NEAFWA RCN Grant #2009-11

2010 4th Quarter Progress Report
1-Oct-2010 — 31-Dec-2010

Development of an Online Database to Enhance the Conservation of SGCN Invertebrates in the Northeastern Region

Dr. James W. Fetzner Jr.
Assistant Curator of Crustacea
FetznerJ@CarnegieMNH.Org

and

Dr. John E. Rawlins
Associate Curator, Head of Section
RawlinsJ@CarnegieMNH.Org

Section of Invertebrate Zoology
Carnegie Museum of Natural History
4400 Forbes Avenue, Pittsburgh, PA 15213-4007

Phone: (412) 622-3259
Fax: (412) 688-8670

January 25, 2011

PROJECT GOALS & OBJECTIVES:

We propose to develop a web-accessible database with corresponding administrative and reporting tools to fill the current data void that exists for invertebrates in the Northeastern Region. The database, as envisioned, will be able to handle data that encompasses all invertebrate taxa. The database schema and tables will be designed for real-world, museum-specimen data, thus allowing individual researchers or institutions to upload their data into a standardized database that will accommodate multiple invertebrate taxonomic lineages. For practical reasons, the diverse lineages represented by the expertise of PI's and staff at CMNH will be emphasized: Crustacea, Lepidoptera, Coleoptera, Diptera, and to a lesser extent Hymenoptera, Hemiptera, Odonata, and other smaller insect orders.

In addition, CMNH will authoritatively identify, database and contribute the first 10,000 data records on some of the ~1,700 SCGN invertebrates listed in the Northeastern Region state wildlife action plans as part of the testing effort and formative evaluative process for this web-based system. These initial specimen records are only a small portion of the holdings currently residing at CMNH, which has significant and diverse collections of SGCN (and other) taxa from throughout the Northeastern Region, the rest of North America, and indeed, from most major bioregions worldwide.

The PI's and CMNH are well-suited for this kind of large-scale databasing, whether it be involvement with the Section of Invertebrate Zoology's BSF program (Biodiversity Services Facility) for USDA APHIS high volume sample screening, for major biotic inventories conducted by CMNH in recent years (Hispaniola especially), or taxon-specific websites for groups like crayfish, crane flies, and other groups. Most institutional collections of invertebrates in the Northeastern Region have SGCN specimens and some of these will be contacted and asked to supply limited specimen data specifically for testing outside data access tools, data conversion, and uploading. We will work diligently with all contributing institutions to build tools that will allow for rapid and efficient upload of specimen data in keeping with collection-specific requirements, data use policies, and concerns for sensitive data.

SUMMARY OF PROGRESS TO DATE (Q1 – Q4):

First Quarter (Q1)

The start date for this project was set as 1-Feb-2010; however, final signed contracts were not received and in place until 10-Feb-2010. So far, we have compiled a list of SGCN taxa from each state wildlife action plan (SWAP) in the Northeastern Region, and this list consists of roughly 2,900 entries. However, many of these were duplicated across multiple states, so this initial list had to be sorted and reduced to a list of unique taxon entries. After this process, there were roughly 1,700 invertebrate species in the list for the Northeastern Region. To arrive at a final list, several taxonomic issues still need to be resolved. These include resolving synonymies and/or updating taxonomic classifications (i.e., for species that have been revised and placed in other genera, etc.). In the meantime, we have also been working on finalizing the database field structure (and associated metadata) and tables that will be used as the core of the SGCN invertebrate website. This field structure should be finalized in Q2, or early Q3.

	Taxon	Counts
1	Amphipoda	49
2	Arachnida	75
3	Chilopoda	4
4	**Coleoptera**	**215**
5	Collembola	26
6	Copepoda	2
7	Crustacea	57
8	Diplopoda	75
9	Diplura	3
10	Diptera	16
11	Echinodermata	1
12	Ephemeroptera	57
13	Hemiptera	19
14	Hirudinea	1
15	Hymenoptera	1
16	Isopoda	29
17	**Lepidoptera**	**477**
18	Mecoptera	2
19	**Mollusca**	**299**
20	Neuroptera	2
21	**Odonata**	**202**
22	Oligochaeta	6
23	Orthoptera	2
24	Ostracoda	1
25	Platyhelminthes	16
26	Plecoptera	49
27	Porifera	4
28	Trichoptera	16
	TOTAL	**1706**

Figure 1. Pie chart showing the distribution of the 1,706 SGCN species by taxonomic group. Also depicted is a list and count of species of greatest conservation need by taxonomic group (i.e., Order) found in state wildlife action plans (SWAPs) in the northeastern region.

