



An interactive learning module for (easily!) teaching ecology students (and professors) about databases for managing and querying large datasets

Becky A. Ball¹, Suzanne W. Dietrich¹, and Don Goelman²

¹School of Mathematical and Natural Sciences, Arizona State University

²Department of Computing Sciences, Villanova University



Purpose

To introduce database skills to ecologists so that they can more efficiently handle large datasets!

- Working with large datasets is an increasingly integral part of ecological research. Students need to be trained in tools & techniques to handle and analyze large amounts of data.
- Most ecologists probably use spreadsheets (e.g. Excel), which is *not* the same as a proper database. They are clunky and inefficient!
- True databases are more efficient, easier to manage, and provide a powerful tool to ask different questions, or “queries”, of that one dataset without constantly changing the spreadsheet (e.g., without re-sorting, copy & pasting, repeatedly typing equations, etc.).
- But it is difficult to teach ecology students about databases, when likely their professors are not experts in databases either.

We present an online tool to introduce students from any major to the basics of databases, from conceptual background to applied design and use.

- The tool assumes no prior knowledge of data science or computing.
- It is web-based, open-source, and free.
- Animations present the conceptual background of databases to teach students how a database is designed and put into use.
- It culminates with hands-on exercises that teach students how to set up and use a database.

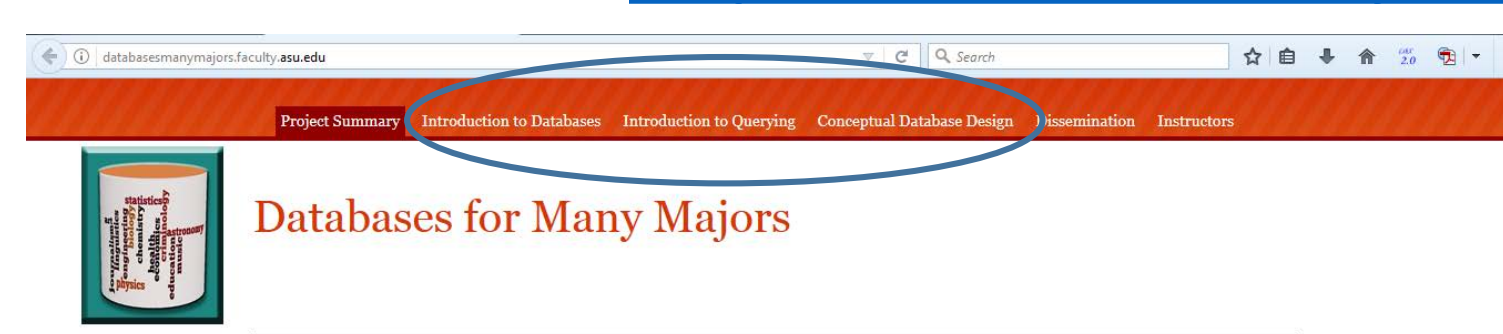
The tool has been customized using an ecological example making the modules relevant to students’ coursework.

- The computer scientists designed a stock set of modules, which were then customized to numerous fields of study.
- In the Environmental Science/Ecology version, we simulate conducting a meta-analysis of data from (fictitious) climate change simulation experiments at Long-Term Ecological Research (LTER) sites.
- Each LTER site has associated characteristics (such as biome, soil type) and has one or several climate change experiments using specific methodology and durations.
- (Presumably there would be numerous columns of associated experimental data, such as effect size of soil C stocks, primary production, microbial biomass, etc.)

