Fragmenta Faunistica 53 (1): 97–115, 2010
PL ISSN 0015-9301© Museum and Institute of Zoology pas

Scope and applicability of information stored in natural history collections – a case study based on the largest museum collections of Tenebrionidae (Coleoptera) in Poland

Piotr Tykarski

Dept. of Ecology, Faculty of Biology, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland, e-mail: ptyk@poczta.onet.pl

Abstract: Collections of specimens in museums of natural history constitute very rich sources of information, containing data that can be analyzed in many ways and answer various types of questions. The aim of this article was to analyze the data content of three big collections of specimens of darkling beetles (Coleoptera, Tenebrionidae) and to show possible uses of different types of analysis with a specific focus to GIS-based methods. The source material was analyzed with regard to collections, species and collecting persons, using a number of aspects, both as summaries and spatial distribution analyses, such as counts of records, species, specimen localities, UTM squares, collection dates.

Key words: Coleoptera, Tenebrionidae, specimen labels, natural history collections, faunistics, spatial analysis, UTM squares, Poland

Introduction

Museums of natural history contain rich sources of biological data, providing material for several fields of scientific research, mostly connected with taxonomy, biogeography, and phylogeny. This is due to the very character of objects that they hold. Specimens of organisms, collected, preserved and deposited in a museum, are physical evidence of the reality of times when they were collected. If a collection is big enough, one can analyze various aspects of information it contains. It is possible under one condition: specimens themselves have to be well documented. The quality and detail level of data given on labels, concerning collecting events, determine feasibility of most of the possible future analyses connected with particular specimens.

Recently, during the last 10 years, the information content stored in museum collections has been given much more attention from science than ever before. We can observe a kind of renaissance of natural history collections, linked with and actually catalyzed by current progress in information technologies. It is so called biodiversity informatics that has increased interest of many scientists in museum collections and caused a change of their perception (Soberón & Peterson 2004, Johnson 2007). This is less and less the old-fashioned, traditional, almost non-scientific field of biological sciences; the modern approach has made it possible to understand multifaceted and complex picture of collections and scientific potential that they have (Krishtalka & Humphrey 2000).

Thanks to computer networks and distributed information systems, it is now possible to aggregate data from a vast number of datasets presenting primary information held in collections all over the world. If data available this way are of sufficient quality, it is possible e.g. to map species distributions or to create regional checklists. Every single specimen represented in a digital form can become a source of valuable information for certain types of analyses, provided that the quality requirements of its description are met (Chapman 2005).

The biggest network of biodiversity information – the Global Biodiversity Information Facility (GBIF) – is in fact, from technical point of view, an extension of several museum networks, sharing data about their specimens, using a common data format (Wheeler 2004). After some improvements, the original concepts and ideas turned out to be suitable for exchanging information about any primary biological records (e.g. observations, molecular data). At the moment, the total count of GBIF records based on museum specimens is ca. 80 million and can be used together with corresponding observation data.

In Poland, this kind of modern approach to resources of national museum institutions started in 2004, when the Polish Biodiversity Information Network (PolBIN, a Polish acronym *KSIB*) was created. The application of technologies developed by GBIF and sharing data with the global network enabled a number of Polish collections to be digitized and properly adjusted to present-day standards and requirements (Kozakiewicz & Tykarski 2007). However, despite PolBIN's achievements (ca. 1,5 million records available on-line, ca. 25% based on collections) this is the beginning of works that are necessary to develop an effective information system that would be useful for a wide range of scientific applications. Only a small percent of collections has been digitized so far. Beside the organizational challenges and necessary understanding of digitization advantages and sharing data by collection owners, the process itself is work and time-consuming.

The properly structured information unit should contain both data (species name, locality and date of the collection event) and metadata (gathering person, identifier, optionally verifier, collection method, etc.). Digitization of the verbatim information is usually only a starting point for a qualified electronic record especially in its geographic part. To be fully functional, it should be georeferenced, enabling a user to include it into geographic analyses (Chapman & Wieczorek 2006). These requirements are still rarely met in databases describing museum collections. Even when the material of a particular collection has all information elements in place, it happens that it is only partially digitized. Post-processing of records, especially adding georeferences, is also not a common practice.

The aim of this article is to show the scope of information contained in data derived from typical entomological collections, as a case study based on tenebrionid collections from three largest sources for this family in Poland. This includes different perspectives and examples of a number of uses of properly structured data.

Table 1. A schema of the data table and example original data entries from the source data to this article. When an entry contained two dates, it was transformed into two records; for single dates only one record was created.

Species	KFP region	Locality	UTM	Date I	Date II	Collector	Museum
Lagria hirta	Baltic Coast	Między-zdroje	VV67	12 VII 1966	23 VII 1966	AO	MIZ
Lagria hirta	Pomeranian Lake District	Bielinek nad Odrą	VU46	2 VIII 1985		PS	ISEA
Tenebrio molitor	Kraków-Wieluń Upland	Kraków-Zwierzyniec	DA14	22 VIII 1968	21 VII 1976	MG & MB	ISEA

METHODS

The data used in this analysis come from the article by IWAN et al. (2010), who described in detail methods used to create this dataset. It is based on the largest collections of Tenebrionidae from three institutions: Institute of Systematics and Evolution of Animals PAS, Cracow (ISEZ), Museum and Institute of Zoology PAS, Warsaw (MIZ) and Upper Silesian Museum, Bytom (USMB), where voucher specimens are deposited. Although the material is rich and carefully prepared, it does not contain full information about all analyzed specimens. In case of series of specimens from the same locality and different dates of collection, only edge dates of the series were noted (Table 1). Yet, lower time resolution caused by this simplification is not an obstacle and has little influence on the requirements of the analyses presented here. Where an original data entry was had two dates, it was transformed into two records, which had different dates and the rest of fields identical. In some cases, dates were absent as they were missing from the original specimen labels. Apart from the text description of a locality, every record in the analyzed material had UTM coordinates (grid 10×10 km) assigned. Additionally, each locality was assigned to a faunistic region, according to the system introduced in Catalogus Faunae Poloniae (KFP), further in the text referred to as a "KFP region". Although the system itself has many disadvantages and should be replaced with more adequate solutions, it is still in use in Polish entomological literature and, therefore, it should be possible to refer to it as a legacy system in future.

Maps were prepared using ESRI ArcGIS 9.3 in the PUWG 1992 coordinate system. All the maps use the same template, presenting boundaries of Poland, bigger towns (in darker gray), borders of KFP regions and the UTM grid: squares 10×10 km with dotted lines and squares 100×100 km with solid lines.

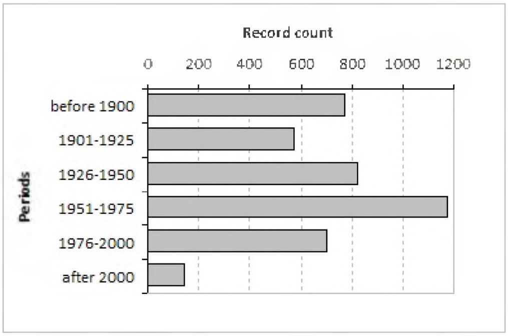

Fig. 1. Distribution of records in the time range of the analyzed material. The data from the XIX century are pooled; newer data divided into 25-year intervals.

