

# Research Between Natural and Cultural History Information: Benefits and IT-Requirements for Transdisciplinarity

KARL-HEINZ LAMPE

Zoologisches Forschungsmuseum Alexander Koenig (ZFMK) and  
Interdisciplinary Center of Complex Systems (IZKS)

KLAUS RIEDE

Zoologisches Forschungsmuseum Alexander Koenig (ZFMK)  
and

MARTIN DOERR

Institute of Computer Science, Foundation for Research and Technology Hellas (ICS Forth)

This article describes an approach to transdisciplinary information integration employing a core ontology. Information is modeled here with an ontology based on the CIDOC-Conceptual Reference Model (ISO 21127). When instantiated with some realistic examples taken from the field of biodiversity (collecting, determination, type creation, expedition, and observation events), the formal specification of semantic concepts makes scientific activities commonly understandable. Ontologies not only allow one to describe the results of scientific activities, such as a description of a biological species, but they can help to clarify the path by which the goal was reached. In particular, they provide a high-level uniform representation of transdisciplinary research activities and results. Ontologies as knowledge representation tools will therefore have strong impact on methodological questions and research strategies for different domains such as biology, archaeology, art history, and socio-economy. They can be regarded as semantic glue between and within scientific and scholarly domains as demonstrated in a series of examples. Following this approach, some requirements for research and development of integrated IT environments between memory institutions are derived.

Categories and Subject Descriptors: D.2.12 [**Software Engineering**]: Interoperability; H.1.1 [**Models and Principles**]: Systems and Information Theory—*Value of information*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; H.2.2 [**Database Management**]: Logical Design—*Schema and subschema*; H.2.8 [**Database Management**]: Database Applications—*Scientific databases*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*; H.3.7 [**Information Storage and Retrieval**]: Digital Libraries; J.3 [**Life and Medical Sciences**]: Biology and genetics; J.5 [**Arts and Humanities**]:

General Terms: Theory, Design, Standardization, Management

This research was partly supported by the German Ministry of Education and Research (BMBF) within the BIOLOG Biodiversity Informatics- and BIOLOG Biota Africa- programmes by funding the DORSA-, DIG- and BIOTA East Africa E15 projects.

Authors' addresses: Dr. Karl-Heinz Lampe and Dr. Klaus Riede, Biodiversity Informatics, Zoologisches Forschungsmuseum Alexander Koenig (ZFMK), Adenauerallee 160, 53113 Bonn, Germany; Dr. Martin Doerr, Institute of Computer Science, Foundation for Research and Technology Hellas (ICS FORTH), Vassilika Vouton, 71110 Heraklion/Crete, Greece.

Author's address: K.-H. Lampe (corresponding author); email: k.lampe.zfmk@uni-bonn-de.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2008 ACM 1556-4673/2008/06-ART4 \$5.00 DOI 10.1145/1367080.1367084 <http://doi.acm.org/10.1145/1367080.1367084>

Additional Key Words and Phrases: Interdisciplinary research, crossdisciplinary search, transdisciplinarity, natural history, cultural heritage, museums, biodiversity, ontologies, information integration, system requirements, semantic enrichment, georeferencing, ISO 21127

**ACM Reference Format:**

Lampe, K.-H., Riede, K., and Doerr, M. 2008. Research between natural and cultural history information: Benefits and IT requirements for transdisciplinarity. *ACM J. Comput. Cultur. Heritage* 1, 1, Article 4 (June 2008), 22 pages. DOI = 10.1145/1367080.1367084 <http://doi.acm.org/10.1145/1367080.1367084>

---

## 1. INTRODUCTION

### 1.1 About Knowledge Management

Information integration supports the creation of new knowledge. Scientific and scholarly disciplines are based on specific methodologies and terminologies. Specific information is usually presented in domain-specific terms. The concept behind these domain-specific terms is hard to understand for nonexperts. That might be one reason that transdisciplinary approaches are so rare.

According to Mittelstrass [2001] “transdisciplinarity is a science theoretical concept which describes particular forms of scientific co-operation and problem solving ...”. The geologists Ryan and Pitman [1998], for example, formed a hypothesis about the Black Sea flood referred to in the Bible. During a period of more than thirty years, they collected evidence and data showing that a great flood had indeed taken place about 5600 BC. Their research was based on various kinds of seafloor investigations, polar ice data, climate studies, coral fossils in the Caribbean, vertical movement of continents under load, behavior of a multitude of fresh and saltwater species, pollen fossil analysis, archaeology of neolithic settlements, history of domestic plants up to ancient mythologies. In the following years, underwater finds in the Black Sea have supported their theory. Transdisciplinary research inaugurates new questions based on new methodologies. This is just one example which illustrates the potential of a transdisciplinary approach based on information integration.

There is a need for high-level ontologies to support information integration. However, current activities with Semantic Web technology mostly deal with mapping or aligning fairly specific terminologies from different domains as demonstrated by Chen et al. [1997] and summarized in Kalfoglou and Schorlemmer [2003]. On the schema level, the predominant approach is to use flat metadata element sets such as the Dublin Core. The latter was developed to simplify and generalize library documentation, allowing for the retrieval of objects by common creation characteristics, subject, and references. In contrast to library information, which normally only refers to primary data, scientific information is directly based on primary data. For information integration, well-structured primary data have to be opened up in their full depth. It is extraordinarily difficult with only existing tools to retrieve information from other domains related in the characteristic way for these kinds of research problems. Therefore a core ontology is expected to be more adequate for scientific information integration. It should be more expressive than flat metadata records and yet be generic enough. In other words, it is necessary to find a platform for a common understanding of domain-specific concepts, independently from their terminology. Ontologies such as the CIDOC-Conceptual Reference Model (CRM) can be used as a common language to map scientific and scholarly concepts. As a result identities, communalities, similarities and differences of concepts can be identified, a prerequisite for an advanced information integration. Expected benefits of information integration are verification of scientific results and the development of new and broader hypotheses (which could not be developed in a single discipline).

According to experiences from ecology, Göttlicher and Bendix [2004] maintain that an interdisciplinary exchange of information can only take place with appropriate data structures. More generally,

we hold that formal ontology [Gruber 1993], in particular capturing the semantics of datastructures and relationships between information elements, is the key to making transdisciplinary questions and requirements explicit and thereby developing scientific methods for their solution. In informatics, ontology is a formal specification of a conceptualization. When instantiated with some realistic examples taken from a particular field such as biodiversity (collecting, determination, type creation, expedition and observation events.), the formal specification of semantic concepts make scientific activities commonly understandable. Ontologies not only describe the results of scientific activities, such as a description of a biological species, but they help to clarify the path by which the goal was reached. Consequently, an objective analysis of the respective scientific argumentation is possible, in particular, how the discipline boundaries are crossed. This opens the way for designing effective knowledge management tools which are demand driven and adequate for the specific discourse in different disciplines, not just based on naive assumptions like most of the current practices. The authors demonstrate the value of such research, using real-life examples of transdisciplinary research associated with biodiversity. The examples are analyzed in terms of a formal core ontology, the CIDOC CRM (ISO21127) [Doerr 2003] and used to describe some initial benefits and requirements for transdisciplinarity.

## 1.2 About Biodiversity and Transdisciplinarity

In view of the enormous ecological and economic problems of the world, there is an urgent need to understand biodiversity in the context of complex systems and to make knowledge directly available. Today we have the technology to realize the vision of a vertical information transfer within biological science, integrating the molecular, organism, and ecosystem level. For the organism level, several international working groups such as the Taxonomic Databases Working Group (TDWG),<sup>1</sup> Biological Collection Access Service for Europe (BIOCASE),<sup>2</sup> and the Global Biodiversity Information Facility (GBIF)<sup>3</sup> develop schemes, protocols, and other helpful tools to make taxonomic data more accessible. They will provide the backbone of organismic classification needed in molecular biology or ecology.

However, for today's scientist, it is not sufficient to be an expert who can best talk to his nearest neighbor. There is a need to share knowledge and expertise and make it understandable. As a result, both scientific and public demands become more and more focussed on transdisciplinarity instead of interdisciplinarity. Another challenge will therefore be a horizontal information transfer between biological and other domains such as libraries and archives. Buzz words like semantic web underline the need of linking biodiversity informatics, geo-informatics and finally the wide area of cultural informatics. This task is mainly directed to archives, libraries, and museums, which can be viewed according to Dempsey [2000] as Memory institutions.

## 2. WHAT KIND OF DATA DO NATURAL HISTORY MUSEUMS HAVE?

Natural history museums are archives of biodiversity. They house millions of specimens collected in space and time, providing firsthand information about geographical and historical presence of organisms, their morphological and other specific characters (e.g. anatomical and genetic ones.), and sometimes additional information about ecological environments or host-parasite relationships, etc.

The classification of these specimens/organisms is based on a taxonomic system that tries to reflect phylogenetic relationships between groups of organisms. It results in subdividing the living world into hierarchical biological units, such as plants, animals, and other kingdoms. Groups of species are classified within higher-ranked taxa, such as genera, families, orders and classes resulting in a Tree of

<sup>1</sup>[http://www.nhm.ac.uk/hosted\\_sites/tdwg/](http://www.nhm.ac.uk/hosted_sites/tdwg/)

<sup>2</sup><http://www.biocase.org/>

<sup>3</sup><http://www.gbif.org/>

Life (for an excellent Web representation see the Tree of Life Project<sup>4</sup>). The leaves of the tree are the comparatively well-defined units of species, subspecies, or other infrasubspecific taxa. Their description and scientific naming procedure is regulated by the International Codes of Nomenclature (for animals, for plants and fungi, for bacteria strains, for virus strains, for cultivated plants). The description of a new species, subspecies, or strain (for bacteria) requires that the new name refers to a single voucher specimen, the so called type specimen.

