AGRICULTURAL RESEARCH FOUNDATION
INTERIM REPORT
FUNDING CYCLE 2021– 2023

TITLE: Eastern Oregon Beneficial Invertebrate Database Project: Compiling Long-Term Datasets on Native Bees to Advance Research and Extension

RESEARCH LEADER: Sandra J. DeBano

COOPERATORS: Hans DeBano and David Wooster

SUMMARY/ABSTRACT:
In the 20 years since its inception, the principal investigators, post-doctoral scholars, and graduate students of the OSU Hermiston Agricultural Research and Extension Center (HAREC) Invertebrate Ecology Laboratory have worked on over one hundred different projects focused on beneficial invertebrates spanning five counties in eastern Oregon. We have accumulated large datasets on many types of beneficial invertebrates, with the two largest groups studied being native bees and aquatic invertebrates. This project focuses on native bees. Data from studies on native bees are currently stored in separate Excel spreadsheets for each individual project. Each spreadsheet contains dozens or more of columns of information that are associated with individual bee specimens collected. The type of information associated with each specimen includes data on its location, taxonomy, date and time collected, collection methods, treatment types, land use, life-history information, a unique identifier number, and more. Each native bee dataset is slightly different, and through the years, we have accumulated hundreds of different Excel files. The bee data and associated specimens represent a valuable and unusual asset. Unfortunately, there is currently no efficient or easy way to extract information from all the datasets simultaneously, which limits our ability to address important research and extension questions. The creation of one database to store the information from all these datasets will allow us to address issues important to a variety of our stakeholders. For example, growers are interested in knowing which native bees exist near their farming operations, when those bees are active seasonally, and what types of plants may support them when crops are not blooming. They can use this type of information to enhance on-farm habitat for pollinators, which can lead to both higher yields and higher quality crops. Stakeholders in natural resources (USFS, BLM, ODFW, and Tribes) are asking for more information on how they can manage working landscapes and undertake restoration projects in ways that meet not only traditional uses of these lands (e.g., livestock production, recreation, timber harvest) but also enhance pollinator habitat at the same time. All of these stakeholders will benefit by an ability to query a database with over 70,000 bee records with associated data to answer a variety of questions. An additional byproduct of the project is that we will be able to document the set of use cases common for similar datasets and the details of individual queries that we and others with similar types of datasets might have.

OBJECTIVE:
Develop a database for native bees of eastern Oregon based on 20 years of research that can be queried for information to address many different types of questions, including research and extension questions.

PROCEDURES:

The general process we are following includes the following steps:

1) Locate all relevant datasets (saved as Excel spreadsheet files) over the last 20 years.
2) Conduct quality assurance and control for each dataset, including insuring proper metadata are recorded.
3) Work with a database consultant to design the best structure for the type of data we have and the questions we wish to address – a step that involves producing a conceptual data model.
4) Use MySQL or other database software to build the database, which includes creating, initializing and populating the database with the datasets from the last 20 years.

SIGNIFICANT ACCOMPLISHMENTS TO DATE:

Significant accomplishments so far include:

1) Forming the project team, and establishing a shared document and data workspace
2) Agreeing upon data format, standards, and use cases
3) Selecting a technological architecture
4) Configuring and testing Amazon Web Services (AWS) for hosting the database
5) Deciding to use MySQL Workbench for our database software and learning how to use it
6) Conceptualizing the overall database design
7) Formulating the database design, including organizing it with five tables (a specimen table, a taxonomic table, a trait table, a location table, and a sample table)
8) Gathering over 70,000 specimen records for inclusion into the database
9) Testing the database structure to determine whether it can link five sample tables and can be used to respond to basic queries, with the ultimate goal of producing a structure that allows us to perform data maintenance and to ask the questions of interest now and in the future
10) Determining the most efficient way to generate sunrise and sunset times for a date range at a particular location, which can then be used to calculate hours of sampling effort
11) Investigating ways to develop the database so that it allows us to produce abundance data adjusted by sampling effort

ADDITIONAL FUNDING RECEIVED DURING PROJECT TERM:
We have not received additional funding for the project at this time.

FUTURE FUNDING POSSIBILITIES:

The creation of the database will allow us to use all of our varied datasets from multiple studies through the last two decades to generate preliminary data to support new grant proposals for competitive programs such as the USDA AFRI programs. Several of the AFRI programs in the "Plant Health and Production and Plant Products and Bioenergy" and "Natural Resources, and Environment" categories have funding for pollinator-focused research in agroecosystems. Our dataset should be functioning in time to generate preliminary data for the August/September 2022 round of deadlines for those programs.