Site	ID	Ecosystem	SoilType	Treatment	TreatmentName	Manipulation	YearsDuration
Konza Prairie	KNZ	Grassland	Medisol	+T	Increased Temperature	cabins	5
Konza Prairie	KNZ	Grassland	Medisol	+UV	Increased UV Radiation	cabins	5
Hartshorn National Forest	CVT	Temperate deciduous	Urbisol	+CO2	Increased Carbon Dioxide	towers	3
Hartshorn National Forest	CVT	Temperate deciduous	Urbisol	+T	Increased Temperature	towers	3
Sonoran Desert	CAP	Desert	Aridisol	+P	Increased Precipitation	sprinklers	2
Sonoran Desert	CAP	Desert	Aridisol	+P	Decreased Precipitation	shelters	2
Sonoran Desert	CAP	Desert	Aridisol	+N	Nitrogen Deposition	power	2
Cedar Creek	CDR	Grassland	Medisol	+P	Increased Precipitation	sprinklers	7
Cedar Creek	CDR	Grassland	Medisol	+CO2	Increased Carbon Dioxide	towers	7
Chihuahuan Desert	SEV	Desert	Aridisol	+T	Increased Temperature	cabins	8
Chihuahuan Desert	SEV	Desert	Aridisol	+P	Decreased Precipitation	shelters	10
Chihuahuan Desert	SEV	Desert	Aridisol	+P	Increased Precipitation	sprinklers	10

Using the Introductory Modules

The modules can be found online at: <http://databasesmanymajors.faculty.asu.edu/>



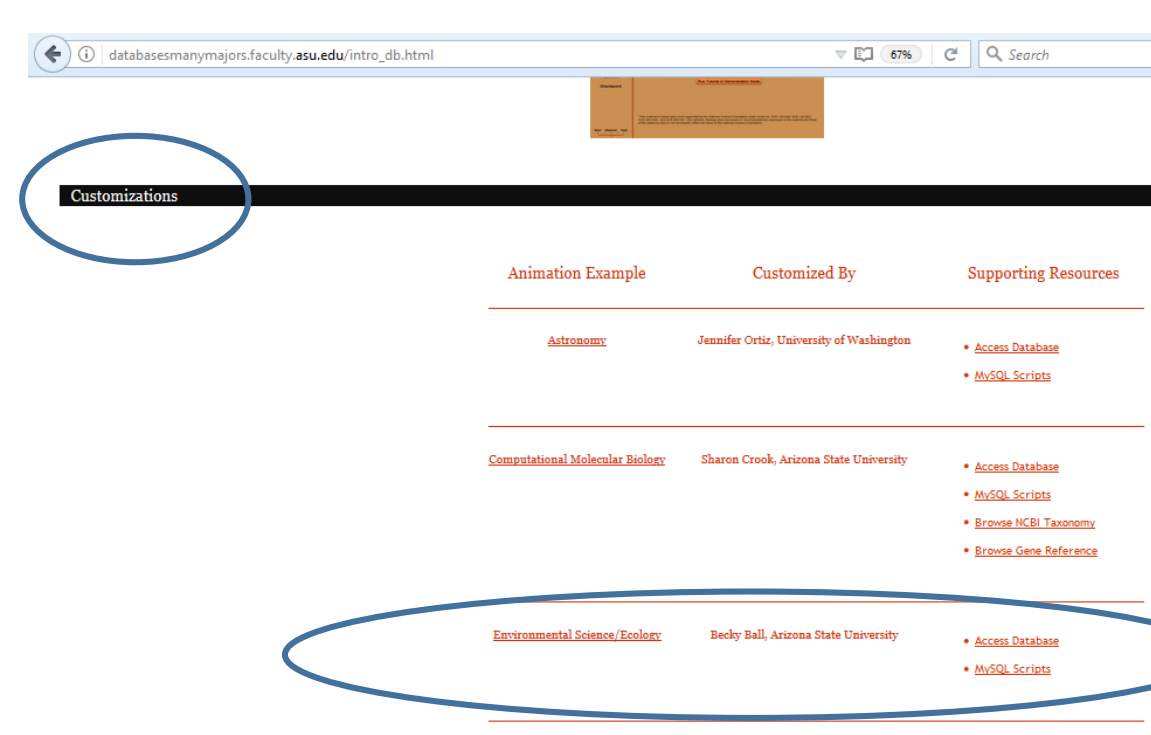
Go to the online module now!



The tabs at the top of the page represent the interactive modules that engage student learning in the three major topics:

- 1) Introduction to Databases:** Promotes the understanding of how a database operates, and how to better utilize the power databases have. Introduces the limitations of spreadsheets in dealing with redundancies, anomalies, missing data, etc.
- 2) Introduction to Querying:** Provides a conceptual introduction to the various operations required to retrieve data from a database to answer a question, and introduces the SQL language to formulate these operations.
- 3) Conceptual Database Design:** Builds on the previous two modules with the introduction to the conceptual design of data, giving an overview to mapping relationships among data entities to design a database based on your dataset.