Second Quarter (Q2)

SGCN Taxon List.— Above is the summary of the taxa listed on state wildlife action plans (SWAPs) in the northeast region (Figure 1). We ended up with a total of 1,706 species, however, there are still some taxonomic issues to investigate, and state heritage programs still need to be contacted to see if they can verify the list for their state. Thus, this list may still change slightly. However, the data shown above gives a quick summary of the species on the lists by taxonomic group. Four groups make up the majority (= 71%) of the species listed by states in the northeast region and include the Lepidoptera, Mollusca, Coleoptera and Odonata.

Databasing.— In Q2, databasing of pinned specimens from the CMNH collection was initiated in order to generate records for testing the various features of the soon-to-be built SGCN website. These specimens represent SGCN invertebrates listed in the state wildlife action plans in the northeastern region. Thus far, specimen label data has been captured from 3,632 specimens (= 36% of total) representing 50 different SGCN beetle species. Additional taxa (and insect Orders) will be captured in Q3.

Database Metadata for Specimens.— We are very close to finalizing the database schema and table field structure for the specimen data and the website. Once completed, work will begin on designing and implementing the SGCN website (See Appendix 1 for current structure of the specimens table).

Third Quarter (Q3)

Databasing.— In Q3, databasing of pinned specimens from the CMNH insect collection continued in order to generate the test dataset that will be used as the initial backbone of the SGCN website. These specimens represent a variety of SGCN invertebrates (insects) that are listed in various state wildlife action plans from the northeastern region. In Q3, specimen label data has been captured from an additional 6,168 specimens (9,800 total = 98% of total proposed), and these represent 92 different SGCN taxa. Additional taxa from other insect Orders will be captured in Q4.

Website Construction.-- At the end of Q3, the construction of the SGCN invertebrate website was initiated, with web page layout and design being the primary goals. The addition of tools and other functionality, such as connecting to the test database, will commence in Q4 once page layouts are finalized and the database records have been proofed and fully geo-referenced.

DIFFICULTIES ENCOUNTERED QUARTER 3: Some of the website tool development has been delayed due to additional time needed for retrospective geo-referencing of specimen records. These website tools should be developed in Q4.

Fourth Quarter (Q4)

Databasing.— In Q4, databasing of pinned specimens from the CMNH insect collection continued, with label data being captured from an additional 7,537 specimens, for a total of 17,337. These specimens represent a total of 113 SGCN taxa, with the majority being beetles (Coleoptera), with a few flies (Diptera) and butterflies (Lepidoptera) mixed in. See Table 1 below for a list of taxa and an associated count of specimens for each. Additional taxa may be added as time permits.

A significant proportion of specimens, especially ones collected prior to the 1990s (pre-GPS), contain associated labels with abbreviated, incomplete or missing locality data (i.e., no Latitude-Longitude coordinates, minimal or ambiguous data, etc.), which makes these specimens difficult to retrospectively geo-reference without some anguish and a significant outlay of time, especially when compared to specimens collected within the last 10-15 years and that have full locality and coordinate data associated with them. However, a majority of the specimens in the CMNH collection are from this earlier time period, so checking, correcting and updating the specimens records has been slow and has thus delayed the creation of the final test database that will be used to demonstrate the website functionalities.

Table 1. List of 113 SGCN taxa, and the number of specimens of each, that have been databased thru Q4.