RESULTS

General information about the data content

The source data comprise specimens of 78 species, collected from 1124 localities at 598 UTM squares by 177 collecting persons (Table 2). The oldest record dates back to

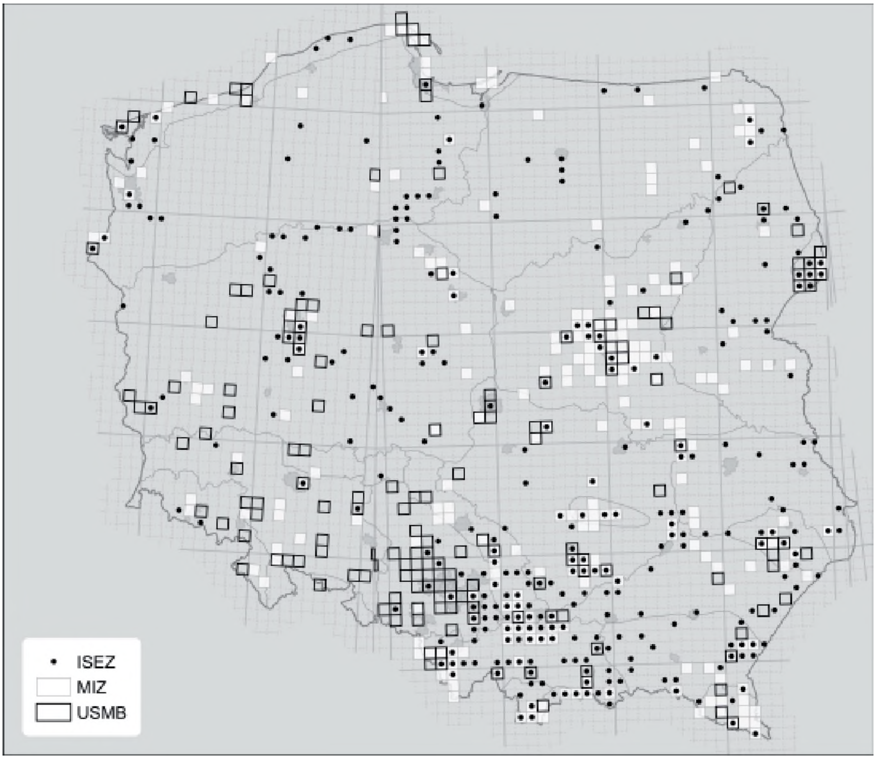

Fig. 2. Distribution of source localities for the three examined collections. Each collection is coded with a different colour.

1859 and the newest data were collected in 2010. A closer look at numbers of records in consecutive periods reveals the highest values in 1951-1975, and subsequent decrease (Fig. 1). If the number of records is proportional to the actual number of specimens, these values might be related to the rate of inflow of new specimens to the collections. The three maximal record numbers were noted in *Opatrum sabulosum*, *Lagria hirta*, and *Tenebrio molitor*. The oldest record is connected with *Isomira murina*, caught in 1859 (leg. A. Waga, ISEZ collection). *Isomira murina*, *Bolitophagus reticulatus*, *Corticeus unicolor*, and *Platydema violaceum* are species with longest ranges of record dates (the first one 151, others 135 years). The next part contains example analyses and summaries connected with GIS analyses of the data, divided for clarity into separate topics.

Comparison of the collections

Total numbers of species were comparable in each of the three examined collections (Table 2) – more than 60 in each case. There are remarkable differences in record counts, what is most probably connected with size of the collections. In comparison with MIZ and ISEA, the USMB tenebrionid collection is visibly more diverse, including similar species count from a much lower number of records and lower group of collectors. The geographical coverage of the collections is comparable and country-wide, although there is a noticeable bias in each one (Fig. 2).The USMB collection is more south-west oriented, ISEA – more south-east and the MIZ collection has more even distribution than the other two, being more dense around Warsaw. In many regions the collections overlap, in some of them suggesting a common history of collecting events or more attractive entomological collection places.

Table 2. Summary of data content in the analyzed material in respect to the three sources of information: ISEA, MIZ and USMB collections.

Museum	Total counts			
	Taxa	Collectors	Records	UTM squares
ISEA	64	83	1661	332
MIZ	66	84	1573	273
USMB	63	51	681	179

Spatial distribution of the data – general summaries

The analysis of record number per UTM square can be a kind of estimation of the research effort that had taken place in an area (Fig. 3). It is clear then which parts of the country receive more attention from coleopterists. According to this material this is Warsaw and its surroundings, Białowieża Forest, vicinities of Cracow, Pieniny Mts, Upper Silesia and Pogórze Przemyskie. There are also single smaller areas that were

Table 3. Summary of data content with respect to species in the source material.

Species name	Data time range			Total counts				Count of KFP regions					
	First record	Last record	Time range	Areas	UTM squares	Re-cords	Colle-ctors	Mate-rial	KFP	Com-mon	Poo-led	Unique for M.	Unique for KFP
Allecula morio (Fabricius, 1787)	1873	2005	133	58	48	92	23	15	16	14	17	1	2
Allecula rhenana Bach, 1859	1903	1975	73	11	10	18	2	6	7	6	7		1
Alphitobius diaperinus (Panzer, 1797)	1938	2003	66	8	7	7	7	6	5	2	9	4	3
Alphitobius laevigatus (Fabricius, 1781)	1880	1938	59	5	3	11	4	4	4	1	7	3	3
Alphitophagus bifasciatus (Say, 1824)	1972	1999	28	3	3	3	3	3	7	2	8	1	5
Blaps halophila Fischer von Waldheim, 1822	1965	1971	7	1	1	4	2	1	2	1	2		1
Blaps lethifera lethifera Marsham, 1802	1903	1931	29	3	3	3	2	3	9	3	9		6
Blaps mortisaga (Linnaeus, 1758)	1876	1994	119	63	60	92	29	15	13	9	19	6	4
Blaps mucronata Latreille, 1804	1935	1955	21	4	4	4	2	3	5	2	6	1	3
Bolitophagus interruptus Illiger, 1800	1911	1912	2	3	3	3	3	1	4	1	4		3
Bolitophagus reticulatus (Linnaeus, 1767)	1873	2007	135	86	62	150	34	19	14	12	21	7	2
Centorus elongatus elongatus (Herbst, 1797)	1900	1900	1	1	1	1	1	1			1	1	
Corticeus bicolor (Olivier, 1790)	1889	1997	109	62	39	114	28	12	18	11	19	1	7
Corticeus bicoloroides (Roubal, 1933)	1930	1932	3	1	1	2	2	1			1	1	
Corticeus fasciatus (Fabricius, 1790)	1889	1993	105	16	15	36	10	10	9	6	13	4	3
Corticeus fraxini (Kugelann, 1794)	1886	2004	119	15	11	18	11	7	10	5	12	2	5
Corticeus linearis (Fabricius, 1790)	1907	1989	83	11	11	18	8	8	12	6	14	2	6
Corticeus longulus (Gyllenhal, 1827)	1894	1994	101	14	13	18	10	6	4	3	7	3	1
Corticeus pini (Panzer, 1799)	1876	1971	96	11	10	11	6	7	9	4	12	3	5
Corticeus suberis (Lucas, 1846)	1938	1938	1	1	1	1	1	1	4		5	1	4
Corticeus suturalis (Paykull, 1800)	1921	2003	83	5	1	5	3	1	2	1	2		1
Corticeus unicolor Piller & Mitterpacher, 1783	1873	2007	135	72	51	131	26	15	20	14	21	1	6
Cryphaeus cornutus (Fischer von Waldheim, 1823)	1937	1937	1	1	1	1	1	1			1	1	
Crypticus quisquilius (Linnaeus, 1761)	1874	2004	131	147	120	207	52	17	14	14	17	3	