Therefore curators in a natural history museum are not only keepers of specimens, they are keepers of names. These scientific names are the only ones valid worldwide. They are internationally standardized. Species names are pivots in digital networks. They connect all kind of scientific (and trivial) information about the animated world ranging from the molecular level via the organism level up to the ecosystem level. Species are based on a concept and therefore not alive.<sup>5</sup> Specimens, however, are pivots in nature's networks. They are the real existing actors in life. A name (a picture, a sound, a gen-sequence, etc.) without a relation to the specimen is nothing more than a statement without a proof. The type specimen on which the species description is based is the reference for the species in the real world.

### 3. WHAT KIND OF DATA DO CULTURAL HISTORY MUSEUMS HAVE?

Cultural history museums and archives cover a wide list of subject areas, such as the history of arts, science, social and political history, archaeology and anthropology. They keep relevant objects to study past and current cultural history and to preserve knowledge about it. They may have multiple reasons to keep objects

- (1) The object itself may be of remarkable artistic value (a masterpiece) or may contain expressions of particular knowledge, such as paintings or manuscripts.
- (2) The object may be a historical heirloom, an object that had been in the hands of an important person of the past without having any other particular value, or a silent witness of an event, such as a watch hit by a bullet and thereby saving its owner's life.
- (3) The object may contain information about a particular situation or part of history, such as a letter.
- (4) The object may be a characteristic representative of its category such as a feather crown from indigenous people, or a ceramic vessel like a Red-figured 5th century B.C. Attic vase. It may illustrate customs, fashions, technology, knowledge, or any other aspect of human life.
- (5) The object may also be a rare or unique representative of its category, such as the roughly 400,000-year old complete wooden spears found in Schoeningen, Niedersachsen, Germany in 1995.

Cultural history museums assess and document the relevance of their objects by describing them in their historical context. In cases (1), (2), and (3), the primary relevance has more to do with the history and social environment of individuals. In cases (4) and (5), the relevance has more generally to do with the living conditions and achievements of a particular social group in a particular period. As some museum curators poetically say, the "objects tell stories." In this sense, cultural history museums preserve both objects as physical evidence of the past and associated stories. They might be seen as keepers of human artifacts and keepers of stories, but categorical information (see case (4)) is also present similar to natural history.

<sup>4</sup><http://www.tolweb.org/tree/phylogeny.html>

<sup>5</sup>Inbedded in a speciation theory, species can be reificated or 'thingified'. This does not make them, however, alive. Species remain a mental product.

#### 4. KNOWLEDGE ASSOCIATED WITH A COLLECTION OBJECT

The knowledge associated with a well-documented collection object is represented in this article as a model [Guarino 1998] instantiating a formal ontology, the CIDOC Conceptual Reference Model or CIDOC CRM [Crofts et al. 2003]. Originally ontology was a branch of philosophy concerned with the study of being, of reality, in its most fundamental and comprehensive forms. In computer science, the term is used for a formal specification of a conceptualization. The CIDOC CRM<sup>6</sup> is a core ontology and ISO standard (ISO 21127) designed for the semantic integration of information from museums, libraries, and archives. It has been developed by the CIDOC CRM Special Interest Group (SIG), an international multidisciplinary team of experts. The CIDOC CRM concentrates on the definition of relationships rather than terminology in order to support mediation, transformation, and integration between heterogeneous database schemas and metadata structures. It is a product of reengineering the dominant common meanings from the most characteristic schema elements in use in these institutions. This has led to a compact model consisting of a class hierarchy of 81 named classes (so-called entities, E1, E2, etc.), which are interlinked by 132 named properties (P1, P2, etc.). These classes and properties turned out to be widely independent of the specific concepts of the original domains. It is not prescriptive, but provides a controlled language to describe common high-level semantics that allow the integration of information at the schema level.

The driving principle is the explicit modeling of events: people, objects, places, and time-spans are interrelated by common events. This principle is based on empirical evidence from numerous data formats analyzed by the developers of the CRM. It is actually the key to describing meaningful relations between facts and knowledge elaborated in different disciplines as demonstrated in this article. It allows for the representation of metadata, such as creation, finding, use, and publication, as well as content summarization and the creation of integrated knowledge bases. From an implementation point of view, models instantiating the CIDOC CRM can be encoded in many forms. The W3C currently recommends RDF or OWL [RDF 2004], but the essence is the semantics of connecting information elements and not the encoding. Therefore equivalent encodings in XML object-oriented databases or even RDBMS are possible.

##### 4.1 Taxonomic Information

People usually see what they want to see. Showing a pinned insect such as in the center of Figure 1 and asking: “What is it?” usually provokes the answer: “A bee!” In this case, the pin and the labels are ignored. A pinned insect, however, is obviously a man-made object. It consists of a biological object, the bee, and an information carrier, a pin carrying labels as documents. Different views are creating different meanings for the same thing. For ecologists, the white label primarily documents a collecting event, for economists, the same label documents an acquisition event, and, for lawyers, it might document a transfer of custody (from nature to a collector). Thus the white label documents a multiple instantiation of three kinds of events defined in the CIDOC CRM, namely E7 (collecting activity), E8 (acquisition event), and E10 (transfer of custody) by one actual activity. But it also shows the location of a particular person, the collector, at a certain time. Similarly, the red label documents a type creation of a new species and a determination or type assignment. The type creation event has created a document, a species description, and a type that is a new class in a philosophical ontological sense with the appellation *Colletes alini* by Kuhlmann, 2000. The bee specimen supported the type creation event in the taxonomic role of a holotype.<sup>7</sup> Type creation events in other disciplines such as in the field of

<sup>6</sup>The CIDOC CRM has been developed to make preexisting different schemas commonly understandable. It was not developed primarily to create new schemas (of course, one can use it for that purpose as a best practice guide).

<sup>7</sup>*Holotype* is the single specimen designated or otherwise fixed as the name-bearing type of a nominal species or subspecies when the nominal taxon is established. (See International Code on Zoological Nomenclature, 4th Edition, 1999).

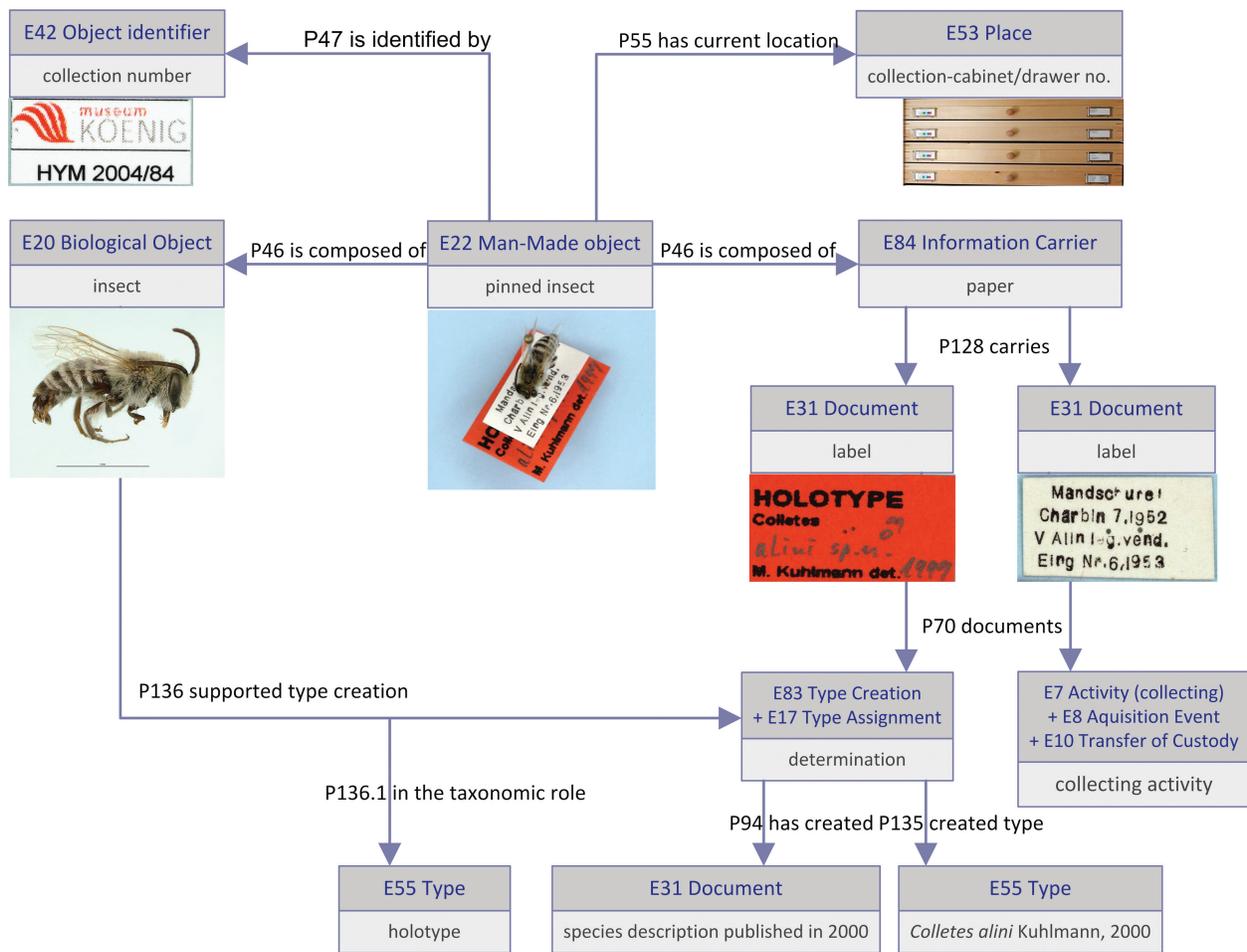


Fig. 1. Knowledge associated with a well-documented collection object in terms of the CIDOC CRM (E1, 2, etc. = Entity number; P1, 2, etc. = Property number).

archaeology are often based on the same or similar methodological approach; the terminology used, however, is different [Lampe and Krause, to appear).

Of course, this knowledge-representation model can easily be extended to include for example, the object identifier and place of storage within a collection (see Figure 1) or refined to a more detailed description such as a collecting event or even an expedition (see Figure 2).