For each module, you can view either the stock version, or scroll down to the “Customizations” section to select the Environmental Science/Ecology example:



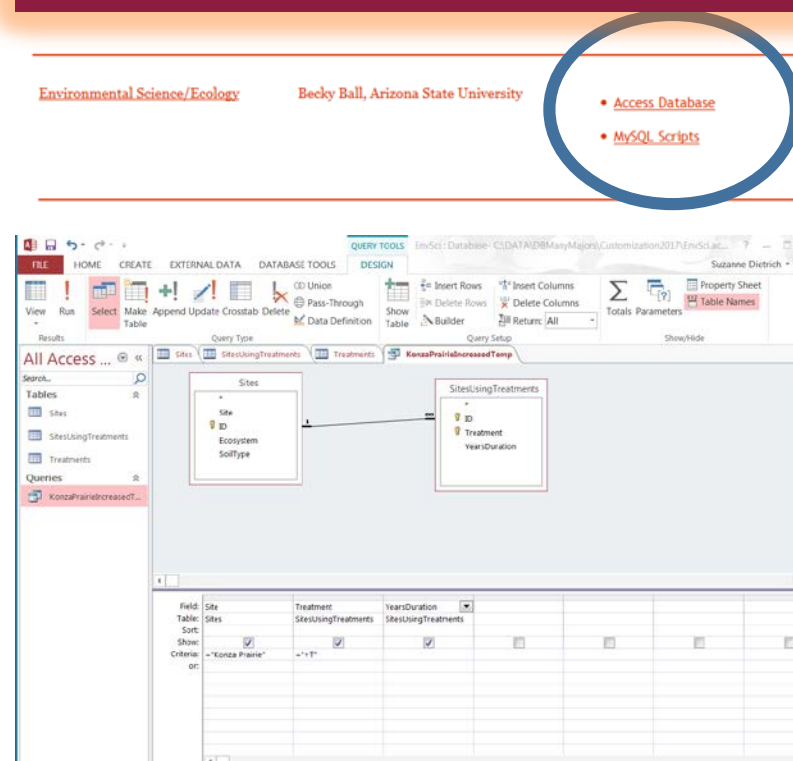
The modules:

- are self-contained and self-paced
- can be completed in about 1 hour
- are targeted at an undergraduate level
- end with self assessment quizzes (referred to as “Checkpoints”) that promote the students’ understanding of the material. If a student answers incorrectly, feedback is given and the question will be asked again later.

We recommend assigning the modules as homework, where students turn in a screen shot of their completed Checkpoints.

