators (@gmail, or .com email addresses are excluded). Content is available in more than 14 languages, including Portuguese, German, Spanish or Vietnamese. The colour reflects the amount of content.

Xper3 knowledge bases and the FAIR principles

Today, the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson *et al.* 2016) are widely promoted. They support and are central to open science. They cover the notions of dissemination, access, standards and data permanence. Xper3 participates in the FAIR approach in several ways (Kerner & Levy 2023).

Findable and accessible

Each creator of an Xper3 knowledge base can decide on the modalities of the dissemination of their work. Sharing content to open up data editing to others is the collaborative principle of Xper3, but it requires each user to have an account in order to log in. However, the interactive keys can be made available to the general public by using the ‘Publish identification’ option in the main menu. A public URL will be generated to access the interactive key without the need for the key user to log in.

Finding the URL is the limitation to using the key. This is why the Xplorer portal (<http://explorer.xper3.fr/>) was introduced. Referencing in Xplorer is currently possible if the creator actively chooses to publish the metadata of the base (and possibly gives access to the interactive identification mentioned above) thus improving its discoverability. It is possible to specify the geographical region or the taxonomic coverage of the identification key, as well as keywords for the general search function. The metadata must be as complete as possible to ensure optimal searching in Xplorer.

In today’s Open Science, there is an increasing number of repositories (Mendeley data, Zenodo data...). These platforms can be used both for archiving and for obtaining a DOI. In France, in the field of environment and biodiversity, the CNRS and the MNHN provide a repository called inDoRES (<https://www.indores.fr/>), which includes a dataverse dedicated to Xper (<https://data.indores.fr/dataverse/xper>). It is recommended that users also make their knowledge base publicly available by publishing a data paper such as the following: Saucède *et al.* (2021); Klopstein *et al.* (2022); Abdellahoum *et al.* (2022); Palomares-Rius *et al.* (2022); Gautam & Sidhu (2023); Linchamps *et al.* (2023); Santos Neves *et al.* (2024).

Interoperable and reusable

To make Xper3 contents interoperable with other identification software, it uses SDD (Structure of Descriptive Data Subgroup 2005), an exchange format supported by TDWG. The SDD format also allows knowledge bases to be imported from Xper2 to Xper3, ensuring continuity between the versions. It is important to note that calculated descriptors do not exist in SDD format and are therefore exported from Xper3 as categorical descriptors, losing the logical rules that generate the calculated values.

The common CSV format is also usable to import and export data, e.g., to import a list of taxa from a taxonomic backbone, to export a descriptive model (standard terminology), to export a taxa-characters matrix, etc. To further content reuse and sharing, preliminary work has been carried out on exporting data in OWL or RDF (Pellen *et al.* 2018; Porcher 2022), although this is not yet fully satisfactory. It is also possible to export Xper3 data in the Nexus format (Maddison *et al.* 1997).

As explained previously, textual details about items, descriptors or states, can include HTML links which allow elements of an Xper3 knowledge base to be mapped to external ontologies and web pages.

The Xper3 platform is free to use and well-documented for the users. On the one hand, the editing interface is regularly improved according to user feedback, and so it is not open source, but licensed under the Creative Commons 3.0 France licence (Attribution, Non-Commercial and ShareAlike).

Indeed, efficient open-source code requires important and up-to-date documentation for developers. On the other hand, the identification API is more stable, open source and well-documented for developers (the code is available on git: <https://gitlab.com/mnhn-lis/public/mkey-plus>, <https://gitlab.com/mnhn-lis/public/mkey-static-client>, <https://gitlab.com/mnhn-lis/public/ikey-plus>, <https://gitlab.com/mnhn-lis/public/ikey-static-client>).

Conclusion

In the era of artificial intelligence (Goëau *et al.* 2021; Koch 2023), big data and integrative taxonomy, access to structured, standardised and validated data are more important than ever. Xper3, in continuity with Xper2 and Xper, was created to assist taxonomists in their tasks, and their feedback has shown the value of such software. This current online version responds to the need for collaborative working.