## 4.2 Geographic Information

Natural history museums document Earth's biodiversity with billions of specimens often collected centuries ago [Duckworth et al. 1993]. Most specimens bear a label, indicating date, collector, and locality. Early naturalists often collected thousands of specimens, which were distributed or sold to institutions all over the world [Raby 1997]. Likewise, cultural artifacts were produced or collected somewhere (part of provenance research). Until the 19th and even into the 20th century, naturalists worked closely with indigenous people and often collected both artifacts and specimens, for example, Spix and Martius in

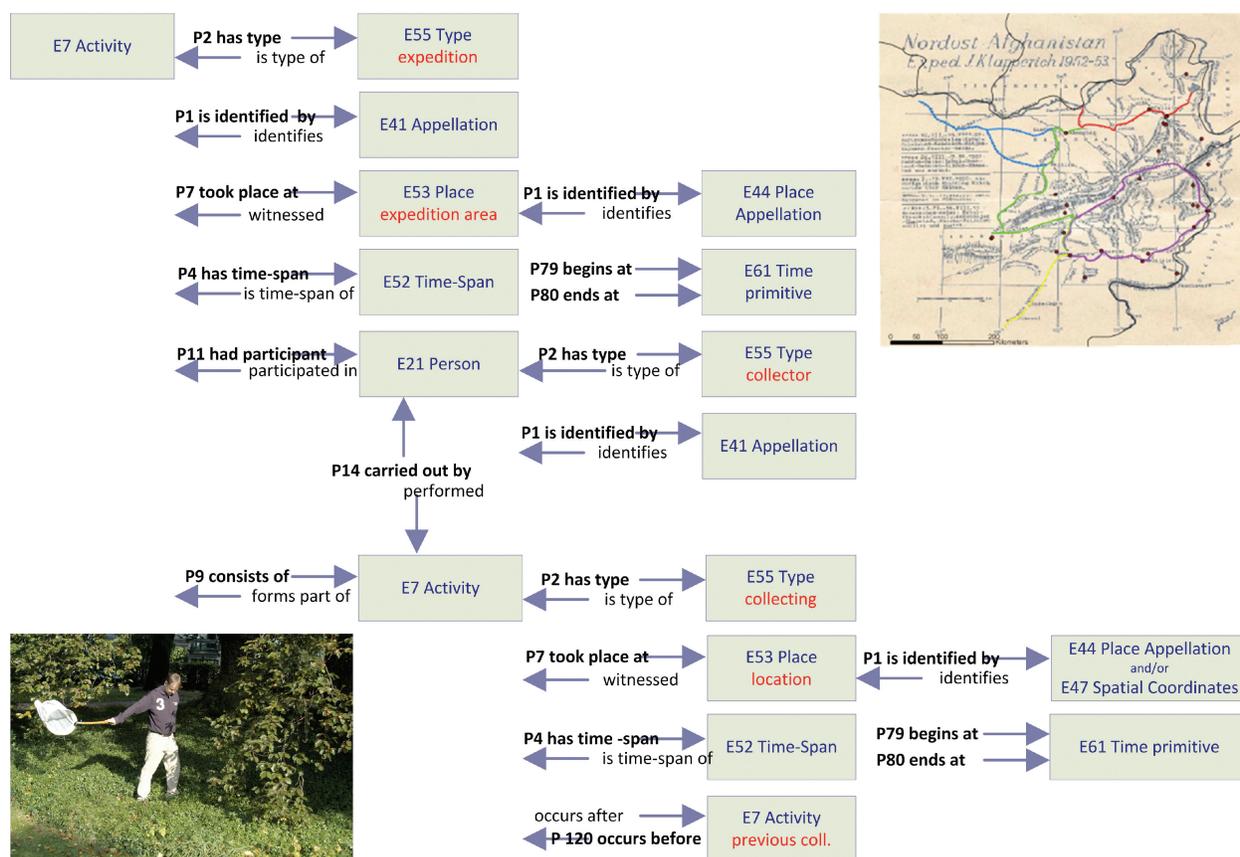


Fig. 2. Expedition and collecting events in terms of CIDOC CRM. Upper right: map of Klapperich’s Afghanistan expeditions (1952-1952). Lower left: Entomological collecting activity.

Brazil from 1817–1820 [Mauthe 1994]. However, geographic information on object labels often consists of names (toponyms) only, without geographic coordinates. The latter are necessary for mapping specimens with geographic information systems and have to be added by a time-consuming georeferencing process [Murphey et al. 2004] consisting of adding latitude and longitude to geographic place names often with the help of gazetteers [Guralnick et al. 2006]. Gazetteers are geographic authority files, such as the Getty Thesaurus of Geographic Names Online (TGN)<sup>8</sup> or the Alexandria Digital Library-Gazetteer Server (ADL),<sup>9</sup> and they deal with geographic names in a geophysical and geopolitical context. Once specimen or object data have been georeferenced, gathering sites (localities) can be visualized by linking to remote map servers such as the Canadian Biodiversity Information Facility - (CBIF)<sup>10</sup> map server or Google Earth.<sup>11</sup>

Facing the huge number of specimens in natural history collections, the efficiency of georeferencing has to be increased by orders of magnitude. Conceptual reference modelling of existing specimen

<sup>8</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

<sup>9</sup><http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp>

<sup>10</sup>[http://www.cbif.gc.ca/mc/index\\_e.php?p\\_cancarte=t](http://www.cbif.gc.ca/mc/index_e.php?p_cancarte=t)

<sup>11</sup><http://earth.google.com/>

databases immediately shows a possible solution: instead of georeferencing each specimen, it is far more efficient to map entire itineraries of collectors (see Figure 2) who often collected hundreds of specimens at the very same place. The resulting event-centric datasets reflect collectors' itineraries and can be visualized as animated maps (e.g., Klapperich's Afghanistan Expedition,<sup>12</sup> using advanced, Web-based GIS tools. Once digitized, these itineraries can be downloaded from repositories and used for semi-automatic georeferencing of all specimens collected by the respective collector, which evidently is much more efficient than georeferencing each specimen. However, the itineraries are usually kept in archives separate from the collection, or even in another institution, which poses a transdisciplinarity information integration problem.

The collecting event is an activity carried out by an actor who is a person, the collector, identified by an appellation. It takes place at a locale identified either by a place name and/or spatial coordinates and lasts a certain length of time. An expedition is an activity consisting of continuous collecting activities carried out by participant collectors. In a similar manner, collector's itineraries can be represented as continuing collecting activities (Figure 2).

Replacing the term collector by observer/photographer or painter/drawer in Figure 2, the same model can characterize observation, photographing, and painting or drawing events (instead of a collecting event). All these different events represent concepts related by basic common semantic relationships. The graphic representation of these relationships as shown in Figure 2 makes domain-specific information units machine processable and understandable to everyone independent of the domain.

Itineraries of naturalists provide a good example to demonstrate how different sources can effectively be combined to clarify object history. Itineraries are equally relevant for other disciplines, such as archaeology, ethnology, or the history of fine art objects.

### 4.3 Natural History Information Related to Cultural Objects

Cultural objects can directly refer to knowledge and studies about natural history, typically in times before biology became a science in the modern sense, or in other cultures. Cultural museums, for example, often house paintings (see Figure 6) or other graphical works depicting the animated world in times when no biological voucher specimens were collected, labelled, and deposited in a museum of natural history. Digitizing and indexing this information in a proper way would complement our knowledge of the history of the animated world.

But even in modern times, logbooks from research voyages and materials collected could be distributed to different archives and museums according to the most relevant discipline to which the material pertains. Therefore important contextual information is inaccessible to other memory institutions if they are not informed about these resources. But even when those archival resources are known, their information is difficult to access due to the lack of a consistent standardized indexing with information categories.

Not only are the features which represent the immediate reasons for keeping objects in a cultural history museum sources of knowledge, but any object feature can provide evidence for other unrelated facts. Paintings, for instance, may show recognizable plant or animal species, customs related to particular species, and environmental conditions as secondary subjects. Objects made for some unrelated use may bear respective decorations, such as porcelain items. Similarly, archival texts may refer to such things in a secondary manner. Customs and feeding habits illustrated by objects may be related to particular species or environmental conditions. The materials of the objects themselves or material traces on them may provide direct or indirect evidence of the presence or abundance of species and

<sup>12</sup><http://www.groms.de/groms/animation/Klapperich.html>

environmental conditions, such as feather crowns, leather, fishskin cloths, animal bones, and other extracts (e.g., insects, plants, and minerals as pigment sources), wooden items, fibers, sea shells, or Victorian frames with butterfly wings. Analysis of the debris of excavated archaeological sites provides evidence about the diet of former occupants.

Finally, the documentation of myths and symbols illustrated by museum objects [Dittrich and Dittrich 2004] may refer to natural history information.

With these examples, we argue for large-scale integration efforts of complementary resources that may serve whatever use within or across disciplines. Actually, there are numerous transdisciplinary scenarios, such as archaeology and botany, ethnology and archaeology, musical instruments and ethnology, musical instruments and furniture, but the authors happen to have best access to real-life examples in this particular field.

## 5. IMPROVEMENT OF DATA QUALITY

Ontology-based information modeling can be used to enhance one's own area of expertise and help to identify additional or overlapping sources of information covered by related fields in the same discipline or even by other domains. It is therefore a useful tool for information integration (transdisciplinarity) and should be used as a best-practice guide to improve already existing databases by means of dynamic linking, cross-verification, and semantic enrichment as well as developing new tools via Web services. Furthermore, it can help to identify other formerly hidden or even obscure information sources from various scientific and scholarly domains. The examples discussed in the following are taken from two complex collection databases dealing with zoological specimens.

