Digitization PEN: Integration of data from the San Diego Natural History Museum with the Lepidoptera of North America Network

Michael Wall (PI) and Pam Horsley (co-PI)

INTRODUCTION AND PROJECT GOALS

The Lepidoptera of North America Network (LepNet) Thematic Collections Network (TCN) was funded with the goal to document diversity in the largest clade of herbivores. It brings together a comprehensive set of entomology collections and a large number of digitized museum records to create a dataset that will prove to be significant for years to come and will be used across many fields of research.

The San Diego Natural History Museum (SDNHM) proposes to serve as a Partner to Existing Network (PEN) for LepNet by contributing valuable historical and geographic occurrence records that will fill geographic and temporal data gaps currently in place despite the wide net cast by LepNet. Due to the combination of historic and contemporary focus on southern California (SoCal) and the peninsula of Baja California (BCP) by the SDNHM, the Lepidoptera records to be contributed represent a continuous time series of specimen data for the region that is not found anywhere else (see Table 1 for geographic area definitions and other key acronyms).

Over the two years of funding, this PEN will:

1. Add occurrence records for at least 150,000 Lepidoptera specimens from North America with significant regional additions to SoCal and BCP (see Gap Analysis). Priority will be given to the taxa outlined by LepNet, but particular attention will be given to taxa that will fill geographical and taxonomic gaps as well as generating multiple species records to inform species distribution modelling potential (see #5).
2. Provide high resolution images of 72 holotype Lepidoptera specimens, including six from Baja California, Mexico in the SDNHM collection.
3. Supervise and train at least six undergraduate students, along with numerous volunteers and high school interns. Over 4000 smartphone images will be produced using the LepSnap smartphone imaging workflow and databasing protocols for LepNet.
4. Generate occurrence data and high-resolution photos to aid in the creation of a Butterfly Atlas of Peninsular California, a developing project that is a part of a museum-wide strategy to create atlases for each taxonomic collection that blend specimen-based records with citizen science observations.
5. Add thousands of Lepidoptera records to a multi-taxonomic conservation risk analysis of BCP being conducted by CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad) in partnership with SDNHM. Fully georeferenced Lepidoptera data will greatly add to this project for which there is little current invertebrate data.

Table 1. Definitions of geographic scope and other acronyms.

Acronym	Definition
SoCal	Southern California (San Barbara, Ventura, San Bernardino, Los Angeles, Orange, Riverside, San Diego, and Imperial Counties)

BCP	Baja California Peninsula (Mexican states of Baja California and Baja California Sur)
SCAN	Symbiota Collections of Arthropods Network
LepNet	A Lepidoptera thematic collection network that use the SCAN portal.

SAN DIEGO NATURAL HISTORY MUSEUM COLLECTION

SDNHM has maintained an entomological collection since the Museum's inception in the 1870s. SDNHM is the only major entomological collection in southernmost California and BCP, an area widely regarded as biogeographically significant and a biodiversity hotspot (Myers, et al., 2000). SDNHM started sponsoring expeditions into BCP in 1926, sometimes combining efforts with The California Academy of Sciences in cooperative research projects until 1966. SDNHM researchers studying specific disciplines (e.g., Fred Thorne, Charles Harbison, and David Faulkner all added significantly to Lepidoptera holdings) rather than through large field parties carried out research in northwestern Mexico after 1966. In 1991, SDNHM refocused its mission to become a regional hub for promoting understanding of the evolution and diversity of SoCal and BCP. Currently, SDNHM conducts a regular program of large scale binational and multidisciplinary expeditions to the BCP. In the last twenty years, SDNHM has
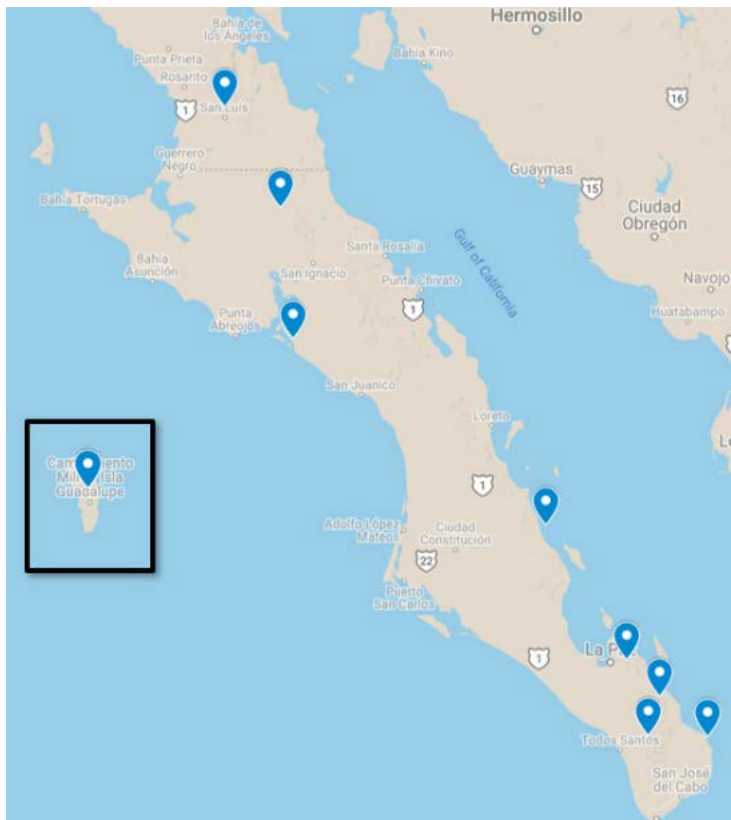

Figure 1. Locations of contemporary (last 20 years) binational multidisciplinary expeditions conducted by SDNHM (Revillagigdo Islands not pictured).

organized/participated in expeditions (Figure 1) to Agua Verde-Punta Mechudo, Isla Guadalupe, Sierra San Francisco, Sierra Guadalupe, Sierra Cacachilas, Sierra la Gata, Sierra de la Laguna, Cabo Pulmo, south coastal Vizcaino Desert, Sierra de la Libertad, and Revillagigedo Islands. In November of 2018, SDNHM will be conducting an expedition to revisit multiple islands in the Gulf of California that the museum last surveyed in 1973. These historic and contemporary expeditions, combined with regional donations and regular research-based acquisitions, have helped to develop an entomology collection of approximately 990,000 pinned and labeled insects and 200,000 insects preserved in 70% ethanol at SDNHM. These specimens are used in support of the Museum's mission: "to interpret the natural world through research, education, and exhibits; to promote understanding of the evolution and diversity of Southern California and the peninsula of Baja California; and to inspire in all people respect for the environment." Although other institutions have large and historically important collections for the region (e.g., California Academy of Sciences, Los Angeles County Museum of Natural History, and University

California Riverside), SDNHM's unique combination of historic and contemporary focus on the region has resulted in an unparalleled continuous time series of specimen data for SoCal and the BCP.