ID	Genus	Subgenus	Species	Subname	Count(*)
1	*Acupalpus*	*Philodes*	*alternans*		16
2	*Acupalpus*		*rectangulus*		72
3	*Agonum*		*crenistriatum*		57
4	*Agonum*		*darlingtoni*		98
5	*Agonum*		*decorum*		187
6	*Agonum*		*moerens*		77
7	*Agonum*		*mutatum*		382
8	*Agonum*		*picicornoides*		4
9	*Agonum*		*punctiforme*		2140
10	*Agonum*		*superioris*		2
11	*Amara*	*Celia*	*chalcea*		11
12	*Amara*		*erratica*		5
13	*Anaedus*		*brunneus*		246
14	*Apristus*		*latens*		57
15	*Atranus*		*pubescens*		70
16	*Atylotus*		*ohioensis*		4
17	*Badister*	*Trimorphus*	*transversus*		26
18	*Bembidion*	*Furcacampa*	*affine*		1191
19	*Bembidion*	*Chrysobracteon*	*carinula*		242
20	*Bembidion*		*cordatum*		2
21	*Bembidion*	*Peryphanes*	*lacunarium*		369
22	*Bembidion*	*Hydriomicrus*	*quadratulum*		55
23	*Bembidion*	*Chaudoir*	*rufotinctum*		80
24	*Bembidion*	*Semicampa*	*semicinctum*		46
25	*Bembidion*	*Peryphus*	*tetracolum*		205
26	*Biethisa*		*quadricoliis*		25
27	*Blethisa*		*julii*		5
28	*Blethisia*		*hudsontica*		12
29	*Brachinus*		*cyanipennis*		515
30	*Brachinus*		*fumans*		51
31	*Brachinus*		*medius*		52
32	*Brachinus*		*ovipennis*		14
33	*Brachinus*		*patruelis*		28
34	*Calathus*		*ingratus*		540
35	*Calosoma*	*Calodrepa*	*wilcoxi*		207
36	*Carabus*	*Carabus*	*goryi*		551
37	*Carabus*		*meander*		26
38	*Carabus*	*Tanaocarabus*	*sylvosus*		112
39	*Carabus*	*Carabus*	*vinctus*		219
40	*Cicendela*		*scutellaris*	*unicolor x rugifrons INTERGRADE*	28
41	*Cicindela*		*ancocisconensis*		298
42	*Cicindela*	*Cicindela*	*duodecimguttata*		735
43	*Cicindela*	*Cicindela*	*formosa*	*generosa*	768
44	*Cicindela*		*rufiventris*	*rufiventris*	637
45	*Cicindela*		*scutellaris*	*lecontei*	608
46	*Cicindela*	*Cicindela*	*tranquebarica*		691
47	*Cryobius*		*brevicornis*		1
48	*Desmocerus*		*palliatus*		172
49	*Dicaelus*		*dilatatus*		292
50	*Dicaelus*		*teter*		332
51	*Dicheirotrichus*	*Trichocellus*	*cognatus*		9
52	*Diplocheila*		*assimilis*		27
53	*Diplocheila*		*impressicollis*		48
54	*Diplocheila*		*striatopunctata*		34
55	*Dychirius*		*politus*		15
56	*Dyschirius*		*brevispinus*		3
57	*Dyschirius*		*erythrocerus*		17
58	*Elaphropus*		*dolosus*		89
59	*Elaphrus*		*fuliginosus*		11
60	*Euphydryas*		*phaeton*		72
61	*Geopinus*		*incrassatus*		148
62	*Goniops*		*chrysocoma*		40

Table 1. Continued.

ID	Genus	Subgenus	Species	Subname	Count(*)
63	*Harpalus*	*Megapangus*	*caliginosus*		298
64	*Harpalus*	*Plectralidus*	*eraticus*		104
65	*Harpalus*		*indigens*		17
66	*Harpalus*		*providens*		86
67	*Helluomorphoides*		*praeustus bicolor*		46
68	*Hybomitra*		*lurida*		10
69	*Hybomitra*		*metabola*		3
70	*Hybomitra*		*trepida*		5
71	*Hybomitra*		*typhus*		1
72	*Lophoglossus*		*scrutator*		13
73	*Loxandrus*		*vulneratus*		22
74	*Lucanus*		*elaphus*		18
75	*Miseodern*		*artica*		2
76	*Nebria*		*lacustris*	*lacustris*	717
77	*Nicrophorus*		*americanus*		22
78	*Nicrophorus*		*marginatus*		31
79	*Notioba*	*Anisotarsus*	*sayi*		17
80	*Notiophilus*		*aquaticus*		3
81	*Notiophilus*		*borealis*		9
82	*Notiophilus*		*nemoralis*		79
83	*Notiophilus*		*novemstriatus*		215
84	*Olisthopus*		*micans*		13
85	*Omophron*		*tesseltatum*		42
86	*Panagaeus*		*fasciatus*		67
87	*Paratachys*		*rhodeanus*		9
88	*Patrobus*		*foveocollis*		32
89	*Patrobus*		*lecontei*	*Chaudoir*	1
90	*Pentagonica*		*picticornis*		26
91	*Pericompsus*		*ephippiatus*		30
92	*Platynus*		*cincticollis*		261
93	*Platynus*		*parmarginatus*		135
94	*Platynus*		*punctiformis*		1
95	*Platypatrobus*		*lacustris*		90
96	*Pseudamara*		*arenaria*		4
97	*Pseudoanophthalmus*		*grandis*		70
98	*Pterostichus*	*Cryobius*	*brevicornis*		449
99	*Pterostichus*		*castor*		81
100	*Pterostichus*	*Euferonia*	*lachrymosus*		347
101	*Pterostichus*		*pinguedineus*		35
102	*Pterostichus*	*Steroperis*	*punctatissimus*		51
103	*Rhadine*		*caudata*		79
104	*Scaphinotus*	*Nomaretus*	*bilobus*		11
105	*Sericoda*		*4-punctata*		4
106	*Tabanus*		*longiglossus ?*		1
107	*Tabanus*		*metabolus*		1
108	*Tabanus*		*trepidus ?*		1
109	*Tabanus*		*typhus*		3
110	*Tetracha*		*virginica*		426
111	*Tetragonoderus*		*fasciatus*		194
112	*Tetraluecus*		*picticornus*		11
113	*Tipula*	*Arctotipula*	*williamsiana*		68
				TOTAL	**17337**