### 5.1 Cross-Referencing

Specimen-based databases deal with real-world objects, such as pinned insects. Ideally, each specimen bears a label documenting additional primary data such as gathering site, collection date, collector's name, and eventually further details such as habitat or collection methods (see Figure 3, lower left). Once analyzed by a specialist, a scientific name is assigned to the specimen and ideally annotated on the label. This process is called determination. If the species is new to science, the specimen is described in the literature and designated as a type specimen as outlined in Section 4.1. Therefore the process of the creation of a new species is based on a specimen description and type designation which provides a cross-reference between the zoological literature and museum collections. However, most of these data are scattered among the literature and not yet available in digital format. Fortunately, there is an increasing number of Global Taxonomic Authority Files, such as Fish Base,<sup>13</sup> the Hymenoptera Name Server,<sup>14</sup> and the Orthoptera Species File. They provide information about the validity of a given taxon name (reflecting human concepts which can change in time), the literature reference of the original description, the *locus typicus* (place where collected), and the depository place of type specimens. This information is primarily compiled from taxonomic publications so that all these databases need a module dealing with bibliography.

Reciprocal referencing between specimen-based databases of type specimens and the respective Global Taxonomic Authority Files (species-based taxonomic databases) allows for validity checks of taxon names or, reversely, comparisons of the original label information of the respective primary type (i.e., object) with the information published in the species description. Some of the few examples of linking such databases is presented here.

<sup>13</sup><http://www.fishbase.org/search.php>

<sup>14</sup>[http://atbi.biosci.ohio-state.edu:210/hymenoptera/nomenclator.home\\_page](http://atbi.biosci.ohio-state.edu:210/hymenoptera/nomenclator.home_page)

The figure consists of four browser windows illustrating the integration of specimen-based information units (DORSA) with species-based authority files (OSF) and geographic data (CBIF).

- Top Left Window (SysTax - database query):** Shows detailed taxonomic information for *Noctitrella plurilingua* Ingrisch, 1997. It includes nomenclatural data, taxonomical data (classification from Animalia Metazoa to Gryllidae subfam. Podoscirtinae tr. Podoscirtini), and literature references. A yellow arrow labeled "OSF > DORSA" points from this window to the top right window.
- Top Right Window (Taxa display):** Shows the species page for *Noctitrella plurilingua* Ingrisch, 1997, including classification, type specimen information, and a link to "Specimens in German collections (DORSA)". A yellow arrow labeled "DORSA > OSF" points from this window back to the top left window.
- Middle Left Window (SysTax - database query):** Shows "data of collections" for *Noctitrella plurilingua* Ingrisch, 1997, including determination, individual data, and a list of pictures. Green arrows point from these picture links to the middle right window.
- Middle Right Window (Query results from the SysTax database):** Displays a large image of a specimen (Habitus lateral view) and a wave file (1573n00p.wav) with a remark: "zus.mit Liara Monkra e.l.". A green arrow labeled "large image" points to the specimen image, and another green arrow labeled "wave file" points to the audio file link.
- Bottom Left Window (SysTax - database query):** Shows a dynamic map service from the Canadian Biodiversity Information Facility (Canada) for *Noctitrella plurilingua* Ingrisch, 1997. A yellow arrow points from this window to the bottom right window.
- Bottom Right Window (SysTax - Noctitrella plurilingua Ingrisch, 1997):** Displays a geographic map of Southeast Asia, showing the distribution area of the species. A yellow arrow points from this window back to the top left window.

Fig. 3. DORSA specimen-based information units with images and sounds connected by internal links (green arrows) and external links (yellow arrows) to the species-based OSF authority file and to the geographic CBIF (Canadian Biodiversity Information Facility) map server.

Within the Digitized Orthoptera Specimen Access (DORSA)<sup>15</sup> project, a virtual museum of Orthoptera type specimens (grasshoppers, locusts, katydids, crickets) housed in major German museums was created. The Global Taxonomic Authority File Orthoptera Species File (OSF)<sup>16</sup> served and still serves as a taxonomic backbone. The OSF was initiated by D. Otte in printed format and is now accessible in digital format [Eades and Otte 2007].

DORSA contains full information about 16,000 specimens, including 2,300 primary types and 6,700 secondary types. 30,000 images and 11,000 sound recordings are linked to their respective specimens [Lampe et al. 2005]. Many of the specimens linked to the sound recordings belong to hitherto undescribed species, which means that they cannot be linked to the taxonomic authority files. They can be considered as “types of tomorrow” [Riede et al. 2006]. The DORSA virtual museum is available through the SYSTAX<sup>17</sup> database infrastructure and major knowledge portals such as BIOCASE or GBIF. Figure 3 shows screenshots of specimen-based multimedia information units, including sound files. The DORSA Web-based repository provides sufficient detailed information, including diagnostic features such as genitalia, to allow, for example, taxonomists to narrow down loan requests from other institutions.

The DORSA project is one example of how related information should be integrated within a virtual museum. Moreover, images and sound files should be linked to their respective specimen files within a domain-specific environment quite similar as to external authority files from other domains (Figure 3).

Figure 4 shows the present online status of the collection management and information system BIODAT, a specimen-based database used in Bonn at the Zoologisches Forschungsmuseum Alexander Koenig (ZFMK) and in Berlin at the Museum für Naturkunde of Humboldt University. The system design is based on events related to biological specimens such as collecting, determination or type creation events. Gathering sites are stored in BIODAT with geographical coordinates, a precondition to check the quality of the respective information against geographical authority files and map servers. Scientific names of biological specimens are internationally standardized and therefore unambiguous and can be validated by cross-referencing them to external taxonomic authority files. In case of name-bearing primary types (Holotype,<sup>7</sup> Syntypes,<sup>18</sup> Lectotype,<sup>19</sup> Neotype<sup>20</sup>) the specimen information can be cross-referenced to their species authority files and vice versa. In contrast to scientific species names, the names of people (collectors, determinators, and authors) are ambiguous. Therefore the identity of a person has to be clarified by checking their life history against biographic authority files (e.g., dates of birth, death, and other events).

BIODAT<sup>21</sup> specimen data and related images are available online via BIODAT’s user-friendly query tool<sup>22</sup> and via the BIOCASE and GBIF portals. It provides a Google-like easy access to fine-grained information about specimens housed in these collections. By selecting preferences, the user can decide what kind of information is represented as well as the number of results on display. The query can refer

<sup>15</sup><http://www.dorsa.de/>

<sup>16</sup><http://osf2.orthoptera.org>

<sup>17</sup><http://www.biologie.uni-ulm.de/systax>

<sup>18</sup>*Syntypes* are specimens of a type series that collectively constitute the name-bearing type. (International Code on Zoological Nomenclature, 4th Edition, 1999).

<sup>19</sup>A *lectotype* may be designated from syntypes to become the unique bearer of the name of a nominal species-group taxon and the standard for its application (International Code on Zoological Nomenclature, 4th Edition, 1999).

<sup>20</sup>*Neotype*: A is single specimen designated as the name-bearing type of a nominal species or subspecies when there is a need to define the nominal taxon objectively and no name-bearing type is believed to be extant (International Code on Zoological Nomenclature, 4th Edition, 1999).

<sup>21</sup><http://www.biodat.de>

<sup>22</sup>The query tool was developed by BIODAT’s programmer, Dirk Striebing.

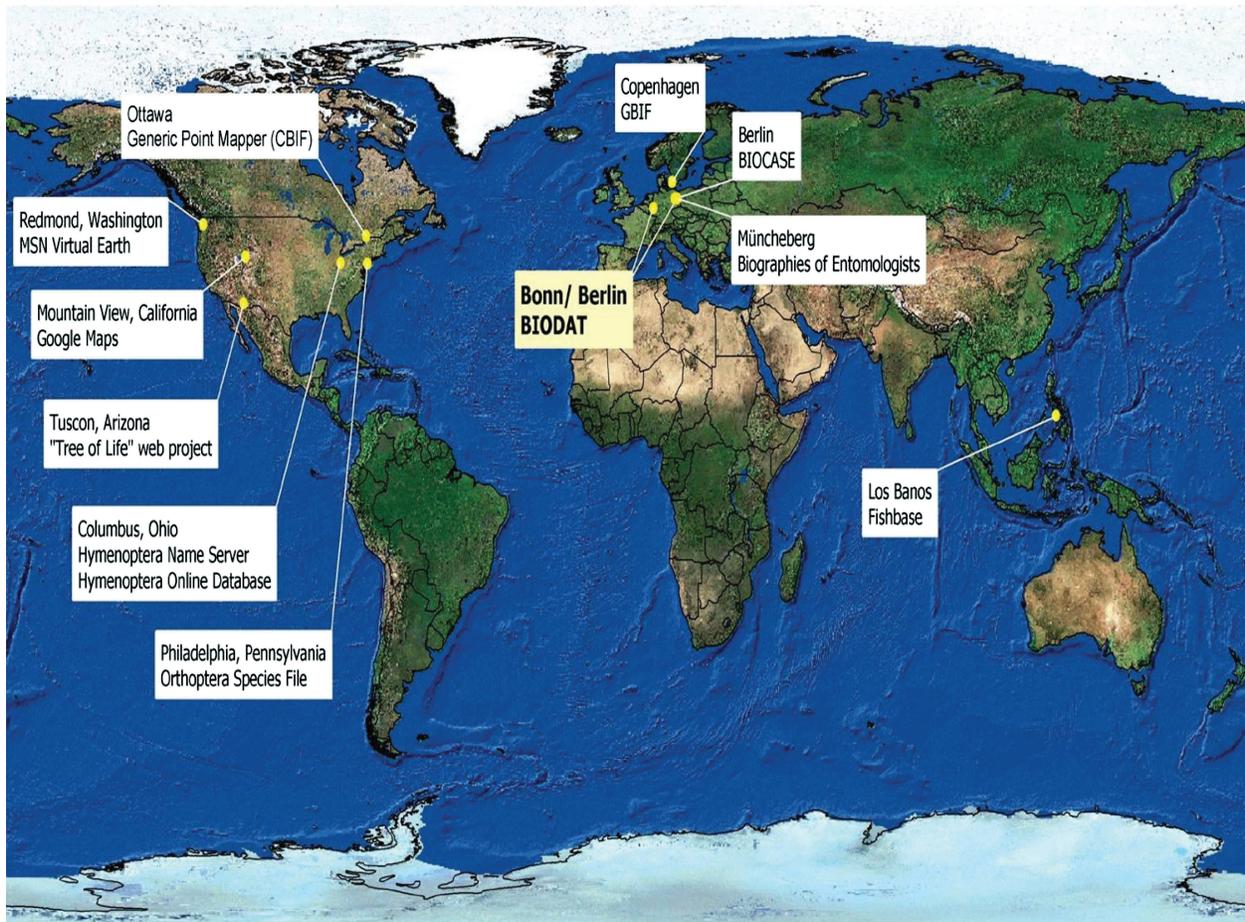


Fig. 4. Cross-referencing of the specimen-based database BIODAT to various global authority files, to map servers and to portals.

to different kinds of information, such as collector and/or gathering site, and systematic information, such as order, family, or genus. The end user has direct access to various global authority files such as taxonomic and biographic name servers as well as geographical map servers by cross-referencing.