Collection Age - The entomology collection at SDNHM can trace its roots to when land surveyor and amateur entomologist O. N. Sanford joined forces with lawyer and amateur botanist D. Cleveland to form the San Diego Society for Natural History in 1874. Specimens collected by Sanford in the 1850's are housed in the museum's entomology collections. Beginning with these few beetle specimens collected locally in the 1850's, the collection has since grown through field collections and donations of significant individual research collections (e.g., 24,000 specimens of the Fred Thorne and William Hedges Lepidoptera collections). The collection also contains voucher specimens obtained as the direct result of regional environmental surveys and forensic investigations. A third of the collection predates 1950, representing a significant historical record of Peninsular California's entomofauna.

Taxonomic Diversity and Type Material - The entomology collection is particularly strong in regional Lepidoptera (30% of the dry collection) reflecting the interests of many previous curators, collection managers, and research associates (e.g., William S. Wright (microlepidoptera), John A. Comstock (CA butterflies), David Faulkner and John Brown (southern California Lepidoptera). Faulkner also built a relatively large and well-curated collection of Neurotpera that specialists in the taxon have described as "the best collection of northwestern Mexico Neuroptera anywhere." (Tauber, pers comm) The Hymenoptera collection contains significant historical material collected and/or curated by F. X. Williams. In his 12 years as curator, PI-Wall has built the contemporary holdings of the collection with an emphasis on Heteroptera, but has also strengthened the collections across taxa through regional faunistic studies using standardized trapping protocols including Malaise traps, pitfall traps, colored pan traps, and UV bucket traps. Recent collections and curation by James Berrian (SDNHM) and Matt Graham (Eastern Connecticut State University) have drastically grown the holdings and increased the taxonomic precision of the arachnid collection. The type collection includes 211 primary types (72 of which are Lepidoptera) and 500 paratypes.

Local Significance - The entomology collection is the only such repository in the county (4261 sq.mi.; population of 3.05 million). The collection contains the raw data reflecting changes due to growth over the last 150 years for San Diego County (see Figure 5) and 100 years for northwestern Mexico. With the increasing concern over the impact of urban growth on the natural environment (McKinney, 2002), particularly in southern California (Radeloff et al., 2005), the collection provides a source of historical distributional data for land use planning research and environmental surveys. Over the years, the collection has been used by the Environmental Health Department, Medical Examiner's Office, US Fish and Wildlife Service and privately-based environmental consultants. Most recently, the USDA Customs and Border Protection Agricultural Inspection office in San Diego has been using the collection as part of their officers' professional development advancement plans.

Collection Use and Related Publications - The SDNHM entomology collection is underutilized for research. In the last decade, it has averaged only nine loans a year (Figure 2). That said, the scientific products that the collection contributes to are meaningful across a variety of domains. In the last decade the collections have been used: 1) to document the spread of invasive species (e.g., Vetter et al. 2015) and emerging forest pests (e.g., Coleman and Seybold 2008), 2) to contribute to the type series of new taxa (e.g., Johnson 2016, Jiménez et al. 2017, etc.), 3) to contribute to faunistic studies (e.g., Sanborn 2014), and 4) as resources for a variety of student dissertations and theses (e.g., Jones 2014, Keller 2014, etc.).

In addition to the scientific use of the collection, the collection is featured in all facets of SDNHM's education, exhibits, and outreach. Specimens and collection information is featured in: 1) permanent exhibitions like *Coast to Cactus in Southern California*, 2) temporary exhibitions like *Unshelved: Cool Stuff from Storage*, 3) add-ons to traveling exhibitions like a display case on regional scarab beetle diversity for *The Discovery of King Tut*, 4) docent led school programming on insect life cycles, and 5) adult programming events like the Museum's 21+ evening program, *The Secret Society of Adultologists*.
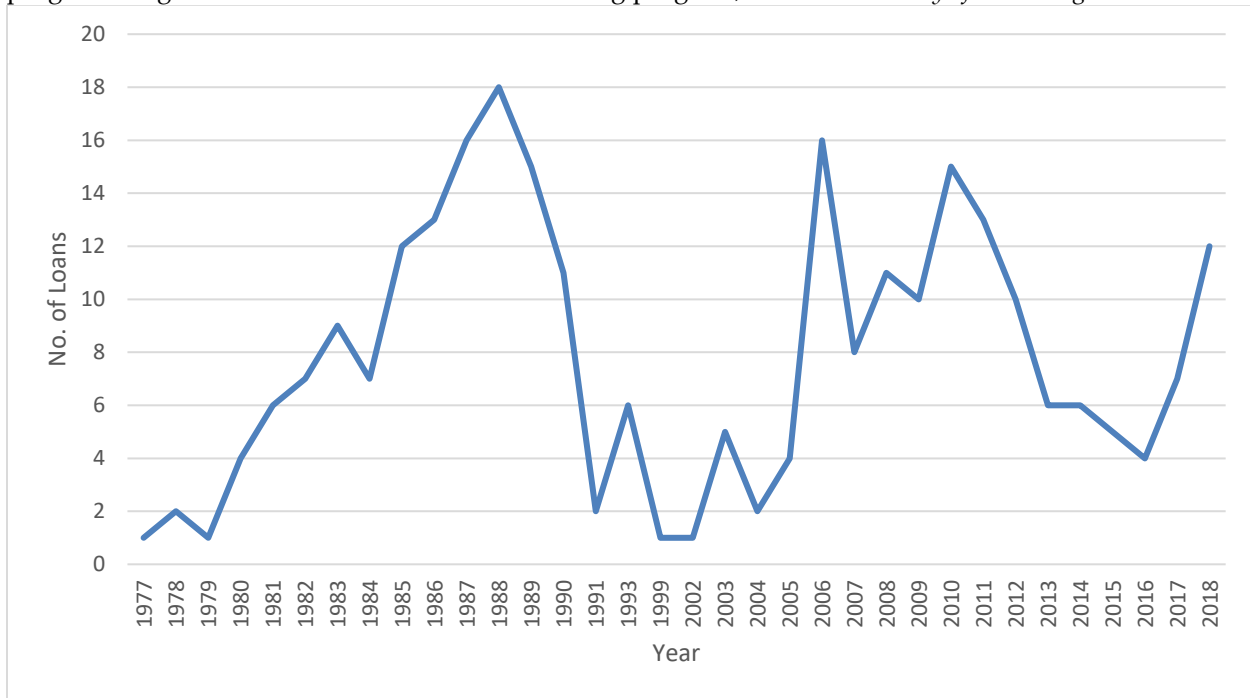