Ctenioplus sulphureus (Linnaeus, 1758)	1882	2009	128	56	43	80	34	14	12	10	16	4	2
Ctenioplus sulphuripes (Germar, 1824)	1920	1920	1	1	1	1	1	1	5	1	5		4
Diaperis boleti (Linnaeus, 1758)	1876	2003	128	90	76	162	36	18	17	12	23	6	5
Eledona agricola (Herbst, 1783)	1873	2003	131	55	49	95	27	15	15	11	19	4	4
Eledonoprius armatus (Panzer, 1799)	1884	1884	1	1	1	1	1	1	3		4	1	3
Gnatocerus cornutus (Fabricius, 1798)	1900	2000	101	8	8	15	8	7	6	4	9	3	2
Gonocephalum granulatum pusillum (Fabricius, 1792)	1896	1896	1	1	1	1	1	1			1	1	
Gonodera luperus luperus (Herbst, 1783)	1879	1999	121	22	21	32	15	13	14	11	16	2	3
Hymenalia rufipes (Fabricius, 1792)	1893	2003	111	26	22	33	17	10	15	8	17	2	7
Hymenophorus doublieri Mulsant, 1851	1959	1959	1	1	1	1	1	1	1		2	1	1
Isomira murina murina (Linnaeus, 1758)	1859	2009	151	157	105	223	36	18	17	14	21	4	3
Laena reitteri Weise, 1877	1876	1992	117	10	8	15	10	2	2	2	2		
Lagria atripes Mulsant & Guillebeau, 1855	1961	2003	43	9	9	12	10	7	3	2	8	5	1
Lagria hirta (Linnaeus, 1758)	1878	2009	132	201	153	280	65	23	22	22	23	1	
Latheticus oryzae Waterhouse, 1880	1938	1956	19	3	3	3	2	3			3	3	
Melanimon tibialis tibialis (Fabricius, 1781)	1874	2007	134	125	98	189	48	16	12	12	16	4	
Mycetocharis axillaris (Paykull, 1799)	1884	2004	121	51	39	80	12	14	14	12	16	2	2
Mycetocharis flavipes (Fabricius, 1792)	1879	1999	121	62	46	104	16	12	16	11	17	1	5
Mycetocharis humeralis (Fabricius, 1787)	1888	1991	104	22	19	35	11	9	9	6	12	3	3
Mycetocharis maura (Fabricius, 1792)	1879	1999	121	30	29	48	13	10	14	10	14		4
Mycetocharis pygmaea (Redtenbacher, 1874)	1904	1904	1	1	1	1	1	1	4	1	4		3
Myrmechixenus subterraneus Chevrolat, 1835	1878	1981	104	13	12	18	6	8	12	5	15	3	7
Nalassus convexus (Comolli, 1837)	1884	1896	13	2	2	2	1	1			1	1	
Nalassus dermestoides (Illiger, 1798)	1874	2006	133	86	69	148	38	17	13	12	18	5	1
Nalassus laevioctostriatus (Goeze, 1777)	1937	1982	46	2	2	2	2	1	6		7	1	6
Neatus picipes (Herbst, 1797)	1888	1995	108	19	13	52	15	5	9	5	9		4
Neomida haemorrhoidalis (Fabricius, 1787)	1878	2003	126	23	21	44	15	11	8	7	12	4	1
Omophlus pubescens (Linnaeus, 1758)	1874	2003	130	45	42	67	26	13	17	12	18	1	5
Oodescelis melas (Fischer von Waldheim, 1823)	1916	1918	3	2	2	2	2	2	3		5	2	3

Species name	Data time range			Total counts				Count of KFP regions					
	First record	Last record	Time range	Areas	UTM squares	Re-cords	Colle-ctors	Mate-rial	KFP	Com-mon	Poo-led	Unique for M.	Unique for KFP
Opatrum riparium Scriba, 1865	1878	1994	117	23	18	34	13	9	11	6	14	3	5
Opatrum sabulosum sabulosum (Linnaeus, 1761)	1878	2003	126	176	131	327	45	19	18	17	20	2	1
Palorus depressus (Fabricius, 1790)	1885	2007	123	7	6	24	5	3	8	2	9	1	6
Palorus ratzeburgii (Wissmann, 1848)	1936	1938	3	2	2	2	1	2	6	2	6		4
Palorus subdepressus Wollaston, 1864	1953	1981	29	2	2	2	2	2	1	1	2	1	
Pedinus femoralis (Linnaeus, 1767)	1880	1996	117	24	18	64	22	6	4	3	7	3	1
Pentaphyllus testaceus (Hellwig, 1792)	1879	1999	121	22	20	47	12	8	11	8	11		3
Phylan gibbus (Fabricius, 1775)	1912	1997	86	21	20	44	16	2	3	1	4	1	2
Platydema dejeanii Laporte & Brullé, 1831	1879	1999	121	26	13	34	12	6	5	5	6	1	
Platydema violaceum (Fabricius, 1790)	1872	2006	135	21	19	25	16	10	12	7	15	3	5
Podonta nigrita (Fabricius, 1794)	1896	1900	5	3	2	3	1	1	2		3	1	2
Prionychus ater (Fabricius, 1775)	1878	2006	129	63	52	101	29	13	17	13	17		4
Prionychus melanarius (Germar, 1813)	1890	1971	82	21	17	29	9	9	9	8	10	1	1
Pseudocistela ceramboides ceramboides (Linnaeus, 1761)	1879	2004	126	73	64	97	23	18	21	16	23	2	5
Scaphidema metallicum metallicum (Fabricius, 1792)	1876	2004	129	91	61	133	37	16	14	11	19	5	3
Stenomax aeneus (Scopoli, 1763)	1879	2003	125	76	60	121	31	16	10	9	17	7	1
Tenebrio molitor Linnaeus, 1758	1878	2010	125	139	111	254	63	22	23	22	23		1
Tenebrio obscurus Fabricius, 1792	1879	1965	87	16	15	32	9	10	13	8	15	2	5
Tenebrio opacus Duftschmid, 1812	1937	1999	63	7	7	14	6	3	3	1	5	2	2
Tribolium castaneum (Herbst, 1797)	1878	1995	118	16	15	21	14	9	11	5	15	4	6
Tribolium confusum Jacquelin du Val, 1862	1899	1972	74	8	7	11	5	7	12	5	14	2	7
Tribolium destructor Uyttenboogaart, 1933	1956	2000	45	9	9	17	12	8	6	3	11	5	3
Tribolium madens (Charpentier, 1825)	1903	1989	87	7	7	13	6	5	9	3	11	2	6
Uloma culinaris (Linnaeus, 1758)	1879	2007	129	45	32	81	25	11	13	9	15	2	4
Uloma rufa (Piller & Mitterpacher, 1783)	1878	2008	131	72	60	99	35	17	12	11	18	6	1

evidently better investigated than others (e.g. Bielinek Reserve at the western border, or Zamość area in the south-east).