Xper3 is part of several citizen science projects (e.g., <http://spipoll.identificationkey.fr>), taxonomic revisions (e.g., Santos Neves *et al.* 2024), university courses (e.g., <https://webiodiv.sorbonne-universite.fr/>) and the enhancement of natural history collections (Hays & Kerner 2020; Gourraud *et al.* 2021). It is also used in initiatives to raise awareness of biodiversity among young audiences.

Today, Xper3 collaborates with the French national network of natural history collections (e-ReColNat, <https://www.recolnat.org/en/>). With other tools such as Annotate-on (Pignal *et al.* 2024), Xper3 will contribute to the preparation of annotated image datasets for AI training on natural history collections.

Future developments planned or underway, in addition to technological developments, include the management of multilingual databases, support in linking and importing data from ontologies or taxonomic backbones, the introduction of hybrid identification with artificial intelligence, and the export of more sophisticated taxonomic files.

Acknowledgements

Xper3 is developed by a biodiversity informatics team hosted by ISYEB, CR2P and BBEES and supported by Sorbonne University, MNHN and CNRS.

We are grateful to the Steering Committee, whose role is to help make decisions and ensure the sustainability of the software.

We would like to thank the other permanent and temporary staff who have contributed to the development of the Xper3 software and without whom this collaborative platform would not exist. In alphabetical order: Antoine Bergamaschi, Thomas Burguière, Florian Causse, Roseline Dupuyau, Nicolas Lebbe, Frédéric Lison, and Vincent Rolland. We would also like to thank all the French and foreign colleagues with whom we collaborate. We dedicate this work to all the users whether they are experts, students or amateurs. We are grateful for their encouragement and feedback.

References

- Abdellahoum Z., Nebbak A., Lafri I., Kaced A., Bouhenna M.M., Bachari K., Boumegoura A., Agred R., Boudchicha R.H., Smadi M.A., Maurin M. & Bitam I. 2022. Identification of Algerian field-caught mosquito vectors by MALDI-TOF MS. *Veterinary Parasitology: Regional Studies and Reports* 31: 100735. <https://doi.org/10.1016/j.vprsr.2022.100735>
- Agathokleous E., Saitanis C.J., Fang C. & Yu Z. 2023. Use of ChatGPT: What does it mean for biology and environmental science? *Science of The Total Environment* 888: 164154. <https://doi.org/10.1016/j.scitotenv.2023.164154>