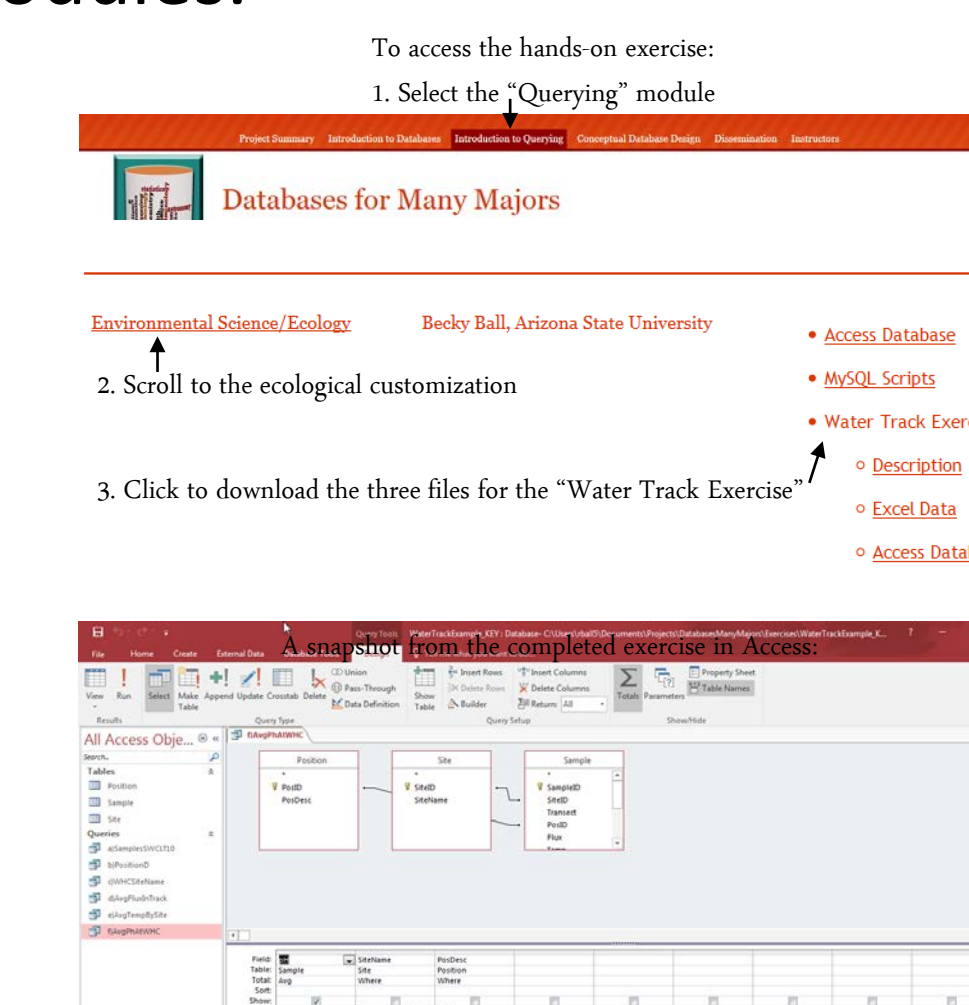
By the end of the three modules, students should know the foundations of databases that will allow them to understand the use of database software such as Microsoft Access.

Using the Hands-On Database Exercises



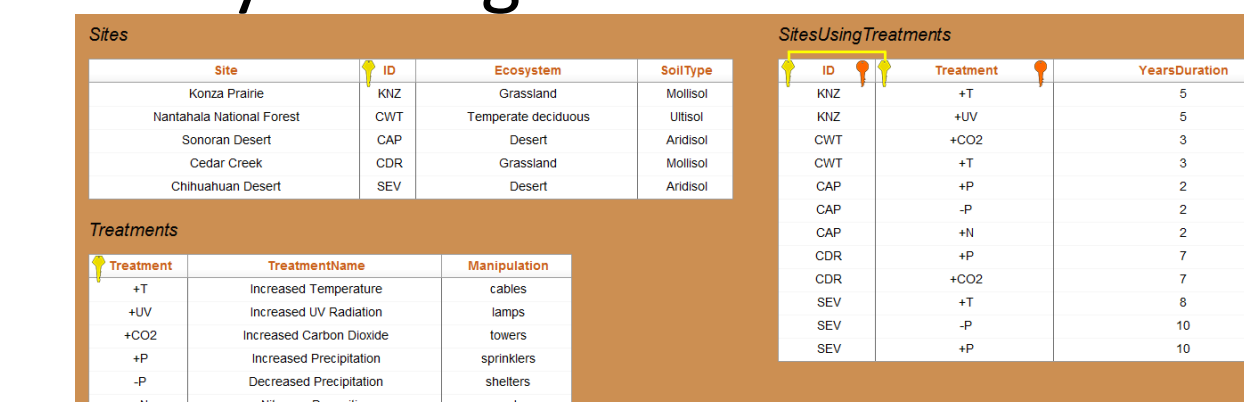
You can find the Access and SQL files of the LTER database alongside the modules, which students can download to work directly with the example database. On the “Instructors” tab, resources are provided to follow up on concepts learned using cooperative learning exercises in class, which can be conducted with the downloaded database (or the stock example provided). This is recommended as an in-class exercise for undergraduates, as follow-up from the online modules.