Similar conclusions may be drawn from the analysis of species numbers recorded from UTM squares (Fig. 4). In this case one can see distinct "hot-spots", showing much higher species richness than in other places. Without doubts, this is connected with higher sampling effort, mentioned earlier. There are a few areas in the analyzed material, showing more than 20 tenebrionid species: Warsaw (DC99, EC19) and Cracow (DA14, DA24) surroundings, Białowieża (FD94), Bielinek reserve, Zamość vicinities (FB30 between Szczebrzeszyn, Biłogoraj and Krasnobród), Piekary Śląskie (CA58).

Both approaches suggest also that most of the country is poorly known – there are no records in none of the three biggest tenebrionid collections, and a high proportion of existing data come from a single collecting event (=a single record).

Species statistics and examples of spatial analyses

Basic statistics for all the species are summarized in Table 3. The three species recorded from the maximum number of localities/UTM squares are *Lagria hirta*, *Opatrum sabulosum*, and *Isomira murina* (Table 3). In the analyzed material *Allecula morio*, *Bolitophagus reticulatus*, *Corticeus unicolor*, *Crypticus quisquilius*, *Eledona agricola*, *Isomira murina*, *Lagria hirta*, *Melanimon tibialis*, *Nalassus dermestoides*, *Platydema violaceum*, *Tenebrio molitor* and *Uloma rufa* have the longest history of research – all of them have been recorded for more than 130 years. 5 species: *Crypticus quisquilius*, *Lagria hirta*, *Melanimon tibialis*, *Opatrum sabulosum*, and *Tenebrio molitor* were collected by more than 40 persons.

Mapping species distributions and spatial analyses may help to summarize current knowledge or even to plan areas and topics for the future research. Here, maps were used to assess representativeness of data about general distribution of species in the analyzed material by comparing them with the respective data from KFP. The analysis of overlapping of species distribution at the level of KFP regions showed four patterns in the compared data. They are presented below together with example species and criteria used to distinguish them:

• high similarity: the analyzed distribution of a species fits well or exactly to the compared distribution in KFP (Fig. 5). Criterion: high proportion of overlapping regions to the total number of regions.

• general low similarity: the analyzed distribution of a species is remarkably different from the compared distribution in KFP (Fig. 6). Criterion: low proportion of overlapping regions to the total number of regions.

• low similarity with high coverage: the distributions differ, but the analyzed coverage is wider than the compared coverage in KFP (Fig. 7). Criterion: high proportion of regions unique for the analyzed distribution to the total number of regions.

• low similarity with low coverage: the distributions differ, but the analyzed coverage is smaller than the compared coverage in KFP (Fig. 8). Criterion: high proportion of regions unique for the KFP distribution to the total number of regions.

Intensity of exploration and occurrence data validity

Colour-coding of collection dates and plotting them on a UTM grid has several uses. A view on dates of the first records from an area (Fig. 10) shows the beginning of the history of research for sites studied over a longer period on one hand, and the areas that have been only recently explored and may need further research, on the other. The analyzed data clearly show areas of Warsaw, Cracow, Upper Silesia, Beskidy, Tatra Mts. and the south-eastern Poland, as well as Białowieża and Szczecin as the first regions that have been explored with respect to the history of knowledge of tenebrionid fauna. An opposite type of observation may be made looking at the regions of Pomorskie and Mazurskie Lakelands – many areas there were investigated very recently.

Unlike in in the previous approach, maps showing dates of the last records (Figs 5-8) can be used to assess for which areas the current data exist, and where information on species occurrence is outdated. The same kind of coding in respect to the pooled data (Fig. 9) revealed areas of the country where the last research (in regard to the analyzed material, of course) took place in the XIX century and no further data exist. Amazingly, there are more than 20 such UTM squares, located also in regions with the longest history of entomological research – near Warsaw, Cracow, Zamość and in Silesia.

When the distribution of the first records may be compared to a bottom surface of the "cloud of information", the distribution of the last records corresponds to its upper surface. Mapping time ranges between the first and the last record from (Fig. 11) would then refer to thickness of the cloud. It may be an estimate of the research effort at sites of interest that can help to answer questions about completeness of knowledge about a particular area.

Activity of the collectors

The whole source material comprised specimens collected by 178 persons (Table 4 shows statistics for only top 30 persons contributing most of the records). The biggest number of records in the analyzed material comes from specimens collected by B. Burakowski (710), S. Stobiecki (363) and Sz. Tenenbaum (236). Burakowski and Tenenbaum, as well as W. Mączyński contributed most in terms of number of species in the collections (respectively 51, 41, 45). The maximum number of source localities included in materials of a single collector was found also for the first two persons and D. Kubisz (144, 72, 71 UTM squares respectively). Most of the collectors had operated in certain regions of the country, or at least it is often easy to show the area of their most intense collection activities, as in cases of B. Kotula or F. Kirsch (Fig. 12). There was also a group of collectors that acquired specimens from numerous localities, scattered all around Poland; the most distinct case represented by B. Burakowski (Fig. 13).

Generally, summarized plot of activities of collectors (Fig. 14) gives a similar view to the formerly presented maps that may be treated as a kind of estimates of research effort (Figs 3, 4, 11). In this case there is maybe the most straightforward relation between attractiveness of an area (whatever was its reason) and the number of collecting visitors.

Table 4. Basic statistics for collectors of specimens in the analyzed material. Only collectors with more than 20 records are shown.