Figure 2. *Screen shot of the home page of the SGCN Invertebrates Database website.*

Website Construction.-- During Q4, the main website page design and layout was completed (see Figure 2). The addition of tools and other functionality, such as connecting to the test database, will commence once the database records have been proofed and fully geo-referenced. The website will be available online at the following web address (URL): http://iz.carnegiemnh.org/SGCNInverts/. However, right now, only the home page is available there, until we are able to get the other functions implemented. We expect to update the site *at least* monthly, until the site is fully implemented and ready for Beta testing by the user community. We hope to have the Beta site ready by the end of May 2011.

DATABASE OUTREACH:

We recently have had contact from several people interested in adding their data to the SGCN Invertebrates database. I think this is mainly from the recent announcement for the 2011 RCN grant competition, which stated that all new data grants need to send data to the SGCN Invertebrates website. This is a good thing and will help the database to grow and have an impact on the conservation of rare invertebrates.

DIFFICULTIES ENCOUNTERED QUARTER 4: As in Q3, some of the website tool development has been delayed due to additional time needed for retrospective geo-referencing and correcting/updating of specimen records. The remaining website tools and functionality will be added during the no-cost extension period.

This delay in generating the specimen database necessitates that we officially request a 1 year no-cost extension of the project. However, we fully expect to have the website functional within the next few months.

ACTIVITIES ANTICIPATED DURING THE NO-COST EXTENSION:

We have already contacted the curator at the Buffalo Museum and will coordinate the databasing of 100 specimens (SGCN taxa) as a test of the new website data import system. The database will be completed during this period and the website tools will also be implemented. We expect the beta website to go live by the end of May 2011, with periodic updates occurring *at least monthly* prior to that date to show progress in implementing website functions.

TIMELINE (ADJUSTED):

Once completed, the website will be implemented on the CMNH web server and will be released for beta testing to a restricted set of potential users/contributors for approximately 1-2 months. We welcome input on potential reviewers from NEAFWA state members. We will then gather feedback on the site and information content, making adjustments as necessary, before the site is available to the general community in the Northeastern Region.

To be completed in First Quarter (Q1):

1) Compile a list of the various SGCN invertebrates from the Northeastern Region member states as reported in their submitted state wildlife action plans.
2) Start discussions and design of the SGCN database structure by defining appropriate tables, fields, and metadata so as to incorporate a variety of invertebrate taxa, localities, and conservation information.

To be completed in Q2:

1) Generate and taxonomically refine a final list of SGCN invertebrates from the northeastern region member states by examining state wildlife action plans and by contacting state heritage programs for input on the final list.
2) Finalize detailed metadata descriptions for each field contained in the developed database (i.e., descriptions of data field contents and formatting).
3) Create the database schema and tables that will hold specimen data, locality records and allow functionality of the website.
4) Start gathering and databasing CMNH insect specimens for inclusion in the SGCN database.

To be completed in Q3:

1) Continue gathering and databasing CMNH insect specimens for inclusion in the SGCN invertebrate database.
2) Begin retrospective geo-referencing of databased specimens to allow map plotting.
3) Begin development of a website interface that interacts with and utilizes the database. Create administrative functions, data upload and download functions, record editing functions, and informational displays.

To be completed in Q4:

1) Finish gathering and databasing CMNH insect specimens for inclusion in the SGCN database.
2) Finish retrospective geo-referencing of databased specimens to allow map plotting.
3) Finish design of the website interface. Create behind-the-scenes administrative functions (i.e., data upload and download functions, record editing functions, and informational displays).