## 5.2 Multiple Verification

**5.2.1 Taxonomic Data / Information.** OSF authority files allow a comparison of published type data with our DORSA specimen data and vice versa. Both databases are linked by a common ID, referring to the name of the type in OSF and to the type specimen in DORSA. The latter is the real-world voucher on which a species description is based. DORSA and OSF are two separate relational databases, optimized to administer specimens and their species names (taxonomic concepts), respectively. Because the species database is based on data mainly from published information, while the specimen database collects its information from the specimens and labels in the collections, inconsistencies must be expected and were found (Table I). Some type specimens were lost because of the destruction of collections by fire. Other inconsistencies required closer investigation, revealing type specimens in museums not yet represented in OSF (Table I, primary types not listed in OSF). As a result, type specimen information in OSF could

Table I. Comparison of Orthoptera Primary Types<sup>23</sup> in German Collections According to OSF and DORSA (After Ingrisch et al. [2004b])

	Global Taxonomic Authority File	Specimen based Museum data	Reciprocal Verification				
			Primary Types not listed in OSF	Taxa not listed in OSF	OSF-Type data not confirmed in Museums	Primary Types lost	Primary Types newly recognized (unlabeled)
	Taxa in OSF (April 2000, of the beginning of the DORSA project)	Primary Types in German Museums checked (December 2002)					
Berlin	1.093	1.423	330	68	15		12
Eberswalde	45	79	34			1	
Dresden	66	117	55	10	6		7
Hamburg	142	135	43	1	5	55	
Halle	38	55	25	4	4		16
Bonn	6	44	12	13			
Frankfurt	82	151	69	5			1
Stuttgart	47	112	65	5			3
Munich	27	54	21	1		1	
Sum	1.546	2.170	654	107	30	57	39

be updated with the information provided by DORSA, while the information in OSF helped to discover type specimens in museum collections that were not properly labeled [Ingrisch et al. 2004a].

In this process, our DORSA taxonomist, Sigfrid Ingrisch, detected a discrepancy level of about 25% between the two sources, such as 654 primary types which had not been listed in OSF and 39 unlabeled, newly recognized primary type specimens within German museums. Besides this, exact data about missing type specimens, such as the loss of 55 name-bearing types in Hamburg during world war II, allow initiation of effective searches to fill these gaps, for example, by designating neotypes.<sup>23</sup>

The example shows that the explicit integration of digital information between disciplines is highly desirable, and equally so between different fields in the same domain and virtually between any information system within a domain. The discrepancies are striking. Because these are core data sets for biodiversity studies, they are expected to be mostly error free. Data-cleaning procedures, such as multiple verification (as outlined previously), are efficient tools for improving biodiversity data quality but require suitable information linking.

**5.2.2 Geographic Data/Information.** In a similar manner, geographic information quality can be improved by comparing collector's itineraries with geographic information from specimen labels. The basic principle is to reconstruct the original way or routes collectors have taken by using isolated original label information on single collecting events. These routes can be compared to published or archival information (e.g., itineraries) if available, and thereby inconsistencies can easily be identified. No one can be at two places at the same time.

The inset in Figure 2 shows part of Klapperich's itineraries reconstructed from his publication [Klapperich 1954]. The black dots symbolize localities extracted from 3,824 collecting events (covering approximately 10,000 insects from the Klapperich material, georeferenced and databased within BIODAT). The Afghanistan routes are clearly visible but coincide only in part with the itineraries reconstructed from literature (colored lines). In a first analysis, outliers make up about around 25% of all points (ten times more than expected) and are due to ambiguous place names, mistakes during georeferencing, and possibly mistakes within the used gazetteers [Riede and Lampe to appear]. This shows that georeferencing based only on place names of labels is less reliable than georeferencing based

<sup>23</sup>Primary types are name-bearing type specimens (compare footnotes 7, 21–23).

on careful reconstruction of itineraries, resulting in a collector-specific thesaurus. Another advantage is that poor geographic label information (e.g., we found Al Wahat meaning Oasis) can be georeferenced easier if the collector's route is known. In addition, the number of localities to be referenced is orders of magnitudes smaller than the number of collected specimens, and therefore much more efficient. Again, to do so, information integration between systems, institutions, and disciplines is required.

### 5.3 Semantic Enrichment

**5.3.1 Georeferencing.** Someday georeferenced museum collections will form a unique contribution to the mapping of the Earth's biodiversity and its historic distribution. Using Geographical Information Systems (GIS), distribution maps can be created and intersected with layers such as political boundaries or vegetation zones, present and former extension of tropical rain forests, or maps reflecting anthropogenic impact (human footprint [Sanderson et al. 2002]). In summary, adding latitude-longitude coordinates to localities can be considered among the most basic procedures of semantic enrichment, facilitating the display of places on a map.

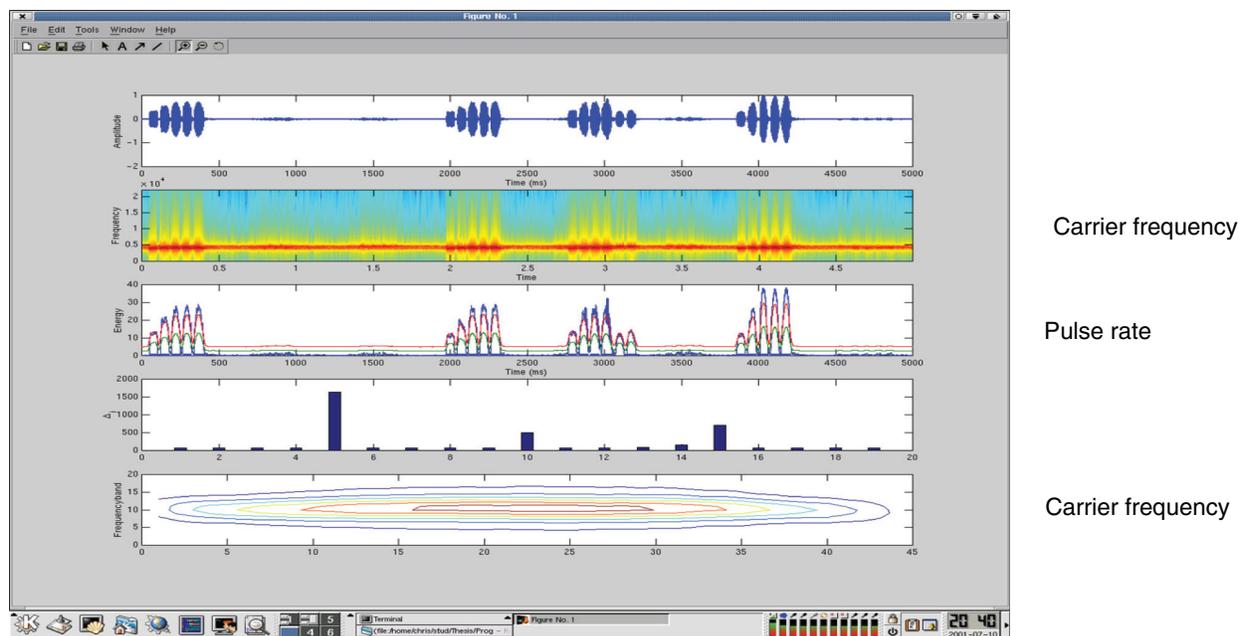
Mapping helps to visualize and reduce ambiguity, particularly in the Americas and Australia, where the same place names may be used for multiple locations. Ambiguity is reduced if the broader geographic information of the location is explicitly mentioned. Moreover, communities may merge or change names more frequently than one expects them to do. Biodiversity conditions normally apply to wider areas than the point of observation. Therefore extraction of intermediate geographical information units such as county or district from archival data provides a geographical-semantic enrichment of the respective specimens files quite similar to the enrichment of sound files. As demonstrated by the BIODAT example, collection management systems should have user-friendly interfaces to gazetteer information. The creation and maintenance of gazetteers normally is not regarded as a task of biologists. We see a strong need for collaboration between biodiversity and the disciplines dealing with gazetteers. The coverage, level of detail, and historical depth of the gazetteers must be appropriate for specimen collection.

**5.3.2 Sound Data/Information.** In a similar manner, multimedia data have to be semantically enriched before their content can be stored and analyzed within databases. In the case of insect songs, carrier frequency and pulse repetition rate are among the basic characteristic features to characterize songs. However, up until now, feature extraction was done manually in a time-consuming process of sound analysis. Within the DORSA project, a software module was developed to extract these basic features and was applied to the entire DORSA sound repository (Figure 5). The calculated sound parameters frequency and pulse repetition rate were then annotated into new columns of the sound file table [Riede et al. 2006]. As a result, important sound parameters now complement the respective sound files. This enables one to search for and evaluate sound parameters by standard SQL queries and retrieve the respective sound files from the repository. Even complex data mining is feasible, such as "Are there any Orthoptera (in our database) from Ecuadorian lowland forests, singing higher than 10 kHz."

More complex verbose descriptions of songs (tzitzi) are also possible and could be really useful, but specially developed speech recognition tools are needed to extract these features automatically.