Collector	First year	Last year	Total counts			
			Species	Records	Areas	UTM squares
Bartoszyński A.	1927	1939	30	77	21	19
Bielawski R.	1947	1973	21	74	54	45
Buchholz L.	1974	1999	18	30	25	21
Bunalski M.	1983	1994	19	33	22	20
Burakowski B.	1930	1997	51	710	258	144
Ciszkiewicz H.	1921	1939	20	55	24	19
Dobosz R.	1989	2007	22	45	25	20
Eichler W.	1911	1963	19	37	16	15
Fejfer F.	1905	1941	22	38	10	7
Gutowski J.M.	1975	2004	26	42	29	24
Iwan D.	1981	1993	20	42	13	12
Jędryczkowski W.	1984	2009	14	34	21	16
Kirsch F.	1913	1942	30	88	26	23
Kotula B.	1872	1898	31	157	46	21
Kubisz D.	1983	2009	38	161	107	71
Mączyński W.	1884	1982	45	185	29	25
Makólski J.	1910	1953	15	39	19	17
Mazur E.	1912	1941	21	72	49	27
Mazur M.	1979	2003	13	61	57	42
Melke A.	1981	1991	15	26	17	14
Młynarski J.K.	1982	1991	14	30	23	14
Mroczkowski M.	1948	1962	17	31	22	20
Nowotny H.	1925	1938	21	41	21	18
Pawłowski J.	1959	2004	21	58	42	34
Popek S.	1907	1939	19	59	26	10
Rybiński M.	1876	1904	24	30	21	6
Stachowiak P.	1974	1987	14	24	13	10
Stobiecki S.	1876	1942	39	363	132	72
Szymczakowski W.	1947	1972	15	127	54	36
Tenenbaum Sz.	1892	1940	41	236	85	46

Conclusions and prospects

The purpose of this article was to give a quantitative summary of the source material on one hand and to explore possible uses of spatial analysis applied to this type of data. Both approaches are supplementary and to some extent overlap, although the same data presented as a map are often much easier to analyze and therefore more informative. A part of the calculations could not be done quickly without using GIS techniques. However, data for such analyses have to be properly structured and cleaned, which is itself a time consuming process. In addition to this somehow trivial statement it is worth to add another one: the more the data, the more useful the effects of an analysis are.

Despite of the long tradition of entomological research in Poland and lots of specimens in numerous collections of many institutions, the modern type of faunistic analyses has not existed so far. The current article is probably the first case of this kind of integrated analysis. It was possible thanks to cooperation of the scientific institutions that initiated compiling data on Polish tenebrionids presented in the source article (Iwan et al. 2010), accompanied by recent achievements of the PolBIN and its integrative initiatives. The source data are planned to become a part of the database of Coleoptera Poloniae programme (coleoptera.ksib.pl), being realized by PolBIN members and aiming at creating an information system on Coleoptera of Poland. The more such datasets will be available, the more complete knowledge of species will be, making results of such analyses more representative and useful.

References

Burakowski B., Mroczkowski M., Stefańska J. 1987. Chrząszcze Coleoptera. Cucujoidea, cz. 3. Katalog fauny Polski 23 (14): 1-309.

Burakowski B., Mroczkowski M., Stefańska J. 2000. Uzupełnienia tomów 2-21. Katalog fauny Polski 23 (22): 1-252.

Chapman A.D. 2005. Principles of data quality, version 1.0 Copenhagen: Global Biodiversity Information Facility.

Chapman A.D., Wieczorek J. (eds.) 2006. BioGeomancer Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility. 90 pp.

Iwan D., Kubisz D., Mazur M.A. 2010. The occurrence of Tenebrionidae (Coleoptera) in Poland based on the largest national museum collections. Fragmenta Faunistica 53(1): 1-95.

Johnson N.F. 2007. Biodiversity informatics. Annu. Rev. Entomol. 52:421-438.

Kozakiewicz M., Tykarski P. 2007. Różnorodność życia na Ziemi - wspólna troska, wspólne działanie, czyli o światowej sieci informacji o bioróżnorodności GBIF i jej polskiej części – KSIB. Wszechświat 108(7-9):172-176.

Krishtalka L., Humphrey P. 2000. Can natural history museums capture the future? BioScience 50: 611-617.

Soberón J., A.T. Peterson. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philos. Trans. R. Soc. London B 359:689-698.

Wheeler Q.D. 2004. What is GBIF? Bioscience 54(8): 717.

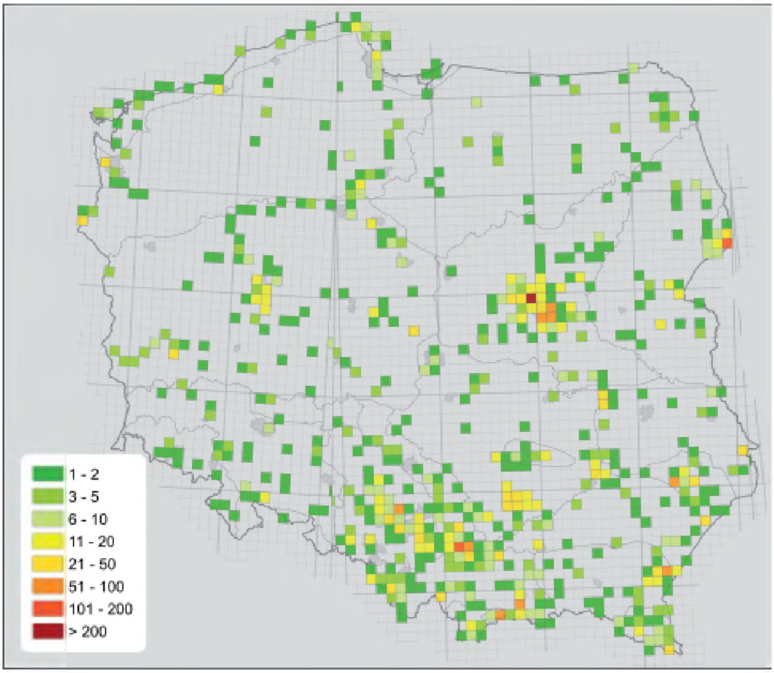

Fig. 3. Numbers of records from UTM squares, divided into classes coded with different colours.

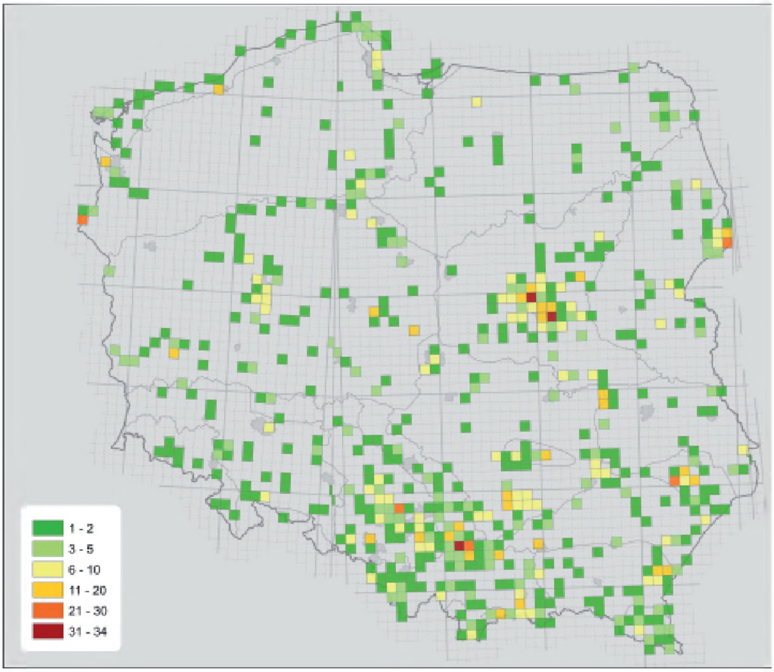

Fig. 4. Numbers of species recorded from UTM squares, divided into classes coded with different colours