Figure 2. Loan activity for SDNHM Entomology Department from 1977 to present.

Despite the valuable scientific and educational impacts listed above, SDNHM's overall collection impact is hampered by the lack of readily available online collection and specimen data. The collection recently joined the SCAN network and will soon be uploading all previously databased historic records, approximately 25,500 beetles, bees, flies, and tortricid moths. The collection will greatly benefit from the impact of joining a collaborative network and projects increased loan and data queries. In addition to filling taxonomic, geographic and temporal gaps in LepNet, this PEN will dramatically increase the amount of online data that will be more readily discoverable for scientific and education queries.

Volunteer and student involvement - The SDNHM operates an extensive and successful Volunteer Program with opportunities to participate in various activities in the museum, including working in the science departments. On a yearly basis, over 750 volunteers log 56,000 service hours, an equivalent of nearly 1.6 million dollars worth of work to the Museum. The Entomology department has been particularly active in engaging high school and college students as interns, along with other volunteers that collectively contribute an average of 3,400 hours per year to the department. It hosts on average three to four high school interns per year with varying lengths of service determined in coordination with their school districts. The department prioritizes gender and ethnic diversity in both internship and volunteer programs (Figure 3).
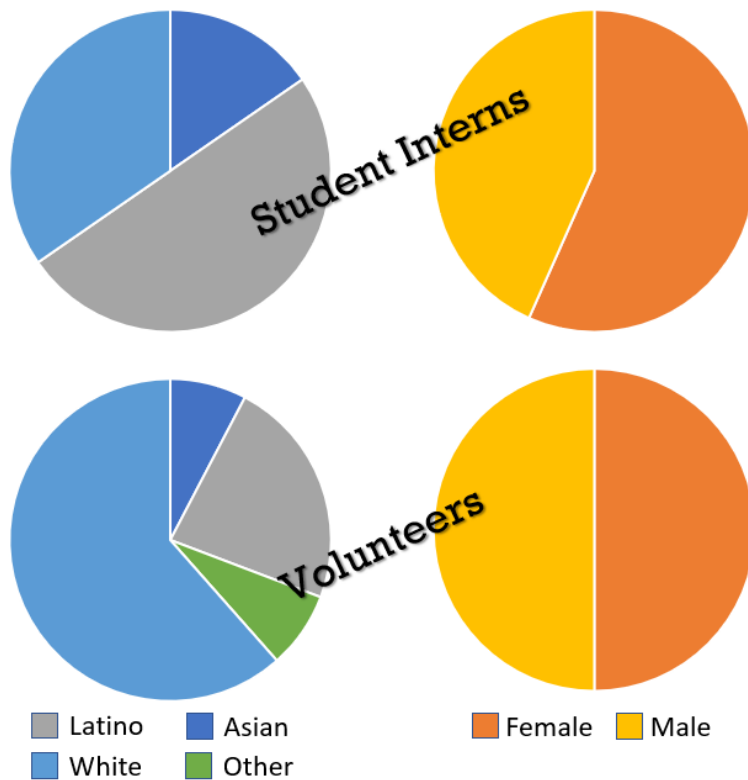
Figure 3. Ethnic and gender diversity amongst volunteers and interns in SDNHM's Entomology Department.

The department has a demonstrable record of engaging students in research. From 2009-2011, with a prior NSF-funded collections improvement grant, the Entomology Department recruited seven undergraduate students, including two Research Experience for Undergraduates (REU), as well as one Research Assistantships for High School Students (RAHSS) to aid in the curation of the collection and complete the goals of the grant. These students were selected from multiple educational institutions in the San Diego area and were of diverse backgrounds. The department keeps students engaged with both the day to day activities of collections-based research (i.e., physical curation, databasing, and georeferencing), while also emphasizing hypothesis-based individual projects (e.g., testing rehydration techniques for ethanol preserved material that had dried out, examining correlations between elevation and species distribution in georeferenced taxa, etc.). The department also emphasizes the importance of science communication. All student interns are required to conduct a brief presentation to interested museum staff about their internship project and experience. Furthermore, some students have presented the findings of their work at SDNHM at professional conferences (e.g., Entomological Collections Network), while others have used this experience to launch further undergraduate and graduate pursuits in Entomology.

INTELLECTUAL MERIT AND CONTRIBUTIONS TO THE LEPNET TCN

Taxonomic and Geographic Scope – This proposal outlines a path to database at least 150,000 records of Lepidoptera from North America. Taxonomically, the collection is composed of 63 lepidopteran families. The collection has particularly large and taxonomically well-curated holdings of butterfly families, Noctuidae, and a variety of microlepidopteran families (Table 2). According to a species-locator inventory completed as part of a previous NSF collections improvement grant, the Lepidoptera collection holds at least 1,870 genera and 5,700 species/subspecies. SDNHM's total Lepidoptera holdings that are identified to family number a little over 180,000. Based on specimen-level inventory work for the Torticidae and other non-lepidopteran taxa (n=25,500), it is estimated that an average of 90% of SDNHM material is from North America. Of the North American material, an average of 80% and 10% are from SoCal and BCP respectively (Table 2). The remaining North American material is primarily from Arizona and Northern California, with an odd assortment of pre-1940 material from other US states (e.g., Washington, Pennsylvania, etc.). Based on these conservative estimates, reaching a goal of over 150,000 North

American records with a clear majority being from SoCal and BCP is achievable within the proposed timeline. Images of at least 4000 exemplar specimens will be captured using the LepNet smartphone specimen imaging protocol. Selection of exemplars will be prioritized in order to capture an image of every species represented with an emphasis on specimens from BCP or SoCal.

