|  |  |
| --- | --- |
| **Data is the New Science:**  **An Introduction to Biodiversity Data Types and their Uses** |  |

## Objectives

Students completing this module will be able to:

* Access data from biodiversity digital data repositories
* Evaluate the research utility of occurrence data derived from different sources.
* Create and interpret a graph.
* Use geo-spatial data to inform biological thinking.
* Describe how a change in a system can impact multiple parts of a system.
* Explain how morphology, behavior, and physiology of a species can impact another species in a co-evolutionary relationship.

**Introduction**

There is a changing landscape for those joining the 21st century workforce. Rapid advances in data research and technology are transforming how we conduct science. The volume and variety of data being generated, the increased accessibility of data for aggregation, the improved discoverability of data, and the increasingly collaborative and interdisciplinary nature of scientific research are driving the need for new skill sets.

The biodiversity sciences, for example, have experienced a rapid mobilization of data that has increased our capacity to investigate large-scale issues. **Biodiversity** is the variety of life on earth. Scientists study biodiversity at many different scales from the variation in the genes of a population to the diversity of species present in an ecosystem. Biodiversity can be described in many ways, including taxonomic, morphological, genetic, ecological, and functional. As we look around our world, all the variable forms of living organisms we see are the product of over 3.5 billion years of evolution in the form, function, and processes of life on earth. In order to address scientific issues of critical national and global importance, such as climate change, zoonotic disease, resource management, invasive species, and biodiversity loss, the 21st century biodiversity scientist must be fluent in integrative fields spanning evolutionary biology, systematics, ecology, geology, genetics, biochemistry, and environmental science and possess the quantitative, computational, and data skills to conduct research using large and complex datasets.

In this module, you will be introduced to some emerging biodiversity data resources. You will be asked to think critically about the strengths and utility of these data resources and then encouraged to think beyond the obvious to how these data could be used to answer big science questions. You will then use what you have learned about these data sources in an example case study, investigating the relationship between two coevolving species.

**Activity 1: Examine Specimen-Based Occurrence Data**

The data we collect about the where and when a species is found is called **occurrence data.** They are information on the presence of an individual from a defined species at a specific place and time. Information from occurrence records has proven to be highly valuable, allowing scientists to examine changes in distributions over time, perhaps in correlation with specific environmental factors, and to compare the distributions of different species. Occurrence data can be gathered by the collection of specimens, which are then preserved in a natural history collection, or may be based on records of observations.

***Specimen-based data*** are based on archived biological specimens housed in a natural history collection. Scientists, naturalists, and explorers have been collecting and preserving specimens for hundreds of years. Today, we can interact directly with the specimens collected by Darwin, Lewis & Clark, and Teddy Roosevelt! Preserved with the specimens is a variety of information about the organism (e.g., collector, collection date, location, habitat, images, community assemblage, phenology). Specimens can be further examined to verify the data point or to yield additional information on the collected organism as scientists build on previous research or identify new questions to investigate. Preserved specimens from natural history collections are a treasure trove of data and information and have been used to study a variety of topics, such as evolutionary relationships, pesticide use, host-parasite evolution, and zoonotic disease transmission, etc. In 2010, museums and other collection holders initiated a massive data mobilization effort to provide digital databases of these archived specimens. This data is aggregated in searchable portals that provide broad access to information on individual specimens, images of the specimens, and associated data. These efforts have vastly increased the accessibility and utility of these data.

1. Go to the following website: <http://libguides.cmich.edu/lifesciencedata/organismal>. Here you will see several groupings of data sets. First, we will examine specimen-based data. We will be using the iDigBio portal for this activity. Click on the [iDigBio link](http://idigbio.org/) provided. Spend some time looking around the iDigBio site to find the answers to the questions below.
2. What is iDigBio short for?
3. How many specimen records are currently available through this portal?
4. Pick a species and search the iDigBio portal for occurrences records. Refer to the iDigBio User Guide as needed to help navigate the portal.
5. What species did you choose?
6. How many records of your species were found?
7. Use a colored pencil to outline the distribution of your species on the map below:



credit: pixaby.com [CC0 Creative Commons License]

1. What is the date that the oldest preserved specimen was collected?
2. Click on the “must have media” button. Once the dataset refreshes, “view” some of the specimen records. What types of media are available for researchers to use?