**5.3.3 Biographical Data/Information.** Most specimen or object-based databases consider only the names of people, but names alone are not sufficient to describe the identity of people. Other information is needed such as birth, death, and other events. This information can be picked up in biographical authority files and enrich the original database with a citation of the source. Biographical Authority Files deal with the biographics of people and their identity. Examples of such files are Biographies of the

## Extraction of sound parameters by using MatLab Software



In cooperation with: Dept of Neuroinformatics, Ulm University; PhD thesis C. Dietrich

Fig. 5. Semantic enrichment of specimen-based databases by extraction of sound parameters. Software tools were developed by C. Dietrich as part of his thesis on automatic identification of cricket songs through neural networks [Dietrich et al. 2004].

Entomologists of the World<sup>24</sup> kept by the German Entomological Institute, the United List of Artists Names<sup>25</sup> of the Getty Foundation and the commercial World Biographical Dictionary of Artists<sup>26</sup> of the Saur publishing company. However, a reciprocal cross-referencing of people-related events (documented in domain-specific databases) to biographic authority files is recommended to help the curator to identify inconsistencies.

**5.3.4 Conservation Data/Information.** IUCN Red List database<sup>27</sup> contains threat status standardized by the well-known Red List information. However, searching the Red List requires considerable biological knowledge of taxonomy and systematics which cannot be expected from all users. For example, imagine a forester querying the Red List database for extinct or critically endangered tropical timber trees. Due to the fact that the life form tree is not encoded in the Red List database, it is practically impossible to retrieve this information (resulting in tree frogs, etc.). Trees can occur in many plant families, and presently all tree families have to be queried, to answer a seemingly simple question. However, once the dataset is semantically enriched by life form (bush, tree, liana), this question is easy to answer.

As shown in the examples of Sections 5.3.1–5.3.4, such semantical enrichment can be partly automated if datasets are ontologically well documented.

<sup>24</sup>[http://www.zalf.de/home\\_zalf/institute/dei/php\\_e/biograph/biograph.php](http://www.zalf.de/home_zalf/institute/dei/php_e/biograph/biograph.php)

<sup>25</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)

<sup>26</sup><http://www.saur.de/akl/english/index.htm>

<sup>27</sup><http://www.redlist.org>

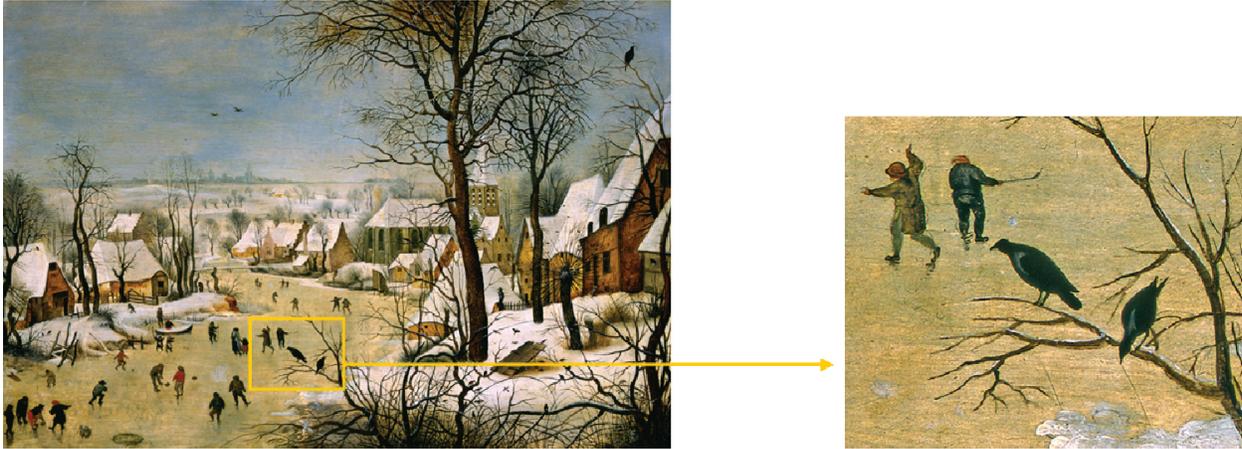


Fig. 6. After Breugel (Pieter Bruegel the Younger) “Winter Landscape with Bird Trap” (1601) with resting hooded crows (enlarged) (Kunsthistorisches Museum Wien/Austria, Inv. No. GG625, used with permission).

## 6. GENERATING SCIENTIFIC ADDED VALUE

As indicated previously, semantically enriched and well-related datasets allow queries from a wider community of users who can easily generate scientific added value. Apart from the examples already sketched, the following examples highlight some additional applications.

### 6.1 The Little Ice Age

Meteorologists detected a general cooling of the climate between the years 1150 and 1460 and a very cold climate between 1560 and 1850, the so called Little or Minor Ice Age, which had a wide impact on agriculture, health, economics, and even art and literature [Mandia 2006]. Neuberger [1970] found evidence for meteorological information in more than 50% of about 12,000 paintings from 41 museums in Europe and the US, painted over a period from 1400 to 1967. Paintings created during the Little Ice Age period such as in Figure 6 show more winter landscapes in comparison to warmer periods.

In addition in several Dutch paintings from the 16th and 17th century, one can find hooded crows *Corvus corone cornix* Linnaeus, 1758 instead of carrion crows, *Corvus corone corone* Linnaeus, 1758 [Misof 2005, personal communication; Dittrich and Dittrich 2004].<sup>28</sup> In Central Europe the hooded crow is found today only east of the Elbe River and in Scandinavia; none can be found in the Netherlands. Climate shifts are therefore not only recorded in traditional methodological approaches such as geological sediments or tree rings.

### 6.2 Turkeys in the Cathedral of Schleswig

After the restoration of paintings in the Cathedral of Schleswig in 1938, Hamkens [1940], published an image of an animal frieze with turkeys (*Meleagris gallopavo* Linnaeus, 1758) painted on medieval plaster of the cloister of the cathedral that was built between 1310 and 1320 (Figure 7). According to Kopp [2000], the record led to speculations about the discovery of America by Vikings in pre-Columbian times because Spanish sailors had not found turkeys in North America and Mexico until 1530 and brought them still later to Europe. In 1952, the restorer of the paintings, Lothar Malskat, admitted having added the turkeys to already existing medieval animal friezes. They therefore were proved to be frauds.

<sup>28</sup>Dittrich & Dittrich created a dictionary of animal symbols by analyzing paintings from the 14th to 17th century.



Fig. 7. Turkey as an element of an animal frieze in the Cathedral of Schleswig. (Used with the permission of EV ang.-luth., Domgemeinde Schleswig.)

This is a complex example of correlated cultural and natural history: species distribution is changed by historical events, and species distribution allows for criticizing historical information, the frieze. It shows how information integration can provide the associations necessary to assess the consistency of biological and historical information.

### 6.3 Wryneck *Jynx torquilla* Linnaeus, 1758, and Love Spell

An example of relating such distinct disciplines as ornithology and anthropology is presented by Vajda [2004] who elucidates the complex relations between a bird (wryneck) and lovespell objects. The word *inyx* in Greek antiquity denoted both the wryneck (*Jynx torquilla* L., belonging to the woodpecker family) and a small wheel-shaped ritual object which could be made to spin, producing a humming noise. Both the bird and the magic wheel played a role in love magic, but the reason for the relation between bird and wheel remains elusive (Figure 8).

A potentiation is observed in some traditions where a caught bird is bound on the wheel to increase the effect of the love spell. The peculiar courtship display, consisting of upright position and slow 180-degree head movements, together with the peculiar (and eventually name-giving) song of the comparatively small bird, might be the underlying reason for the strange association. The author explores the comparatively wide geographic distribution of this lore, diagnosing its absence in Egypt where it is not mentioned in the otherwise exhaustive text of the magic papyrus texts, which dedicate their entire fifth part to love spells [Abt 1908, p. 179; Graf 1996, p. 161]. A possible explanation becomes evident by a glance at the Wryneck's distribution map (Figure 9) extracted from the Global Register of Migratory Species (GROMS<sup>29</sup>). The breeding distribution (yellow) of the bird does not extend into Egypt. However, the much wider wintering distribution (blue) does extend into Africa but probably does not relate to the spell because the strange head movements are only observed during courtship, hence within the

<sup>29</sup><http://www.groms.de>

The screenshot shows the Museum of Fine Arts, Boston website. The main content area displays search results for a collection search. The object shown is a terracotta lynx-wheel, a circular object with eleven birds (wry-necks) seated on the rim of a four-spoked wheel. The object is described as being from the Greek, Late Geometric Period, 750-700 B.C. The page also includes a 'Please note' section regarding the use of images and a 'Provenance/Ownership History' section.

**collections**  
 Advanced Search  
 Art of Asia, Oceania + Africa  
 Art of Europe  
 Art of the Americas  
 Art of the Ancient World  
 Contemporary Art  
 Musical Instruments  
 Prints, Drawings + Photographs  
 Textile + Fashion Arts  
 Conservation + Collections Care  
 Online Tours  
 Recent Acquisitions  
 Resources

**collections search results**  
 Collections Search Results  
 Individual object from advanced search  
 return to results

Collection Search  
 330,321 artworks  
 Advanced Search

Done with your search?  
 Tell us what you think.

**Please note:**  
 If you have comments or questions regarding objects in the collection or about the results of your search, please write to [webmaster@mfa.org](mailto:webmaster@mfa.org). Note that some of the electronic records indicate that they have not been reviewed recently by curatorial staff and might need revision; also, please note that a small percentage of the MFA's collection is not presently searchable online.

We are pleased to share images of objects on this Web site with the public as an educational resource. While these images are not permitted to be used for reproduction, we encourage you to do so by [visiting our image rights page to submit a request](#).