Table 2. Lepidoptera holdings at SDNHM for the 20 most specimen rich families with percentile estimates for grant-specific geographic areas.

Family	Total Specimens	90% from NA	10% of NA from BCP	80% of NA from SoCal
Nymphalidae	27,910	25,119	2,512	20,095
Noctuidae	26,427	23,784	2,378	19,027
Hesperiidae	23,233	20,910	2,091	16,728
Geometridae	19,436	17,492	1,749	13,994
Lycaenidae	19,024	17,122	1,712	13,697
Pieridae	14,706	13,235	1,324	10,588
Tortricidae	7,416	6,674	667	5,340
Pyralidae	6,476	5,828	583	4,663
Erebidae	6,076	5,468	547	4,375
Papilionidae	5,359	4,823	482	3,858
Gelechiidae	5,287	4,758	476	3,807
Riodinidae	3,488	3,139	314	2,511
Sphingidae	2,579	2,321	232	1,857
Saturniidae	2,187	1,968	197	1,575
Tineidae	1,765	1,589	159	1,271
Oecophoridae	1,309	1,178	118	942
Lasiocampidae	1,142	1,028	103	822
Elachistidae	1,073	966	97	773
Notodontidae	806	725	73	580
Sesiidae	804	724	72	579
42 other families	4,772	4,295	429	3,436
Total	**181,275**	**163,148**	**16,315**	**130,518**

Gap Analysis - SDNHM's Lepidoptera holdings will add significant geographic, temporal, and taxonomic diversity to LepNet's current specimen-based records.

Taxonomic and Geographic Gaps - SDNHM's over 140 year long geographic focus on SoCal and BCP has resulted in a collection that well represents the region's lepidopteran diversity across geographic space and time. In general, 85-90% of SDNHM's entomological holding are from the southwestern United States with an emphasis on SoCal, and northwestern Mexico with an emphasis on BCP. For example, for the Tortricidae, the only lepidopteran family for which all SDNHM specimens (n=3754) are databased,

60% of specimens are from the southwestern United States, 30% from northwestern Mexico, and the remaining 10% are from other areas in the United States and Central America.

This geographic focus is complementary to LepNet's current coverage. For instance, a query of LepNet/SCAN for Lepidoptera in BCP resulted in 6,766 specimens representing approximately 300 species/subspecies. Depending on the family, it is estimated that 5-30% of SDNHM's holding are from the BCP. Digitization of the taxa proposed here will result in anywhere from 9,000-54,000 records for this underrepresented region. Furthermore, SDNMH's tradition of binational expeditions to remote under-collected areas will fill in specific geographic gaps in the middle of the peninsula (Figure 4).

More surprising than the underrepresentation of BCP in LepNet is the relative lack of data from southern California, a gap that SDNHM is primed to fill. For example, a SCAN query of the Tortricidae of California resulted in 4,291 specimens of which only 345 (8%) were from SoCal. In sharp contrast, 2,092 (94%) of SDNHM's 2,214 California Tortricidae are from SoCal. The only current California member of the LepNet consortium, the Bohart Museum of Entomology at the University of California, Davis, continues to regularly add to the LepNet database, but

Figure 4. Heatmap of LepNet/SCAN Lepidoptera specimen density. Note the presence of a mid-peninsula collecting gap.

their current Lepidoptera data is biased towards central and northern California with only 130 (10%) of their Lepidoptera records coming from SoCal. Inclusion of the regional Lepidoptera collections of the SDNHM will fill a critical gap in an area that harbors incredible biological diversity in the backyards of two of the United States' most populous cities.

Temporal Diversity – SDNHM entomology collections stand out due to strong pre-WWII holdings and consistency of collection across time. To date, the department has databased approximately 25,500 historic (i.e., pre-2006) specimens records, which can be contrasted with the SCAN/LepNet Lepidoptera records for southern California (Figure 5). The first paid curator of the department was the lepidopterist, W.S. Wright, who is primarily responsible for the strength of numerous collections from the 1920s (24%, Figure 5). Charles Harbison took over the collection after WWII and maintained consistent collection during his tenure and participated in several expeditions throughout BCP, including important collections from the peninsula's Pacific and Gulf islands. David Faulkner is responsible for the very
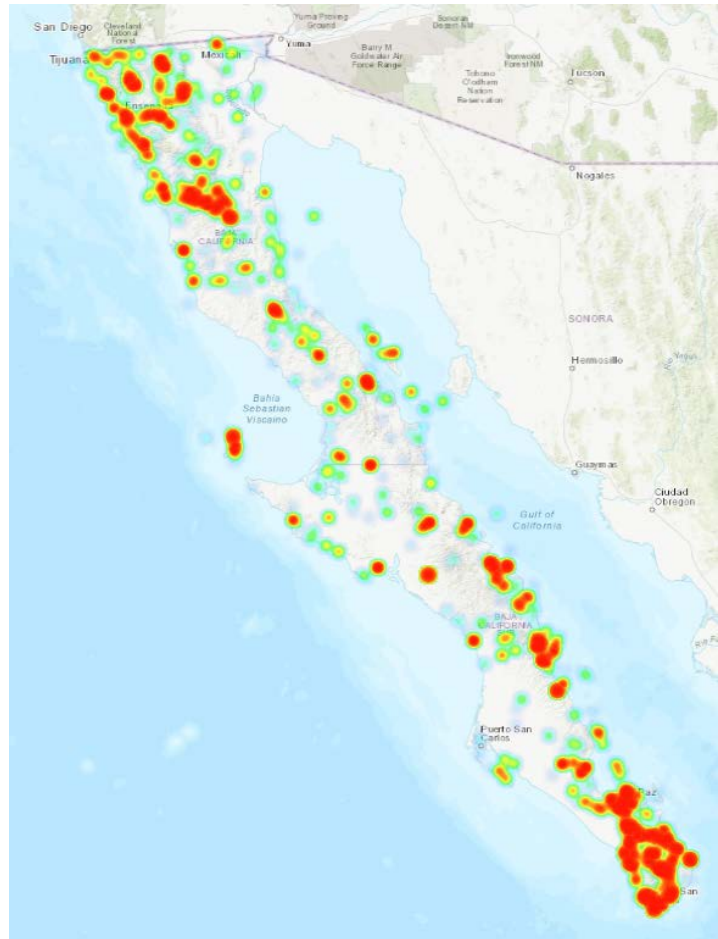
strong collecting spike starting in the 70s that extends to the 90s, a period underrepresented in SCAN/LepNet records (Figure 5).