- Aide T.M., Corrada-Bravo C., Campos-Cerqueira M., Milan C., Vega G. & Alvarez R. 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1: e103. <https://doi.org/10.7717/peerj.103>
- Bartlett P., Eberhardt U. & Beker H.J. 2022. Demystifying *Hebeloma*: introducing hebeloma.org and its database. *IMA Fungus* 13 (1): 18. <https://doi.org/10.1186/s43008-022-00105-2>
- Beentje H. 2010. *The Kew Plant Glossary: An Illustrated Dictionary of Plant Terms*. Royal Botanic Gardens Kew, UK.
- Bodin S.C., Scheel-Ybert R., Beauchêne J., Molino J.-F. & Bremond L. 2019. CharKey: an electronic identification key for wood charcoals of French Guiana. *IAWA Journal* 40 (1): 75-S20. <https://doi.org/10.1163/22941932-40190227>
- Burguiere T., Causse F., Ung V. & Vignes-Lebbe R. 2013. IKey+: a new single-access key generation web service. *Systematic Biology* 62 (1): 157–161. <https://doi.org/10.1093/sysbio/sys069>
- Dallwitz M.J. 1980. A general system for coding taxonomic descriptions. *TAXON* 29 (1): 41–46. <https://doi.org/10.2307/1219595>
- Dallwitz M.J., Paine T.A. & Zurcher E.J. 2018. Principles of interactive keys. Available from <https://www.delta-intkey.com/www/interactivekeys.htm> [accessed on 25 May 2023].
- Darmoni S.J., Lebbe J., Vignes R. & Poynard T. 1993. MAKEY: génération automatique d'un algorithme clinique pour aider le diagnostic sérologique de l'hépatite virale aiguë. *Informatique et Santé* 15: 49–55.
- Del Rio C., Stull G.W. & De Franceschi D. 2020. Survey of the fruits and endocarps of Icacinaceae (Lamiids, Icacinales). *European Journal of Taxonomy* 645: 1–130. <https://doi.org/10.5852/ejt.2020.645>
- Dubuisson J.-Y., Bouquin S. & Vignes Lebbe R. 2018. Computer assisted identification: Xper and orchids of France. *L'Orchidophile* 219: 311–318.
- Gautam N. & Sidhu M.C. 2023. An alternative key for identifying three prostrate species of the genus *Euphorbia*. *Botanica* 29 (1): 21–27. <https://doi.org/10.35513/Botlit.2023.1.3>
- Gérard D. & Vignes Lebbe R. 2010. MyKey: a server-side software to create customized decision trees. *In: Nimis P.L. & Vignes Lebbe R. (eds) Tools for Identifying Biodiversity: Progress and Problems: 107–112*. Trieste.
- Goëau H., Bonnet P. & Joly A. 2021. AI-based identification of plant photographs from herbarium specimens. *Biodiversity Information Science and Standards* 5: e73751. <https://doi.org/10.3897/biss.5.73751>
- Goharimanesh M., Stöhr S., Mirshamsi O., Ghassemzadeh F. & Adriaens D. 2021. Interactive identification key to all brittle star families (Echinodermata; Ophiuroidea) leads to revised morphological descriptions. *European Journal of Taxonomy* 766: 1–63. <https://doi.org/10.5852/ejt.2021.766.1483>
- Gourraud A., Vignes Lebbe R., Kerner A. & Pignal M. 2021. Architectural pattern: study of orchid architecture using tools to take quick measurements of virtual specimens. *Biodiversity Information Science and Standards* 5: e75752. <https://doi.org/10.3897/biss.5.75752>
- Hagedorn G. 2007. *Structuring descriptive data of organisms — Requirement analysis and information models (Strukturierung organischer Beschreibungsdaten – Anforderungsanalyse und Informationsmodelle)*. Universität Bayreuth.
- Hagedorn G., Rambold G. & Martellos S. 2010. Types of identification keys. *In: Nimis P.L. & Vignes Lebbe R. (eds) Tools for Identifying Biodiversity: Progress and Problems: 59–64*.

- Hays L.-M. & Kerner A. 2020. Virtual systematics with annotate: the Mexican archaeocyaths example. *Biodiversity Information Science and Standards* 4: e59106. <https://doi.org/10.3897/biss.4.59106>
- Hebert P.D.N. & Gregory T.R. 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54 (5): 852–859. <https://doi.org/10.1080/10635150500354886>
- Hilgenhof R., Gagnon E., Knapp S., Aubriot X., Tepe E.J., Bohs L., Giacomini L.L., Gouvêa Y.F., Martine C.T., Orejuela A., Orozco C.I., Peralta I.E. & Särkinen T. 2023. Morphological trait evolution in *Solanum* (Solanaceae): evolutionary lability of key taxonomic characters. *Taxon* 72 (4): 811–847. <https://doi.org/10.1002/tax.12990>
- Jouveau S., Delaunay M., Vignes-Lebbe R. & Nattier R. 2018. A multi-access identification key based on colour patterns in ladybirds (Coleoptera, Coccinellidae). *ZooKeys* 758: 55–73. <https://doi.org/10.3897/zookeys.758.22171>
- Kerner A. 2022. De rat de bibliothèque à pro de la souris. In: *Paléontologie d'aujourd'hui*: 122–123. EDP Sciences.
- Kerner A. & Levy C. 2023. E-taxonomy, Open Science, and FAIR Principles: a comparison of how taxon descriptions are computerized. *Biodiversity Information Science and Standards* 7: e110459. <https://doi.org/10.3897/biss.7.110459>
- Kerner A. & Vignes Lebbe R. 2016. Knowledge base with XPER3 in palaeontology: calculated descriptors offer new functionalities. *TDWG 2016* Santa Clara de San Carlos, Costa Rica.
- Kerner A. & Vignes Lebbe R. 2019. Multi-context knowledge base using calculated descriptors from Xper3: the archaeocyaths knowledge base example. *Biodiversity Information Science and Standards* 3: e37083. <https://doi.org/10.3897/biss.3.37083>
- Kerner A., Bouquin S., Portier R. & Vignes Lebbe R. 2021. The 8 years of existence of Xper3: state of the art and future developments of the platform. *Biodiversity Information Science and Standards* 5: e74250. <https://doi.org/10.3897/biss.5.74250>
- Klopfstein S., Broad G.R., Urfer K., Vårdal H. & Haraldseide H. 2022. An interactive key to the European genera of Campopleginae (Hymenoptera, Ichneumonidae) and 20 new species for Sweden. *Entomologisk Tidskrift* 143: 121–156.
- Koch W. 2023. Capturing and using taxonomic knowledge for species identification. *Biodiversity Information Science and Standards* 7: e110957. <https://doi.org/10.3897/biss.7.110957>
- Lebbe J. 1991. *Représentation des concepts en biologie et en médecine: introduction à l'analyse des connaissances et à l'identification assistée par ordinateur*. Université Pierre et Marie Curie.
- Lebbe J. & Vignes R. 1991. Génération de graphes d'identification à partir de descriptions de concepts. In: *Induction symbolique et numérique à partir de données*. Cepadues: 193–239. Cepadues, Toulouse, France.
- Lebbe J., Vignes R. & Dedet J.P. 1989. Computer-aided identification of insect vectors. *Parasitology Today* 5 (9): 301–304. [https://doi.org/10.1016/0169-4758\(89\)90025-2](https://doi.org/10.1016/0169-4758(89)90025-2)
- Linchamps P., Avery D.M., Cornette R., Denys C., Matthews T. & Stoetzel E. 2023. Quaternary rodents of South Africa: A companion guide for cranio-dental identification. *PLoS ONE* 18 (11): e0289812. <https://doi.org/10.1371/journal.pone.0289812>
- Lombard N., Le Roux M.M. & van Wyk B.-E. 2021. Electronic identification keys for species with cryptic morphological characters: a feasibility study using some Thesium species. *PhytoKeys* 172: 97–119. <https://doi.org/10.3897/phytokeys.172.53484>