**Activity 2: Examine Observation-Based Occurrence Data**

***Observation-based data*** is information gathered by scientists, naturalists, and citizen scientists and represents an observation of an organism. Observation data is not linked directly to a physical specimen. While some data may be associated with just a location and date, other data types may be accompanied by a photo and detailed information (e.g., collection methods, environmental conditions, geographic location, associated species, weather, behavior, abundance, phenology). Similar to specimen-based data these data can provide a wealth of information on biodiversity and are now included in many online databases accessible to researchers, educators, and the public.

1. Return to this website: <http://libguides.cmich.edu/lifesciencedata/organismal>. The next set of questions relates to the GBIF data sets. These data sets include both specimen and observation-based occurrence records. We will be using GBIF for this activity. Click on the [GBIF link](https://www.gbif.org/) provided. Spend some time looking around the GBIF site to find the answers to the questions below.
2. What does GBIF stand for?
3. How many occurrence records are currently available through this portal?
4. What is Darwin Core?
5. Search the GBIF portal for occurrences for the same species you used in the iDigBio exercise. Refer to the GBIF User Guide as needed to help navigate the portal.
6. How many records of your species were found?
7. What other information and resources are provided with the results of your search?
8. Use a different color pencil to outline the distribution of your species on the map you made of specimen based iDigBio data.
   1. How do the two distributions compare each other?
   2. What factors might contribute to any differences you found between the two distributions?
   3. Which data source, iDigBio or GBIF, do you think would be better to use to build a model of your species’ distribution? Justify your answer by discussing the volume of data available, types of records (occurrence- vs. specimen-based) included, accuracy of identification, and the ability to verify records.

**Caution**: Occurrence data, like all data types, have limitations. Some limitations of occurrence data include uneven sampling, observation or collector bias, and incorrect identification. Part of data literacy is considering the limitations of the data and the suitability of the data for addressing the research question of interest. For example, when using occurrence data, the researcher needs to keep in mind that the absence of an occurrence record for a species at a location does not prove that it has never existed there, only that an occurrence was never recorded there. It is also important to take into consideration the biology of your species of interest. Considering the biology and ecology of your organism or ecosystem of interest, combined with an understanding of the sampling methods for gathering data and the complexity of the data records, are all critical aspects of working with occurrence records.

**Activity 3: Examine Environmental Data Sets**

**Environmental data** are available from multiple sources and include detailed information on a variety of abiotic variables (e.g., soil types, climate variables, hydrology, land use). Environmental data is available for different regions, with varying levels of resolution, and for different time frames. This data can be used to address long-term and short-term questions related to anthropogenic disturbance, climate variation, historical weather patterns, changing land use, environmental contaminants, etc.

Go to the following website: <http://libguides.cmich.edu/lifesciencedata/environmental>. Here you will see several environmental data sets. Pick a dataset to investigate. This set of questions relates to the environmental data sets. Use that data link to address the questions below:

1. Which environmental data source did you review?
2. What types of data are available at your selected data source?
3. What types of questions could be addressed using this data? List three questions you think could be addressed with your environmental data set.
4. Question I:
5. Question 2:
6. Question 3:

**Activity 4: Using Biodiversity Data to Investigate Plant-Pollinator Coevolution – Distribution**

|  |  |
| --- | --- |
| *Ophrys sphegodes*, the early spider orchid, has flowers with yellow-green sepals and petals and one highly modified, velvety brown central petal (a **labellum**) at the base of the flower. Each flower contains both an androecium and a gynoecium that is fused into a structure called a column or **gynostemium**. There is single male anther is found at the tip of the column and instead of producing individual pollen grains, the pollen is concentrated in two masses (called **pollinia**) at the end of the anther (you can see the green column curved over the top of the brown labellum in the picture below). The stigma is on the underside of the column just below the anther. | Macintosh HD:private:var:folders:TL:TLeomGpAH24WJTp1NCqapk+++TM:-Tmp-:TemporaryItems:Early-spider-orchid-flower-Dorset.jpg |

|  |  |
| --- | --- |
|  | *Ophrys sphegodes* is pollinated by the solitary bee *Andrena nigroaenea*. These orchids attact pollinators by sexual deception. They produce the sex pheromones of female bees and also provide visual and tactile cues, attracting the males for “mating”. During pseudocopulation, the pollinia are transferred to the male bee’s head and pollen can be transferred to the next flower visited. The flowers don’t provide nectar to the bees as a reward, so there is a net energy loss to the bees who are fooled by the deception. |