In the mid-1990s, the department conducted a three-year focused survey of the Lepidoptera of Marine Corps Air Station Miramar (see specimen spike centered on 1996 in Figure 5) that documented approximately 621 species of Lepidoptera based on over 30,000 specimens. The department aggressively sought global specialists to work with the material and ultimately identified 162 species of Noctuidae, 90 geometrids, 75 pyralids, 58 tortricids, and over 230 other species from 35 different families. Sampling efforts for the project were year-round, providing a fantastic phenological data set. This extremely well curated collection is a one of a kind snap-shot of the Lepidoptera of SoCal but remains undatabased. Collectively, all of these collection efforts are complementary to LepNet's current temporal profile for SoCal and BCP (Figure 5). Furthermore, the collection is exceptionally strong in pre-WWII holdings, representing a snapshot of the ecological record prior to runaway population growth and development in SoCal that can be contrasted with the exhaustive 1990s Miramar collection (Figure 5).
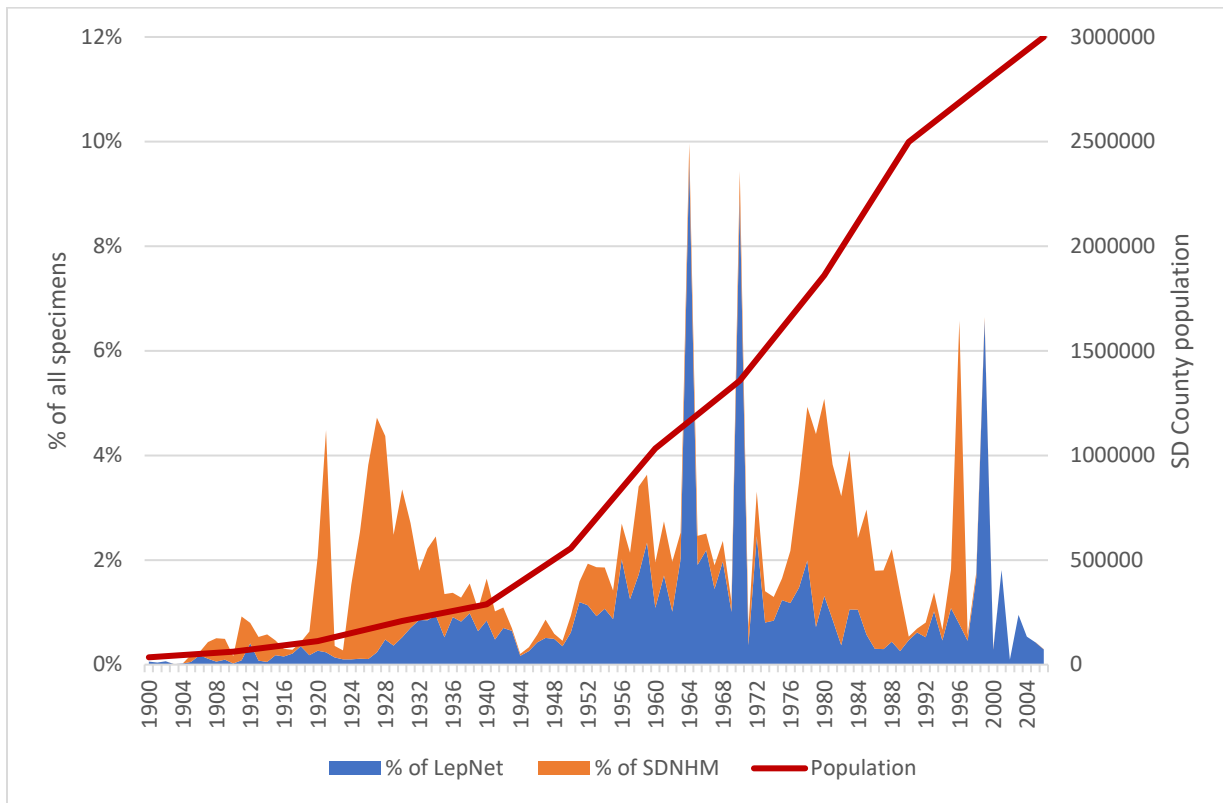

Figure 5. Contrasting temporal profiles for LepNet/SCAN southern California holdings and SDNHM holdings mapped against regional population growth. Note that cumulatively over one-third of the collection predates World War II.

PROJECT MANAGEMENT PLAN

The successful completion of the Entomology Department's NSF-funded collections improvement grant (2009-2011) transformed the physical storage and accessibility of the SDNHM entomology collection. A complete species locator inventory was recorded, non-standard unit trays were eliminated, and a majority of specimens were re-housed in plastazote-bottomed unit trays and new drawers. During this

time, the SDNHM Lepidoptera collection was curated to allow for more efficient digitization of specimen label data. Specimens were organized geographically within unit trays allowing for quick determination of the pertinent specimens for this project. Multiple systematists visited the collection to help identify unsorted material in their respective taxonomic specialties. Most relevant to this proposal, renowned Lepidopterists Jerry Powell, Paul Opler, John Brown, and others visited the collection for a few days of Lepidoptera identification and sorting. Their expertise increased the accessibility of the collection and resulted in many drawers of newly identified taxa.