**Ritual object with eleven birds on a wheel**  
 Greek, Late Geometric Period, 750-700 B.C.  
 Place of Manufacture: Athens, Attica, Greece  
 Height: 6.50 cm (2 9/16 in.); diameter: 21.5 cm (8 7/16 in.)  
 Terracotta  
**Classification:** Sculpture  
 Lynx-wheel. Eleven iynx birds (wry-necks) with folded wings are seated on the rim of a four-spoked wheel. One bird is restored. Linear geometric designs on upper and lower parts of the wheel. On the underside, two groups of ten and twelve apotropaic eyes (leaves), cross-hatching cross-hatched triangles, and checker-board.  
 Museum of Fine Arts, Boston  
 John Michael Rodocanachi Fund, 1928  
**Accession number:** 28.49  
**Provenance/Ownership History:** By date unknown: said to come from Phaleron; by 1928: with Seltman; purchased by MFA from Seltman, March 1, 1928, for \$ 463.30

Fig. 8. Picture of the lynx wheel found in the collection database of the Boston Fine Arts Museum (<http://www.mfa.org>), using the accession number 28.49 in the advanced search. The accession number is cited in Nelson [1940, p. 443, footnote 2].

breeding range. Consequently, one could predict that the Indian version of the myth should correlate with the rather limited North Indian breeding range of the bird. A somewhat easier correlation between geographic distribution and culture could be based on local names. It is hypothesized that the scientific genus name *lynx* is based on an onomatopoeic description of the song. Any other (or similar) local names based on the song should only be found in languages overlapping with the (former) breeding range.

#### 6.4 Repatriation of Knowledge

Another advantage of georeferencing specimen data is that one can easily create a map illustrating the countries of origin of type specimens (Figure 10). German researchers share a significant part in exploring the Orthoptera fauna of the world. Examples of significant historical expeditions are those of Hemprich and Ehrenberg to North Africa and West Asia [Klug 1830–1832], Preuss 1891 to

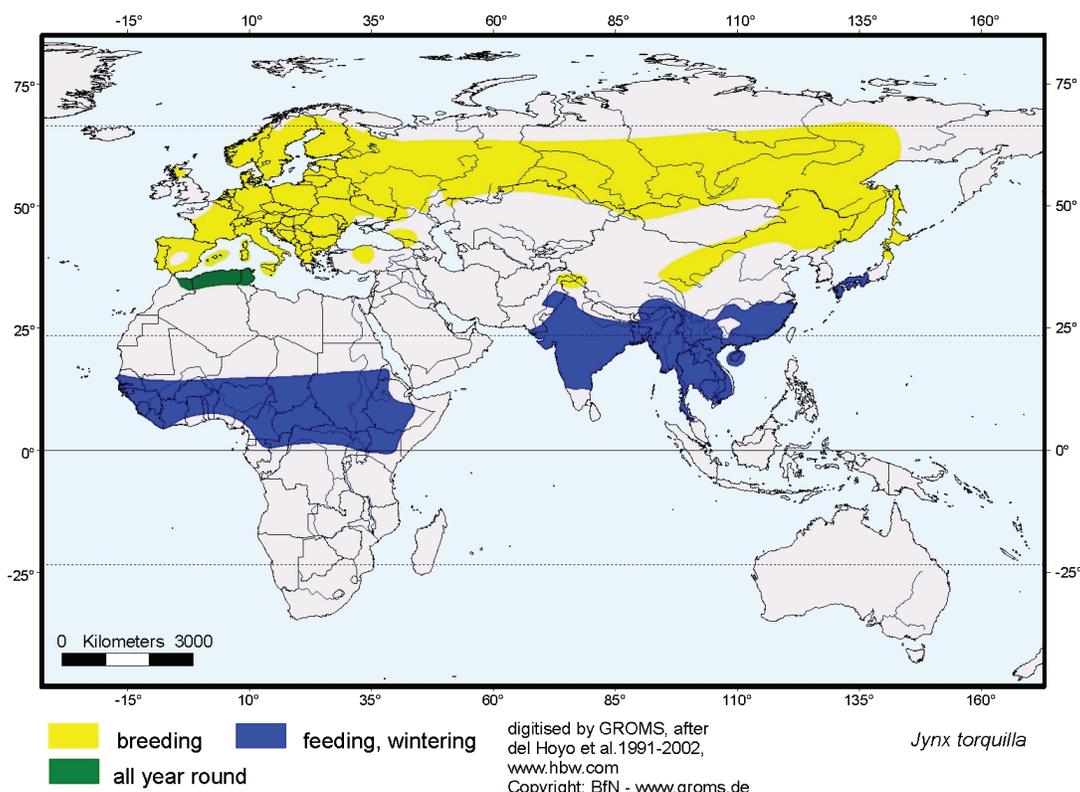


Fig. 9. Distribution map of the Wryneck *Jynx torquilla* (Linnaeus 1758).

Cameroon [Karsch 1891(a), (b)], the Kaiserin-Augustafuss-Expedition to New Guinea 1912–1913 (e.g., Karny [1928]), and Ramme's revisions of African, Southeast Asian, Southeast European and West Asian Acrididae (e.g., Ramme [1929, 1941, 1951]). Due to these expeditions and many others, including modern ones, German museum collections hold type specimens from all over the world (Figure 10). Hence, publishing type specimen data and images through a virtual museum signifies the repatriation of knowledge to countries of type origin.

## 7. CONCLUSION AND PERSPECTIVES

This article refers to fundamental core ontologies that allow for making sense across domains. Concerning a research agenda, it aims at identifying needs, challenges, possible approaches, and respective feasibilities.

Complex phenomena or problems, such as those described in the field of biodiversity, cannot be satisfactorily explained or solved within a single scientific or scholarly domain. Transdisciplinary approaches by means of information integration provide a better understanding of this complexity. Through real-life examples, we show how information integration can provide transparency of sources of evidence in such a way as to enable everyone to check the arguments for verification of information from multiple sources.

We have shown that machine-supported semantic information integration in the described computer applications has a huge potential. We found surprisingly large inconsistencies (about 25%) in the

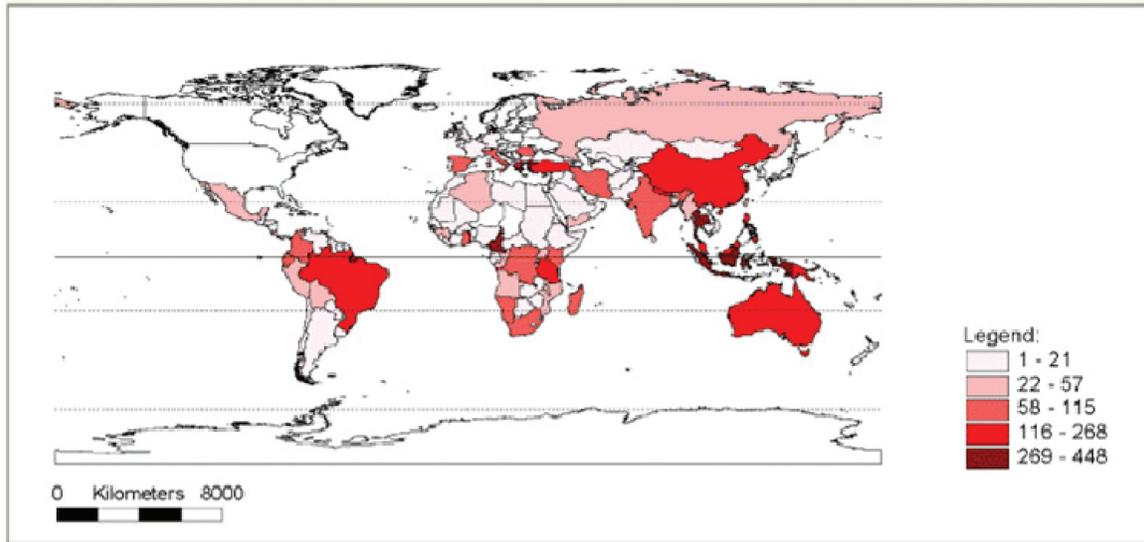


Fig. 10. Countries of origin of Orthoptera type-specimens housed in German Museums, collected during significant historical expeditions.

information content and handling within systems of the same domain, but with a different focus. Information integration, therefore, provides the means to improve the quality of biodiversity core data. We expect similar problems in all disciplines where information is transferred manually between systems such as in archaeology from field observation to museum collections. Furthermore, transdisciplinary information integration also leads to new knowledge and understanding of the relationships between nature and culture.

Whereas other published applications of the CIDOC CRM [Crofts 1999; Doerr et al. 2004] have referred to information integration within an application domain, we have demonstrated how ontology-based modeling and, in particular, the event-centric ontology of the CIDOC CRM provides a suitable approach to describing explicitly the relevant relations of information to be integrated within and across specific disciplines, expanding integration beyond cultural information in the narrower sense. The integration of particular information is not simply a matter of terminology as is widely assumed, but is rather more a matter of the identity of particular specimens, people, and places. Having described the scientific benefits of integration of specific kinds of information, it is probable that similar cases are prevalent within all cultural heritage disciplines and descriptive sciences.

Following these experiences, we recommend that, for integrated information systems developed for research museums and descriptive sciences, a move from a class and object-centric to an event-centric approach be instituted. For example, a virtual worldwide catalogue of collecting events could serve as a backbone for an accelerated data capture of comprehensive collection information and for improving data quality, and thereby the identity of persons, objects, places, and concepts would be ontology-based in a domain-neutral way. These concepts help to speed up or even automate semantic enrichment which is a necessary prerequisite to making databases more user-friendly and preparing them for advanced Web services. In addition, Web services should allow easy, uniform access methods to merge local domain-specific and global authority information.

## ACKNOWLEDGMENTS

We thank Dr. Siegfried Krause, Germanisches Nationalmuseum (GNM) Nuremberg, for helpful discussions and comments and Dr. Steven Stead, Paveprime Ltd., London, for checking the ontologies. Pastor Pfeifer, Domgemeinde zu Schleswig, supplied an image of the turkey in Figure 7. Dr. Bernhard Misof, ZFMK Bonn, led us to the phenomenon of hooded crows on paintings from the 16th and 17th century, and the ornithologist Dr. R. van den Elzen (ZFMK) confirmed the respective bird determinations. Dr. Dagmar Hirschfelder (GNM) and Dr. Gerhard Muehe (BIOTA South Africa project, Uni. Hamburg) helped in investigating literature. We thank the programmer of BIODAT, Dirk Striebing, Museum für Naturkunde, Berlin, our colleagues Dr. Sigfrid Ingrisch, the DORSA data capture team, Dipl. Biol. Birgit Klasen, Dirk Rohwedder and the BIODAT- and BIOTA E15- data capture teams, all ZFMK, for their permanent support. Finally we thank Dr. Bradley Sinclair (ZFMK) and Mike Shaw, Bonn, for checking the English text.