- Maddison D.R., Swofford D.L. & Maddison W.P. 1997. Nexus: an extensible file format for systematic information. *Systematic Biology* 46 (4): 590–621. <https://doi.org/10.1093/sysbio/46.4.590>
- Nimis P.L. & Vignes-Lebbe R. 2010. *Tools for Identifying Biodiversity: Progress and Problems*. Università di Trieste, Trieste, Italy.
- Norton G.A., Patterson D.J. & Schneider M. 2000. LucID: a multimedia educational tool for identification and diagnostics. *International Journal of Innovation in Science and Mathematics Education* 4 (1): 15–22.
- Open Science 2023. Trends and best practices in open science. Available from <https://openscience.eu/> [accessed on 15 May 2024].
- Palomares-Rius J.E., León-Ropero G., Clavero-Camacho I., Archidona-Yuste A., Cantalapiedra-Navarrete C. & Castillo P. 2022. A new interactive web-based polytomous key for species identification of pin nematodes of the genus *Paratylenchus* Micoletzky, 1922 (Nematoda: Paratylenchinae) with the use of ribosomal and mitochondrial genes. *Journal of Zoological Systematics and Evolutionary Research* 2022: 6216266. <https://doi.org/10.1155/2022/6216266>
- Pankhurst R.J. 1970. Key generation by computer. *Nature* 227 (5264): 1269–1270. <https://doi.org/10.1038/2271269a0>
- Pellen F., Bouquin S., Mougnot I. & Vignes-Lebbe R. 2018. Building an OWL ontology with Xper3. *Biodiversity Information Science and Standards* 2: e25614. <https://doi.org/10.3897/biss.2.25614>
- Picek L., Šulc M., Patel Y. & Matas J. 2022. Plant recognition by AI: deep neural nets, transformers, and kNN in deep embeddings. *Frontiers in Plant Science* 13: 787527. <https://doi.org/10.3389/fpls.2022.787527>
- Pignal M., Laudereau C. & Laudereau P.-L. 2023. Liste annotée des Goodyerinae Klotzsch (Orchidaceae) de Nouvelle-Calédonie avec la description de deux nouvelles espèces. *Adansonia* 45 (19): 307–325. <https://doi.org/10.5252/adansonia2023v45a19>
- Pignal M., Bertin G., Chupin L., Perez Pimparé E., Klasnja S., Vignes Lebbe R. & Dusoulier F. 2024. Récolnat Annotate-On: a tool to improve your experience with virtual collections. *Adansonia* 46 (13): 133–148. <https://doi.org/10.5252/adansonia2024v46a13>
- Pinel A., Bouquin S., Bourdon E., Kerner A. & Vignes-Lebbe R. 2017. Three years of Xper3 assessment: towards sharing semantic taxonomic content of identification keys. *Proceedings of TDWG* 1: e20382. <https://doi.org/10.3897/tdwgproceedings.1.20382>
- Porcher M.-L. 2022. *Terminologie associée au processus de diagnose des espèces : un service de formalisation pour l'interopérabilité sémantique*. Master 2 thesis, Sorbonne Université-MNHN.
- Saïd A.H., Hennequin S., Rouhan G. & Dubuisson J.-Y. 2017. Disentangling the diversity and taxonomy of Hymenophyllaceae (Hymenophyllales, Polypodiidae) in the Comoros. *European Journal of Taxonomy* 313: 1–53. <https://doi.org/10.5852/ejt.2017.313>
- Santos Neves M.S.A., Motta M.A., Maciel-De-Freitas R., Da Silva Xavier A., Lourenço-De-Oliveira R. & Silva-Do-Nascimento T.F. 2024. Illustrated identification key to females of the genus *Sabethes* Robineau-Desvoidy recorded from Brazil (Diptera: Culicidae), in dichotomous and interactive formats, including an updated list of species and new records for the states. *Zootaxa* 5406 (2): 253–287. <https://doi.org/10.11646/zootaxa.5406.2.3>
- Saucède T., Eléaume M., Jossart Q., Moreau C., Downey R., Bax N., Sands C., Mercado B., Gallut C. & Vignes-Lebbe R. 2021. Taxonomy 2.0: computer-aided identification tools to assist Antarctic biologists in the field and in the laboratory. *Antarctic Science* 33 (1): 39–51.