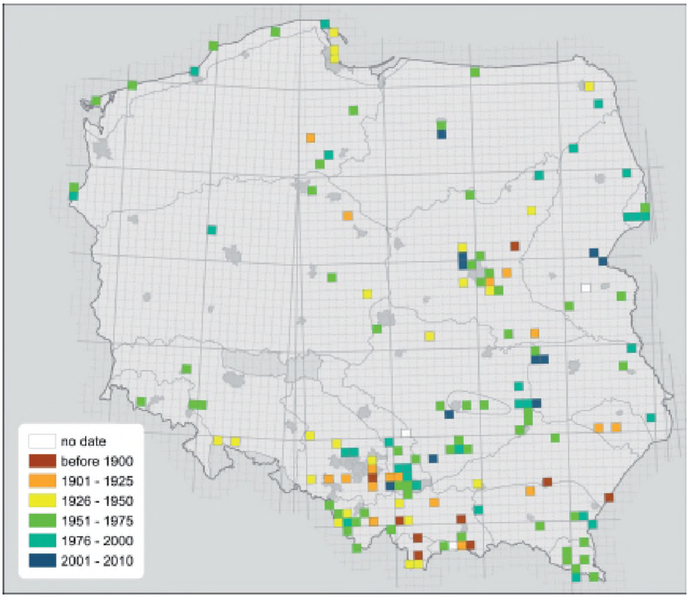

Fig. 5. Distribution of *Lagria hirta* – an example of a high similarity of general distribution information to the distribution data in KFP: the species was recorded in similar number of regions in the examined material and in the KFP. The colours code the year of the last record in a UTM square. Regions in light-grey were noted in KFP.

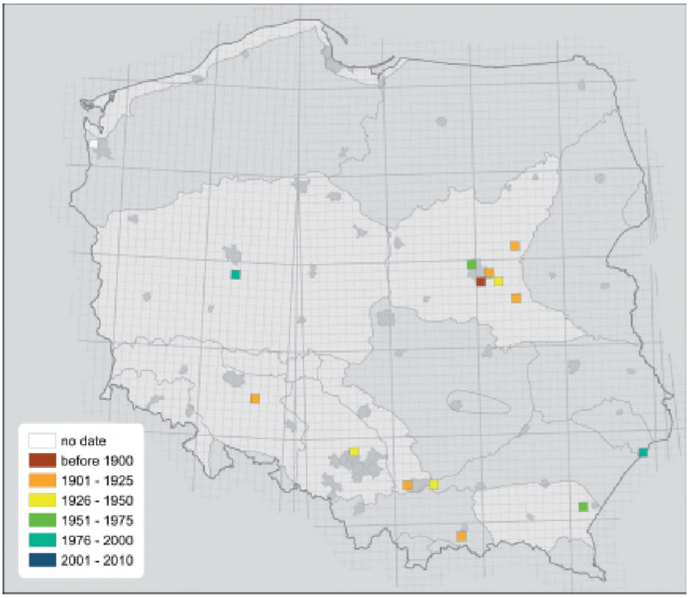

Fig. 6. Distribution of *Corticeus fasciatus* – an example of a low similarity of general distribution information to the distribution data in KFP: the species was recorded in different regions in the examined material and in the KFP. The colours code the year of the last record in a UTM square. Regions in light-grey were noted in KFP.

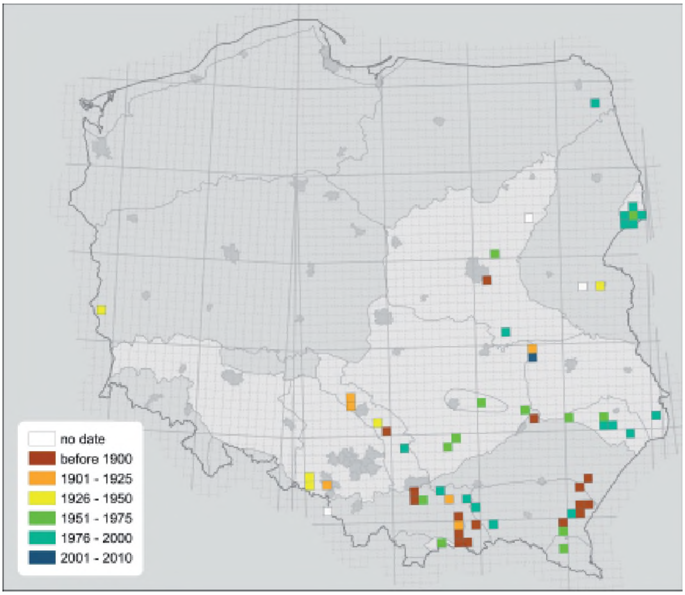

Fig. 7. Distribution of *Stenomax aeneus* – an example of a low similarity of a distribution in the analyzed material to the respective KFP data, with good coverage: the species was recorded in many regions that were not noted in KFP. The colours code the year of the last record in a UTM square. Regions in light-grey were noted in KFP.

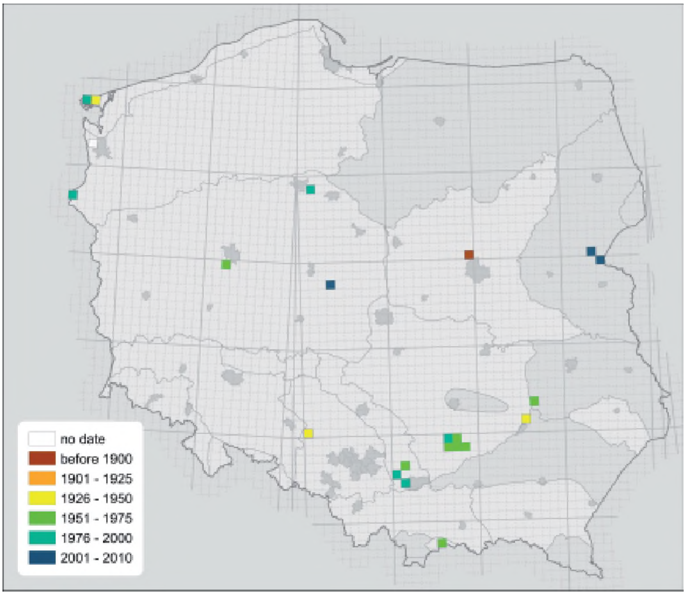

Fig. 8. Distribution of *Hymenalia rufipes* – an example of a low similarity of a distribution in the analyzed material to the respective KFP data, with low coverage: the species was recorded in many regions in KFP for which there are no data in the examined material. The colours code the year of the last record in a UTM square. Regions in light-grey were noted in KFP.