Organization and Responsibilities - The SDNHM will work within the existing framework of LepNet. LepNet PI Neil Cobb (Northern Arizona University) is responsible for collaborations and overall project management with participating institutions. PEN Lead PI Wall will manage the SDNHM component. Co-PI Horsley will serve as Training and Database Manager. Horsley will be responsible for training incoming personnel on museum curation and digitizing techniques and will also spend time on data quality control, social media, and batch georeferencing of specimen records. The Volunteer Coordinator/Lab Technician will oversee the day-to-day management of digitization work flows, coordinate scheduling, and be responsible for photographing type specimens, as well as contributing 50% of their effort to specimen digitization. In addition, to accomplish the goals of this PEN, participation of three undergraduate students and approximately seven dedicated volunteers per year will be needed (see Task Analysis and Timeline section below). An overview of responsibilities, milestones, and project timeline are in Table 3.

Table 3. Task, responsibilities, timeline, and milestones for span of project.

Personnel	Tasks	Summer 2019	Fall 2019	Spring 2019	Summer 2020	Fall 2020	Spring 2020
Wall - PI	Grant supervision/administration	█	█	█	█	█	█
	Attend LepNet or ADBC Summit meeting		█				
Horsley - Co-PI and Training and Data Manager	Training students and volunteers	█	█	█	█	█	█
	SDNHM taxonomy update	█	█				
	Data quality control/cleanup		█			█	█
	Batch georeferencing		█			█	
	Download DwC-A backup		█	█		█	
	Social media management #MothyMondays	█	█	█	█	█	█
	Attend LepNet or ADBC Summit meeting		█		█		
	Butterfly Atlas of Peninsular California					█	█
Undergraduate & Volunteer Coordinator	Supervision and scheduling of digitizers	█	█	█	█	█	█
	Image type specimens	█		█	█	█	█
	Specimen data entry (50% effort)	█	█	█	█	█	█
Undergraduate students & Volunteers	Specimen data entry	█	█	█	█	█	█
	Smartphone imaging of 4000 exemplars	█	█	█	█		
	Leps by Fieldguide trial		█		█		

Data Capture Workflow – This PEN will benefit tremendously from the infrastructure, tools, and experience already generated via Symbiota, SCAN, and other LepNet institutions (Figure 6). The SDNHM will follow well-established workflows already in place by LepNet. Should updated and more efficient workflows become available, the SDNHM will re-assess and adapt accordingly. The SDNHM will capture label data using Symbiota for at least 150,000 pinned adult specimens of Lepidoptera from southern California and northeastern Mexico. Each year, approximately 7-10 individuals will be recruited to digitize for at least six hours a week, and as such an additional laptop computer and smartphones for image capture is requested in order to maintain digitizing capabilities for these individuals and ease scheduling conflicts.

Digitizing New Specimen Data

1. **Training**. Students and volunteers will be trained by the Training and Data Manager on proper databasing and imaging protocols in line with those provided by LepNet. They will be educated on the principles of classification and nomenclature, become familiar with the SDNHM entomology collection, and learn proper specimen handling and museum protocols. Prior to digitization, review of Symbiota and LepNet protocols and how-to videos will be required in order to provide more efficient hands-on databasing tutorials.

2. **Taxon assignment**. Students and volunteers will each be assigned a Lepidoptera higher level taxon to work on by the Training and Data Manager. Each digitizer will mark the outside of drawers where they are currently working and work with the Lab Technician to maintain a manifest of digitization. Digitizers will begin with specimens identified to the genus and/or species level. Upon completion, digitizers will continue with those specimens identified only to family, where they may be able to provide tentative determinations to lower taxonomic levels. The determinations will be flagged with low confidence in Symbiota. Digitizers will then use the Leps by Fieldguide app to image exemplars of these specimens, which will serve the dual purpose of testing the app's identification capabilities and reinforcing the digitizer's knowledge of Lepidoptera taxonomy. Taxa and specimens from outside of North America will not be digitized.

3. **Barcodes**. Digitizers will add archival unique identifier barcode labels to each specimen just prior to data entry in Symbiota. The SDNHM uses small labels that include both a human readable code (SDNHM acronym, plus number) and a machine-readable matrix.

4. **Data entry**.

    a) *Label data* - Label data capture will be conducted continuously over the 2-year period by a team of digitizers working on multiple Lepidoptera families at once. There will be four workstations available in the SDNHM Entomology lab for data entry in Symbiota. As Symbiota is web-based, digitizers will also have the option to use their own laptops when desired. The Training and Data Manager will customize the LepNet Data Entry guidelines to help reflect specific SDNHM label and geographic idiosyncrasies.

    b) *Exemplar species imaging* - Exemplars from each taxon will be photographed using LepNet smartphone imaging protocols resulting in approximately 4000 high quality images. Digitizers will be provided with an appropriate smartphone to ensure the image resolution is within the guidelines set by LepNet.

5. **Data editing/cleanup**. The Training and Data Manager will monitor student and volunteer specimen label data entries continuously and communicate issues that arise with digitizers. A more thorough data quality control check will be performed twice a year prior to batch georeferencing and again a final time upon completion of the PEN project.

6. **Georeferencing**. Following data entry and editing, the Training and Data Manager will batch georeference specimens primarily using Symbiota-GEOLocate following the best practices developed through iDigBio. Although there are protocols for using the georeferenced records available in iDigBio, the ability to assign coordinates for some of the historical and obscure specimen records that will be contributed by the SDNHM may require alternate measures. The Training and Data Manager will work with Larry Gall (LepNet) and Nelson Rios (GeoLocate) at Yale University to evaluate GeoLocate's reference database for Baja California. In an effort to increase the efficiency of georeferencing Baja California Lepidoptera records, SDNHM will cross-reference localities with the "Historic Place Names of Baja States" as available on the BajaFlora website hosted by the SDNHM (BajaFlora.org). These localities will likely contribute a significant number of new records for GeoLocate to incorporate into their database and will make the program more powerful for the future.