To Be Completed During the No-Cost Extension Period:

1) Create interaction with geo-referenced database records that allows GIS map plotting functionality (using imbedded Google maps).
2) Gather specimen data on SGCN invertebrates from other institutional collections to test website functionality for outside users (e.g., Buffalo Museum).
3) Provide diagnostic tools, well-illustrated, to assist others (i.e., non-professional taxonomists) in the identification of SGCN taxa contained in the database (typically specimen photos, genitalic preps (if available), etc.).
4) Initiate a network of museums, NGOs, state and federal agencies in the Northeastern Region that might contribute data on SGCN invertebrates (initially limited to one or a few "beta testers").
5) Release "***beta***" SGCN invertebrate website (for 1 – 2 months) to a selected group of beta testers for their initial reaction to the web interface and for testing and exploration of the various implemented tools and features.
6) Use beta feedback to fix bugs and improve the web interface.
7) Release the "final" SGCN website to the Northeastern Region conservation community.

APPENDIX 1. Fields and metadata for the Specimen table in the SGCN database.

Field	Type	Length	Description
SpecimenID	Integer	10	Primary Key
UniqueNo	VarChar	10	Carnegie Museum Invertebrate Zoology Unique Number
Source	Text		the source of the specimen, can be new acquisition, acquired from another institution, donated, etc.
CMAccession	VarChar	25	CMNH Accession Number
Repository	VarChar	150	The institution that currently houses the specimen.
Genus	VarChar	25	Taxonomic Generic Name
Subgenus	VarChar	25	subgenus name
Species	VarChar	60	species name
Sublevel	VarChar	25	typically set to subspecies, ssp.
Subname	VarChar	25	subspecies name
DetBy	VarChar	35	the initials and last name of the person that identified the specimen.
Sex	VarChar	10	the sex of the specimen
SampleCode	VarChar	45	records the sample or site code for a project or survey.
Country	VarChar	30	country name
State	VarChar	30	state name
County	VarChar	20	the county name
Locality	VarChar	150	short locality description, typically a vector and distance from nearest named place.
Datum	VarChar	35	the datum used to measure the coordinates. If datum is not recorded, use "not recorded" and determine error accordingly.
DecLat	VarChar	10	the latitude, in decimal degrees
DecLong	VarChar	10	the longitude, in decimal degrees
Latdeg	VarChar	15	the degree part of latitude
Latmin	VarChar	15	the minute part of latitude
Latsec	VarChar	15	the second part of latitude, a real number
Latdir	VarChar	3	the latitude direction, N or S
Longdeg	VarChar	15	the degree part of longitude
Longmin	VarChar	15	the minute part of longitude
Longsec	VarChar	15	the second part of longitude, a real number
Longdir	VarChar	3	the longitude direction, E or W
MaxErrorDistance	VarChar	15	the maximum error distance for the Lat-Long determination
MaxErrorUnits	VarChar	15	the units of the error distance.
LatLongRemarks	Text		comments about coordinate determination and error.
Elevation	VarChar	16	elevation in meters
TrapStartDate	Date	10	For specimens collected over a range of trap dates, this is the trap starting date
CollectDate	Date	10	Date that specimens were collected, or if collected over a date range, this is the trap end date.
DateComment	Text		comments about dates, especially difficult to read or interpret from labels.
Collectors	Text		initials and last name of people who collected the specimen
Habitat	VarChar	150	habitat description where the specimen was captured.
ReproState	VarChar	15	reproductive stage of the specimen, typically "adult"
OtherData	Text		Other data that may be associated with the specimen
OtherLabels	Text		Other label data that might be associated with the specimen
Comment	LongText		general comments about the specimen
Associations	Text		lists any other associations of the specimen
ConStatus	VarChar	12	Conservation Status of the species
CollectingMethod	VarChar	150	Method of collection used, e.g., trap type, lure, etc.
KillingAgentUsed	VarChar	150	records the killing agent used, e.g., cyanide, ethyl acetate, tetrachloroethane, etc.
FieldPreservationMethod	VarChar	255	method used to preserve the specimen, e.g., drying, ethanol, papered, pinned, etc.
ProjectName	VarChar	255	Name of project specimen was collected under.
DNAVoucher	VarChar	10	indicates whether a DNA sample was taken and gives its DNA sample number.
SlidePrep	VarChar	10	Slide preparation number
ImageVouch	VarChar	10	Image voucher number
CultureNo	VarChar	25	Associated culture number
Types	VarChar	30	Indicates type status for the specimen.
MultiSpec	VarChar	10	Indicates >1 specimens, or lots.
PrepMethod	VarChar	255	Method used to prepare the specimen.
dbBy	VarChar	40	stores the initials and last name of the person who databased the specimen
dbDate	Date	10	indicates the date the specimen was databased.
LastUpdated	Timestamp	20	indicates the date the specimen record was last updated.
LastUpdateBy	VarChar	40	Name of the person that last modified the specimen record.