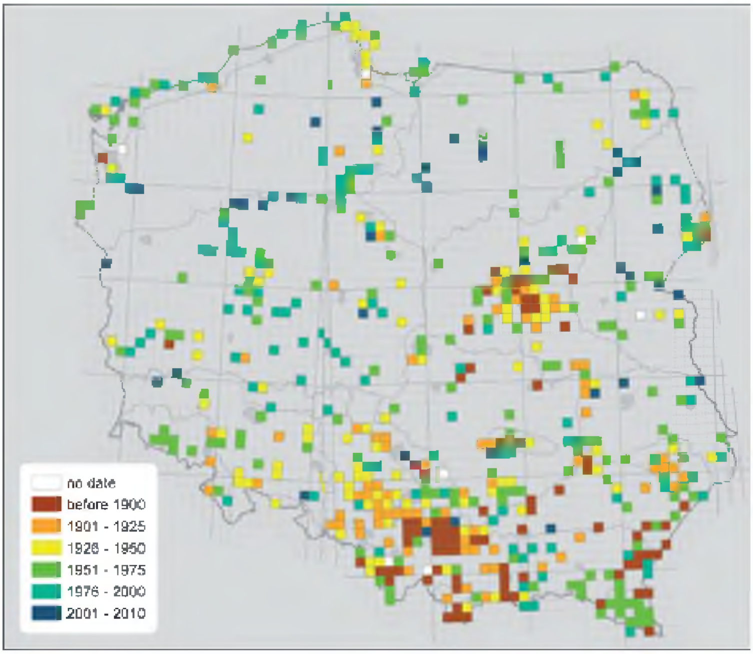

Fig. 9. Beginning of exploration at UTM squares of the source data. The colours code the year of the first record in a square.

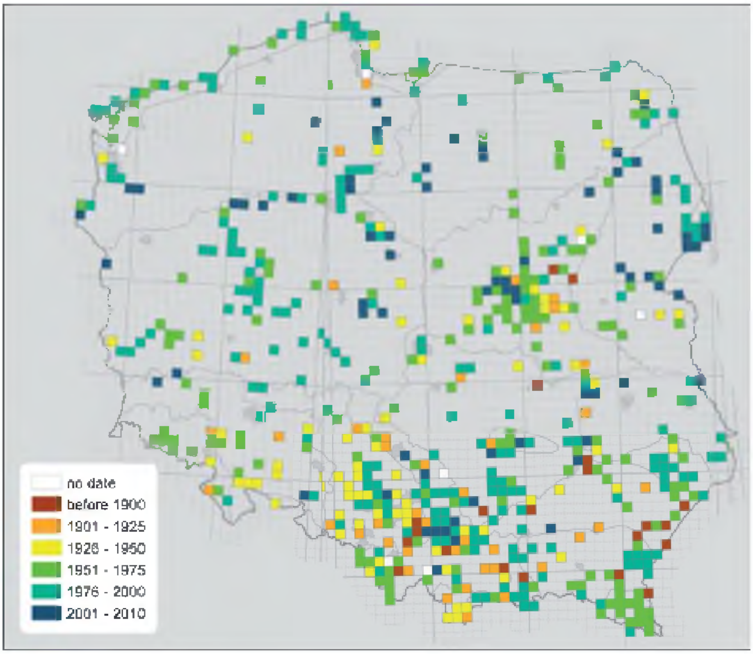

Fig. 10. Recent research at UTM squares of the source data. The colours code the year of the last record in a square.

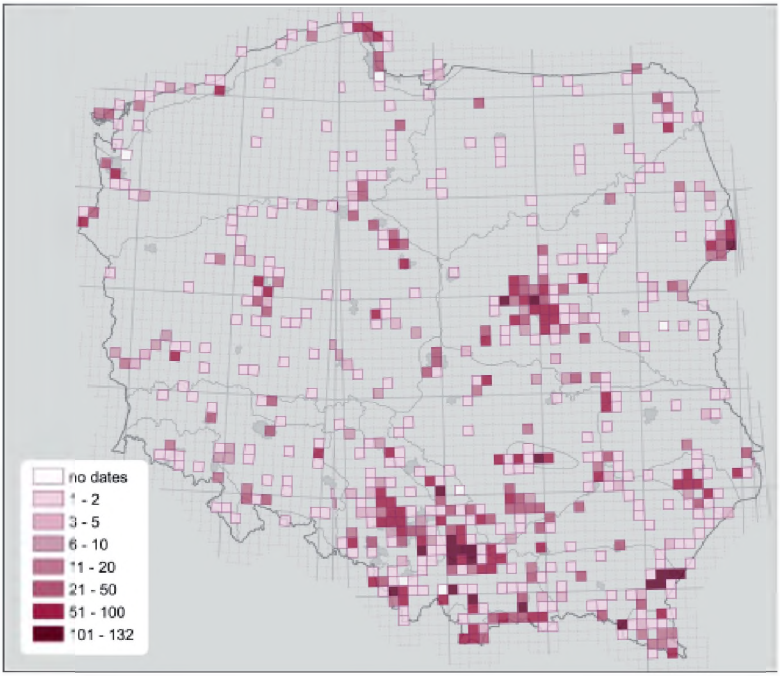

Fig. 11. Time range of research at UTM squares according to the source data – the period from a year of the first record to a year of the last record. The colours code classes of length of research.

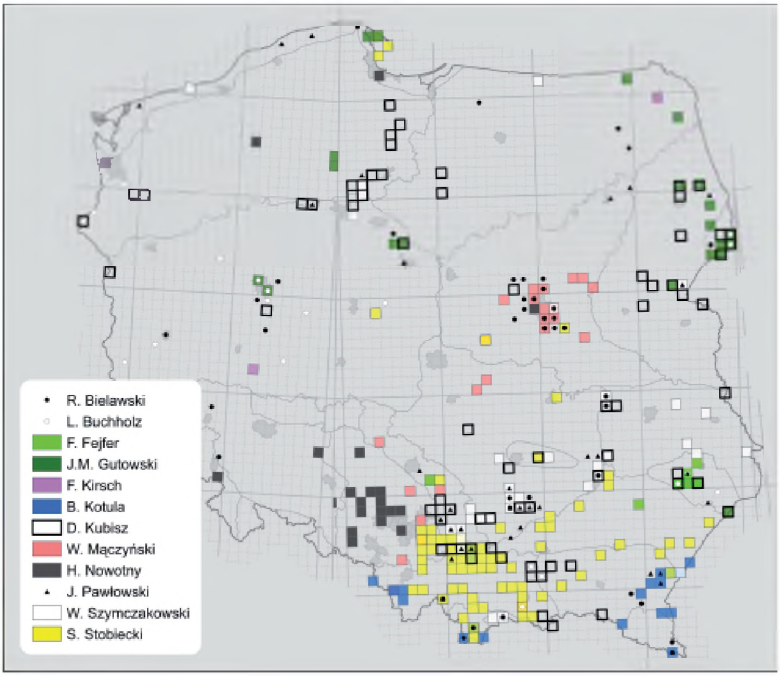

Fig. 12. Distribution of areas of activity of selected collectors – part 1.