Schneider C., Minor M.A. & D’Haese C.A. 2023. A new group of species of the genus *Megalothorax* (Collembola, Neelidae) with Gondwanan distribution, and introducing an open interactive identification key of *Megalothorax* species. *Zootaxa* 5228 (2): 101–121. <https://doi.org/10.11646/zootaxa.5228.2.1>

Structure of Descriptive Data Subgroup 2005. Structured Descriptive Data (SDD). Biodiversity Information Standards (TDWG). Available from <https://www.tdwg.org/standards/sdd/> [accessed on 15 May 2024].

Taberlet P., Coissac E., Pompanon F., Brochmann C. & Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21 (8): 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>

Vignes Lebbe R., Chesselet P. & Diep Thi M.H. 2016. Xper3: nouveaux outils pour le travail collaboratif, la formation et la transmission des connaissances sur les phénotypes végétaux. In: *Botanists of the Twenty First Century: Roles, Challenges and Opportunities: Based on the Proceedings of UNESCO International Conference, 22–25 September 2014, Paris, France*. UNESCO: 228–239. UNESCO, Paris, France.

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., Da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., ’T Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., Van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., Van Der Lei J., Van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J. & Mons B. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>

Manuscript received: 23 May 2024

Manuscript accepted: 12 December 2024

Published on: 18 April 2025

Topic editor: Tony Robillard

Section editor: Frank Zachos

Desk editor: Chris Le Coquet-Le Roux

Printed versions of all papers are deposited in the libraries of four of the institutes that are members of the *EJT* consortium: Muséum national d’Histoire naturelle, Paris, France; Meise Botanic Garden, Belgium; Royal Museum for Central Africa, Tervuren, Belgium; Royal Belgian Institute of Natural Sciences, Brussels, Belgium. The other members of the consortium are: Natural History Museum of Denmark, Copenhagen, Denmark; Naturalis Biodiversity Center, Leiden, the Netherlands; Museo Nacional de Ciencias Naturales-CSIC, Madrid, Spain; Leibniz Institute for the Analysis of Biodiversity Change, Bonn – Hamburg, Germany; National Museum of the Czech Republic, Prague, Czech Republic; The Steinhardt Museum of Natural History, Tel Aviv, Israël.