1. Go to [www.iDigBio.org](http://www.idigbio.org)/portal. Search for *Ophrys sphegodes*.
2. How many specimen records were located?
3. Outline their distribution on the map below:



credit: pixaby.com [CC0 Creative Commons License]

1. Add *Andrena nigroaenea* to your search. Outline their distribution on the map in question 2 in a different color.
2. How many *Andrena nigroaenea* specimen records were located?
3. How does their distribution compare to *Ophrys sphegodes*?
4. Go to <https://www.gbif.org/>. Search for *Ophrys sphegodes*. Outline their distribution on the map in question 2 in a third color.
5. How many specimen records were located?
6. Add *Andrena nigroaenea* to your search. Outline their distribution on the map in question 2 in a fourth color.
7. How many *Andrena nigroaenea* specimen records were located?
8. Looking at the GBIF data, how does their distribution compare to *Ophrys sphegod*
9. Use all data you have entered on the map to compare the distributions of these two species.
10. Do the distribution of *Ophrys sphegodes* and the range of *Andrena nigroaenea* overlap? What does this mean biologically?
11. If the distribution of *Ophrys sphegodes* and the range of *Andrena nigroaenea* did not overlap, what would this tell you about the pollination of *Ophrys sphegodes*?

**Activity 5: Using Biodiversity Data to Investigate Plant-Pollinator Coevolution – Phenology**

**Phenology** is the study of the timing of cyclical events in an organism's life cycle, such as the flowering of plants, emergence of worker bees from the hive or the migration of birds. The timing of critical life stages can be triggered by external environmental clues such as seasonal temperature change, photoperiod, or precipitation. As global weather patterns alter or fluctuate due to climate change, an organism's phenology may shift in response to a change in triggers. If species that interact closely (e.g. pollinator and plant, predator and prey) respond to triggers that are no longer synced, the interaction may be disrupted and have negative impacts on one or both species.

To address the impacts of climate change on phenology we need to see long-term historical patterns and trends in both the environmental conditions and phenology of the species of interest. No problem! We are in the data rich era of science. Environmental Data and Occurrence Data from natural history collections, professionals, and citizen scientists provide scientists the necessary information to investigate phenological changes.

In 2014, Robbirt et al. examined these types of data to determine if the interaction between *Andrena nigroaenea* and *Ophrys sphegodes* was being disrupted by climate change. The researchers examined 102 herbarium specimens of *Ophrys sphegodes* from the Natural History Museum in London. For each specimen, the percent of open flowers was determined. This data was used to determine the peak flowering dates for each year for which there was adequate data.

To see how the timing of the bee reproduction aligns with peak flowering dates, the researchers examined field records of the flight date (first flight of spring) for *Andrena nigroaenea* from the Bees, Wasps, and Ants Recording Society (BWARS). BWARS is a citizen science project in Britain and Ireland. Volunteers are trained to make and record accurate observations of these organisms into a shared database. The researchers also examined over 350 *Andrena nigroaenea* specimens at the Natural History Museum and Oxford University Museum. Based on these two sources, the researchers determined the average date of first flight for each year.

Climate data for all years were obtained from the UK Meteorological Office and mean March to May (Spring) temperature was calculated for each year. Summaries of these data are included in Table 1 below.

Table 1. Comparison of flowering time and bee emergence at different spring temperatures. [Data extracted from Robbirt et al., 2014]

|  |  |  |
| --- | --- | --- |
| March – May Temperature  (°C) | Peak Flowering Time  (days after March 1) | First Bee Flight  (days after March 1) |
| 7.0 | 85.1 | 92.9 |
| 7.5 | 81.8 | 85.0 |
| 8.0 | 78.5 | 77.3 |
| 8.5 | 75.4 | 69.7 |
| 9.0 | 72.2 | 61.8 |
| 9.5 | 69.0 | 54.1 |
| 10.0 | 65.7 | 46.2 |

1. Plot March – May temperature (x-axis) vs. peak flowering time and first bee flight (y-axis) on the same graph. Use Excel to graph your data. Print your graph and turn it in with your lab packet.
2. Based on the graph, do you think that the interaction between the orchid and bee is being disrupted as temperature changes? Explain your evidence.
3. Which species would you predict would be more harmed by this disruption? Provide evidence for your prediction.

Reference:

Robbirt K, Roberts DL, Hutchings MJ, Davy AJ. 2014. Potential disruption of pollination in a sexually deceptive orchid by climatic change. Current Biology 24, 2845–2849.