## REFERENCES

- ABT, A. 1908. Die Aplogie des Apuleius von Madura und die antike Zauberei. *Religionsgeschichtliche Versuche und Vorarbeiten IV*, Heft 2, Giessen.
- CHEN, H., MARTINEZ, J., NG, T. D., AND SCHATZ, B. R. 1997. A concept space approach to addressing the vocabulary problem in scientific information retrieval: An experiment on the worm community system. *J. Amer. Soc. Inform. Sci.* 48, 1, 17–31.
- CROFTS, N. 1999. Implementing the CIDOC CRM with a relational database. *MCN Spectra* 24, 1.
- CROFTS, N., DOERR, M., GILL, T., STEAD, S., AND STIFF, M. 2003. Definition of the CIDOC conceptual reference model. [http://cidoc.ics.forth.gr/definition\\_cidoc.html](http://cidoc.ics.forth.gr/definition_cidoc.html).
- DEMPSEY, L. 2000. Scientific, industrial, and cultural heritage: A shared approach: A research framework for digital libraries, museums and archives. *Ariadne* 22, <http://www.ariadne.ac.uk/issue22/dempsey/intro.html>.
- DIETRICH, C., PALM, G., RIEDE, K., AND SCHWENKER, F. 2004. Classification of bioacoustic time series based on the combination of global and local decisions. *Int. J. Patt. Recogn.* 37, 2293–2305.
- DITTRICH, S. AND DITTRICH, L. 2004. *Lexikon der Tiersymbole. Tiere als Sinnbilder in der Malerei des 14.-17. Jahrhunderts*. 2nd Ed. Imhof, Petersberg, Germany.
- DOERR, M. 2003. The CIDOC CRM—An ontological approach to semantic interoperability of metadata. *AI Magazine* (Special Issue on *Ontologies*) 24, 3.
- DOERR, M., SCHALLER, K., AND THEODORIDOU, M. 2004. Integration of complementary archaeological sources. In *Proceedings of Computer Applications and Quantitative Methods in Archaeology*. Prato, Italy.
- DUCKWORTH, W. D., GENOWAYS, H. H., AND ROSE, C. L. 1993. Preserving natural science collections: Chronicle of our environmental heritage. National Institute for the Conservation of Cultural Property, Washington, DC.
- EADES, D. C. AND OTTE, D. 2007. Orthoptera species file online. Version 2.0/3.2. <http://osf2.orthoptera.org/HomePage.aspx>.
- GILL, T. 2004. Building semantic bridges between museums, libraries and archives: The CIDOC Conceptual Reference Model. *First Monday* [http://www.firstmonday.org/issues/issue9\\_5/gill/](http://www.firstmonday.org/issues/issue9_5/gill/).
- GÖTTLICHER, D. AND BENDIX, J. 2004. Eine modulare Multi-User Datenbank für eine ökologische Forschergruppe mit heterogenem Datenbestand. *Zeitschrift für Agrarinformatik* 4/04, 95–103.
- GURALNICK, R. P., WIECZOREK, J., BEAMAN, R., HJLMANS, R. J. AND THE BIOGEOMANCER WORKING GROUP. 2006. BioGeomancer: Automated georeferencing to map the world's biodiversity data. *PLoS Biology* 4, 11, 381.
- GRAF, F. 1996. *Gottesnähe und Schadenzauber. Die Magie in der griechisch-römischen Antike*. München, Germany.
- GRUBER, T. R. 1993. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.* 43, 907–928.
- GUARINO, N. 1998. Formal ontology and information systems. In *Proceedings of the 1st International Conference on Formal Ontology in Information Systems*, N. Guarino, Ed. Trento, Italy, IOS Press, 3–15.
- HAMKENS, F. H. 1940. Der Truthahn im Dom zu Schleswig. *Germanenerbe* 5, 155–156.
- INGRISCH, S., RIEDE, K., AND LAMPE, K.-H. 2004. DORSA—Virtuelles Museum deutscher Orthopterensammlungen online. *Mitteilungen der Deutschen Gesellschaft für allgemeine und angewandte Entomologie* 14, 479–482.
- INGRISCH, S., RIEDE, K., LAMPE, K.-H., AND DIETRICH, C. 2004. DORSA—A “Virtual Museum” of German Orthoptera Collections. *Memorie della Societa Entomologica Italiana* 82, 2, 349–356.
- KALFOGLOU, Y. AND SCHORLEMMER, W. M. 2003. Ontology mapping: The state of the art. *Knowl. Engin. Rev.* 18, 1, 1–31.

- KARNY, H. H. 1928. Ueber die Gryllacriden der Deutschen Kaiserin Augustafuß-Expedition 1913. *Mitteilungen aus dem Zoologischen Museum in Berlin* 14, 3–4, 83–114.
- KARSCH, F. 1891a. Uebersicht der von Herrn Dr. Paul Preuss auf der Barombi-Station in Kamerun gesammelten Locustodeen. Als Anhang: Ueber die Mecopodiden. *Berliner Entomologische Zeitschrift* 36, 317–346.
- KARSCH, F. 1891b. Verzeichniss der von Herrn Dr. Paul Preuss in Kamerun erbeuteten Acridiideen. *Berliner Entomologische Zeitschrift* 36, 1, 175–196.
- KLAPPERICH, J. 1954. Auf Forschungsreisen in Afghanistan. *Entomologische Blätter* 50, 107–118.
- KLUG, J. C. F. 1830–1832. Symbolae physicae seu icones et descriptiones Insectorum, quae ex itinere per Africam borealem et Asiam occidentalem Friderici Guilelmi Hemprich et Christiani Godofredi Ehrenberg medicinae et chirurgiae doctorum studio novae aut illustratae redierunt. In *Symbolae Physicae Zool. 2 (Insect. 3), Decas prima, fol. a-i, pls 1-10 [no Orthoptera](1829); Decas secunda, fol. a-f, pls. 11-20 (1830); Decas tertia, fol. a-l, pls. 21-30 (1832)*, F. G. Hemprich, and C. H. Ehrenberg, Berlin, Germany, 1929–1945.
- KOPP, P. 2000. Die Truthähne im Schleswiger Dom. *Archäologisch*. <http://www.archaeologisch.de/truthahnschleswig.html>.
- LAMPE, K.-H., RIEDE, K., AND INGRISCH, S. 2005. Repatriation of knowledge about insects and types through the DORSA virtual museum. *Beiträge zur Entomologie* 55, 2, 477–484.
- MANDIA, S. A. 2006. Influence of dramatic climate shifts on european civilizations: The rise and fall of the Vikings and the Little Ice Age in Europe. <http://www2.sunysuffolk.edu/mandias/lia/index.html>.
- MAUTHE, G. 1994. Die Österreichische Brasilienexpedition 1817–1820. In *Brasilianische Reise 1817–1820. Carl Friedrich Philipp von Martius zum 200. Geburtstag*. J. Helbig, Ed. Hirmer Verlag, München, Germany, 13–27.
- MITTELSTRASS, J. 2001. On transdisciplinarity. In *Proceedings of Science and the Future of Mankind. Science for Man and Man for Science*. Vatican City, 527, XV tables, 495–500.
- MURPHEY, P. C., GURALNICK, R. P., GLAUBITZ, R., NEUFELD, D., AND RYAN, J. A. 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the mountain and plains spatio-temporal database-informatics initiative (Mapstedi). *PhyloInformatics* 3, 1–29.
- NELSON, G. W. 1940. A Greek votive Iynx-wheel in Boston. *Amer. J. Archaeol.* 44, 4, 443–456.
- NEUBERGER, H. 1970. Climate in art. *Weather* 25, 46–56.
- RABY, P. 1997. *Bright Paradise: Victorian Scientific Travellers*. Princeton University Press, Princeton, NJ.
- RAMME, W. 1929. Afrikanische Acrididae. Revisionen und Beschreibungen wenig bekannter und neuer Gattungen und Arten. *Mitteilungen aus dem Zoologischen Museum in Berlin* 15, 247–492.
- RAMME, W. 1941. Beiträge zur Kenntnis der Acrididen-Fauna des indomalayischen und benachbarter Gebiete (Orth.) mit besonderer Berücksichtigung der Tiergeographie von Celebes. *Mitteilungen aus dem Zoologischen Museum in Berlin* 25, 1–243.
- RAMME, W. 1951. Zur Systematik, Faunistik und Biologie der Orthopteren von Südost-Europa und Vorderasien. *Mitteilungen aus dem Zoologischen Museum in Berlin* 27, 1–432.
- RIEDE, K., NISCHK, F., DIETRICH, C., THIEL, C., AND SCHWENKER, F. 2006. Automated annotation of Orthoptera songs: First results from analysing the DORSA sound repository. *J. Orthoptera Resear.* 15, 1, 105–113.
- RDF. 2004. Resource Description Framework (RDF). <http://www.w3.org/RDF/>.
- RYAN, W. AND PITMAN, W. 1998. *Noah's Flood: The New Scientific Discoveries About the Event That Changed History*. Simon & Schuster, New York, NY.
- SANDERSON, E. W., JAITEH, M., LEVY, M. A., REDFORD, K. H., WANNEBO, A. V., AND WOOLMER, G. 2002. The human footprint and the last of the wild. *Bioscience* 52, 891–904.
- VAJDA, L. 2004. Wendehals und Liebeszauber. *Acta Antiqua Academiae Scientiarum Hungaricae* 44, 185–198.

Received December 2006; revised December 2007; accepted December 2007