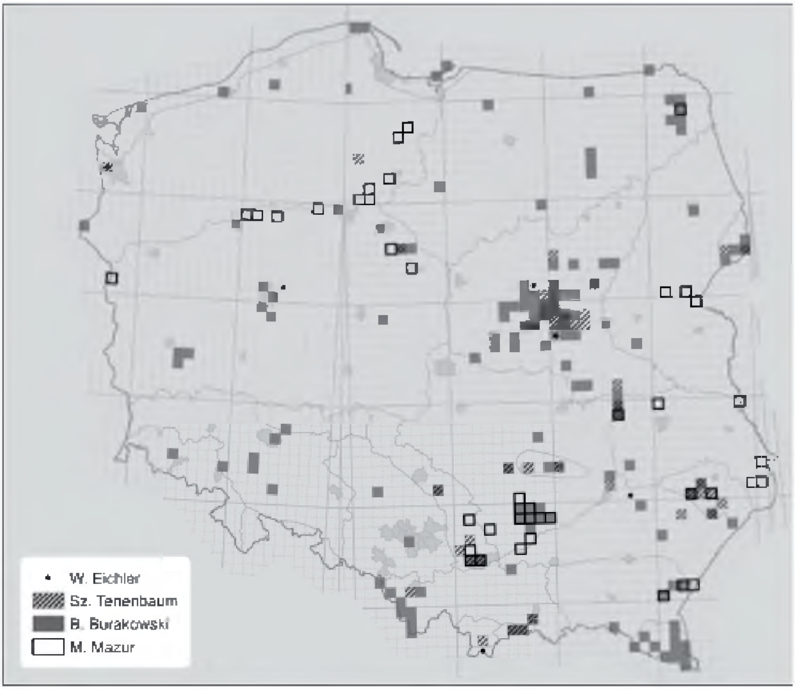

Fig. 13. Distribution of areas of activity of selected collectors – part 2.

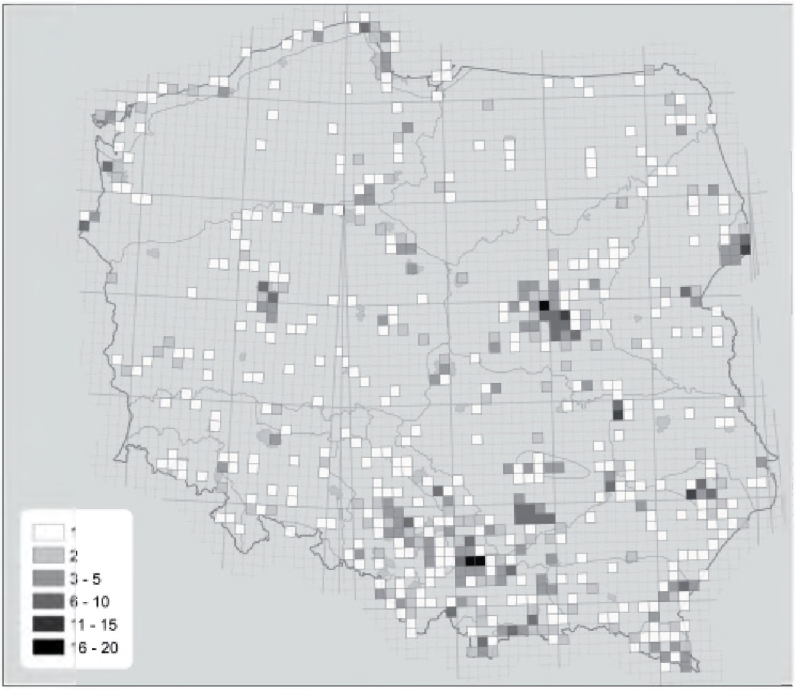

Fig. 14. Numbers of collectors active at UTM squares, divided into classes coded with different colours.

Streszczenie

[Zakres i możliwości wykorzystania informacji zawartych w kolekcjach przyrodniczych – analiza na przykładzie największych polskich muzealnych kolekcji Tenebrionidae (Coleoptera)]

Kolekcje okazów w muzeach przyrodniczych stanowią bogate źródło informacji i zawierają dane, które mogą znaleźć rozmaite zastosowania i być podstawą różnego typu analiz. Celem tego artykułu było zademonstrowanie możliwości wykorzystania potencjału tkwiącego w tego typu zbiorach informacji poprzez analizę zawartości danych z trzech kolekcji czarnuchowatych (Coleoptera, Tenebrionidae), ze szczególnym uwzględnieniem korzyści zastosowania analiz przestrzennych. Analiza materiału źródłowego objęła podsumowania i zestawienia danych względem kolekcji (instytucji), gatunków i osób zbieraczy, dotyczące liczby rekordów, gatunków, lokalizacji, kwadratów UTM i dat pochodzenia okazów. Mapy ukazujące rozkład badanych wartości w różnych aspektach analizy okazały się bardzo przydatne jako syntetyczny sposób prezentacji danych, często dużo wygodniejszy niż tradycyjna forma tabelarycznych zestawień. Ponadto bez wykorzystania technik GIS część obliczeń nie mogłaby zostać wykonana, przynajmniej w sposób umożliwiający uzyskanie wyników w sensownym czasie.

Jest to najprawdopodobniej pierwsza w kraju tego typu analiza materiału faunistycznego, wykorzystująca doświadczenia i metodologię wypracowaną w ostatnich latach w dziedzinie „biodiversity informatics". Jest to obszar na styku nauk przyrodniczych i informatyki, w którym wykorzystuje się możliwości, jakie dają współczesne technologie informatyczne, do analiz danych przyrodniczych, między innymi tych uznawanych za tradycyjne i nienowoczesne. Techniki te stały się dostępne w Polsce m.in. poprzez rozwój międzynarodowej sieci GBIF i działania Krajowej Sieci Informacji o Bioróżnorodności. Niezależnie od aspektu technicznego tego typu analiz i związanych z tym wymagań co do jakości danych, podstawowym warunkiem zaistnienia samej możliwości ich przeprowadzenia jest zgromadzenie w jednym miejscu dużej ilości odpowiednich danych. Wymaga to efektywnej współpracy, zwykle na poziomie międzyinstytucjonalnym, takiej jaką zaprezentowali autorzy opracowania źródłowego dla niniejszego artykułu.

Dane z trzech omawianych kolekcji staną się elementem większego systemu informatycznego – bazy Coleoptera Poloniae – zaplanowanego na długie działanie programu KSIB, mającego za cel utworzenie pełnego źródła danych o krajowej faunie Coleoptera. Gdy dostępnych będzie więcej tego typu zbiorów danych, wyniki analiz rozmieszczenia gatunków będą bardziej reprezentatywne i znajdą szereg rzeczywistych zastosowań, niemożliwych do zrealizowania w tej chwili.

Accepted: 30 November 2010