Imaging - The SDNHM will follow the 15 imaging standards and "Guide to knowing when to image the ventral surface of Lepidopteran wings" already adopted by LepNet and available online. High resolution images of the 72 holotype specimens will be completed using an existing in-house Visionary Digital BK Lab imaging system (see Facilities and Equipment supplemental document for details). Following guidelines provided by LepNet, at least 216 publication-quality images of dorsal/ventral/key structures will be produced and aid in future research. These images will be served to LepNet via the iDigBio image appliance bulk-uploading feature and image copies will be backed up on the SDNHM server.

Students and volunteers will be responsible for producing over 4000 smartphone images of exemplar species over the course of the project. These images will be stored on the SDNHM server and will be served to LepNet via Fieldguide iOS and Android apps that are freely available. The Leps by Fieldguide app is integrated with the LepNet database so that catalog numbers of specimens are linked to images in the existing record. The app also provides an instantaneous "suggested" identification via a convolution neural net process. This suggested identification can be verified or nullified against human identifications, thereby increasing the predictive power of the app over time.
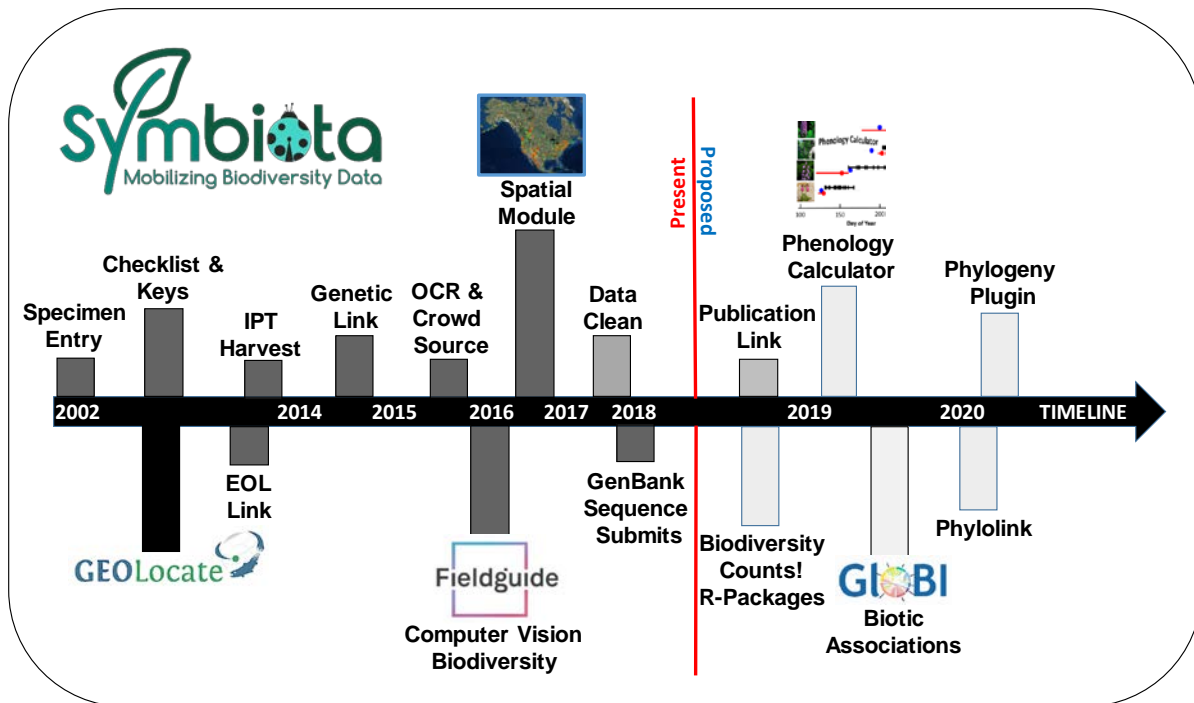
Figure 6. Functions built into Symbiota are listed above timeline bar and functions that are linked (API) are shown below the bar. Shading represents degree of development (darker = more developed), length of bar indicates complexity of function.

Training Plan - The Training and Data Manager will recruit and train students and volunteers. They will work alongside and coordinate with the Volunteer Coordinator/Lab Technician who will be responsible for the day-to-day management of digitization. The Volunteer Coordinator/Lab Technician will also oversee the scheduling of students and volunteers, as well as spend 50% of their effort digitizing specimens. As has been proven successful by previous projects, recruitment periods for undergraduate students will occur at the beginning of each semester and will target the multitude of educational institutions in the San Diego area. Recruitment of volunteers will be ongoing and on an as-needed basis via the SDNHM Volunteer Department.

Coinciding with recruitment periods, the Training and Data Manager will aim to group train hired undergraduate students and will also provide periodic volunteer training sessions. These training sessions will educate students and volunteers on the principles of scientific classification and nomenclature, familiarize them with the entomology collection, and teach them proper specimen handling and museum protocols. In order to facilitate training and standardize techniques, digitizers will be required to become familiar with Symbiota and LepNet protocols, how-to videos, and other relevant information prior to meeting with the Training and Data Manager. Exposure to these informative materials prior to the training sessions will decrease the initial training time required and allow for more efficient databasing efforts.

Task Analysis - Based on an average databasing rate between 20-25 specimens/hour, SDNHM estimates that it will take approximately 6,000 hours to complete databasing of the 150,000 target specimens (Table

4). These databasing rates are based off SDNHM's recent digitization of 25,500 specimens and have been confirmed by other LepNet institutions (Cobb – NAU & Cognato – MSU, pers. comm.). The Volunteer Coordinator/Lab Technician will allocate 50% of their effort to digitization of specimens and will develop a higher average digitization rate over time. SDNHM also requests funds to hire three undergraduate students at 400 hours each per year who will conservatively attain a digitization rate of 22 specimens per hour based on SDNHM and LepNet institution experience. The remaining specimen records will be digitized with the help of at least seven volunteers per year. The SDNHM requires volunteers to commit a minimum of 6 hours per week with volunteers averaging approximately 44 weeks each year (taking into account their possible vacation time and schedules). This represents 1848 volunteer hours per year and easily accomplishable within the 3400 yearly volunteer hours that the department currently maintains. While seemingly ambitious, based on the successful undergraduate and volunteer recruitment record at the SDNHM, the PIs are highly confident that the department can maintain this level of efficiency throughout the course of the PEN and meet proposed digitization goals.

Table 4. Digitization rates and specimen output bases on task analyses and data from other LepNet institutions.

Personnel	Specimen digitization rate/hour	Hours contributed	No. of individuals	Years	Specimens
Volunteer Coordinator/Lab Technician	25	520	1	2	26000
Undergraduate students	22	400	3	2	52800
Volunteers	20	264	7	2	73920
					152720

Travel - As per program requirements, travel to iDigBio for Wall and/or Horsley is budgeted for year 1 of the project. LepNet will be approaching its final year of funding, and subsequent network meetings are unknown. The PIs will communicate with LepNet to coordinate activities with those of the network.

BROADER IMPACTS

General Public Engagement – In the words of the original LepNet proposal, "Butterflies and many moths are consummate invertebrate biodiversity ambassadors; no other arthropod group captures the interests and affection of the public more than Lepidoptera. These groups serve as gateways to appreciate the natural world, and the vast networks of clubs in our region are testament to the immense education and outreach potential for Lepidoptera." The San Diego Natural History Museum hosts over 450,000 visitors every year with an additional 500,000 visiting the website. Over the two years of this grant period, the Entomology Department will work with education and outreach partners to integrate butterflies and moths into existing programming in which the Entomology Department already participates (i.e., behind the scenes tours, Annual Meeting Open House, Nat at Night, summer camps, Secret Society of Adultolgists 21+ programming, and more). The department will maintain participation metrics for all these activities.

Student Involvement - SDNHM proposes to train three to six paid undergraduates and at least six high school student volunteer interns as part of this grant. The Museum has a proven track record of engaging ethnically and gender diverse students in departmental activities. While the bulk of this grant is undeniably a lot of data entry, PIs plan to develop modest hypothesis driven projects that are dependent on the student's data entry efforts in order to complete. The PIs will mentor students in informal science communication. All students will present the results of their efforts at internal seminars and through SDNHM's social media channels.

Collection Accessibility - The digitization efforts outlined here will increase SDNHM's available specimen data from 1.7% to 11.5%. According to a survey of entomological collections (Cobb unpublished), private non-profit museums like SDNHM contain 30% of entomological specimens, yet have the lowest digitization rate (3%) relative to university and government supported collections. This effort will not only increase the quantity of data and improve its availability, but will also build institutional knowledge and capacity to be carried over into future digitization efforts for other taxa.

Improved Taxonomic Curation - This effort will improve the taxonomic curation of SDNHM's Lepidoptera collections. Of the over 180,000 Lepidoptera specimens in SDNHM's collection inventory, 17% are only identified to family. SDNHM proposes to use LepSnap (the image recognition AI integrated with SCAN) to provide tentative identifications to unidentified material. Specimens with these AI generated determinations will be flagged within LepNet/SCAN with an "ID confidence" of "0 – ID requested" and the "Identified by" field will be labeled as "LepSnap AI". This smartphone-based workflow will be a case study using students and volunteers to provide tentative identifications for expert evaluation amongst the LepNet community.

Butterfly Atlas of Peninsular California - Data from this digitization project are critical for the development of The Butterfly Atlas of Peninsular California, a project that the Entomology Department is currently developing. The San Diego Natural History Museum is well known for its regional "Atlas" books and websites. Publication of the San Diego County Bird Atlas (Unitt 2004) was followed by the development of virtual botanical (Rebman 2007) and herpetological (Hollingsworth 2013) atlases. Most recently, the Museum published the San Diego County Mammal Atlas (Tremor 2017). These projects have all blended citizen science with voucher-based museum records to create synthetic educational resources about regional biodiversity. As SDNHM looks to integrate these taxonomic atlases into a singular multi-taxon resource, the Museum has prioritized the digitization of butterflies as an initial entomological representative in this collective effort. With over 90,000 specimens of butterflies in SDNHM holdings, LepNet digitization efforts will significantly increase the voucher-based collections of this project.

Social Media - Entomology staff and interns will work with SDNHM's marketing staff to create a weekly social media outreach campaign centered on #MothyMondays. Using images generated from this proposal, staff will create a weekly post of pictures combined with an interesting fact, and links back to the LepNet/SCAN taxon page. SDNHM currently has close to 43,000 followers on Facebook and posts average close to 6,300 total impressions. Assuming 52 posts a year for two years, 600,000 LepNet related impressions should easily be generated.

Data for Binational Conservation Management - Data from this project will be crucial to adding invertebrates to multi-taxonomic conservation risk analysis of BCP being conducted by CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad). Starting in June 2017, the SDNHM and CONABIO began a collaborative project with other regional biodiversity researchers and NGOs to

strengthen understanding of the status of research and conservation efforts in the peninsula of Baja California. This project aims to put in one place all the information that SDNHM and partner have to date about the distribution of species in the peninsula and make it available online as maps and GIS files. This project is now being taken to the next level as SDNHM works with CONABIO to conduct peninsula-wide species distribution modeling across multiple vertebrate and plant groups. Unfortunately, the only invertebrate data incorporated in the analysis is for a few dozen scorpion taxa. Data from SDNHM LepNet efforts will be directly shared with CONABIO, adding hundreds of invertebrate taxa and thousands of records to a dataset currently dominated by plants.

SUMMARY

The San Diego Natural History Museum proposes to database at least 150,000 specimens for incorporation in the LepNet TCN. This material is geographically and temporally complementary to the holdings of LepNet's current contributing institutions, and includes important pre-WWII records for southern California, a large number of records from the Baja California Peninsula, and exhaustive collections (621 species/30,000 specimens) from a three-year survey of a Marine base in San Diego County. Multiple high-school students, undergraduates, and volunteers will be trained in the basics of collections-based research and Lepidoptera taxonomy. In addition to creating a lasting legacy of online collections data for perpetual use by researchers and educators, these databased records will be of immediate utility in binational conservation risk analysis, and a collections-based citizen science oriented atlas to the butterflies of Peninsular California.