Additionally, we provide a hands-on Environmental Science/Ecology database exercise using a real dataset. Students download the database containing data from a published field experiment conducted in the McMurdo Dry Valleys of Antarctica. A worksheet walks students through (1) re-enforcing the conceptual information learned in the online modules to design the database and (2) working directly with the data in Access to conduct queries, simulating how a database is actually used for practical data analysis in ecology and developing skills in a common database software. This exercise is most appropriate for advanced undergraduates and graduate students.

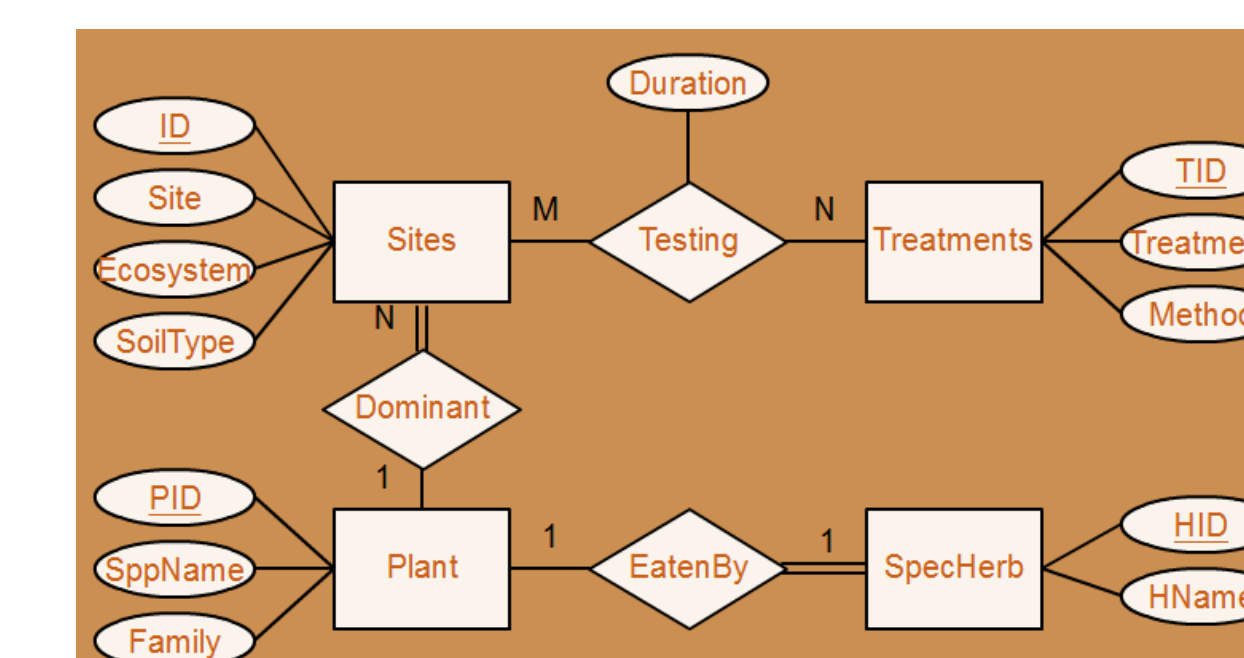


Why databases?

- Spreadsheets (e.g. Excel) are clunky and inefficient compared to a proper database (e.g. Access).
- A database breaks the dataset down into tables, which are more easily managed.



- The content of the tables are linked, according to your design.



- Adding, modifying, or deleting data, and other data “anomalies”, are more easily managed. As a simple example, if we realize the SEV experiments were conducted on Entisols rather than Aridisols, we change it once in the database table, rather than correcting every SEV entry in the spreadsheet individually.
- Also, conducting basic analyses on the data are easily conducted in Access using drop-down menus for “queries”, rather than repeatedly reorganizing the datasheet and typing equations in Excel.
- This early exposure and introduction to database skills allows our students to become more efficient data-handlers than many of their ecology professors!

Contact

For questions, please contact:

The computer scientists that designed the modules:
Suzanne W. Dietrich (dietrich@asu.edu)
Don Goelman (don.goelman@villanova.edu)

The ecologist that customized the ecology example:
Becky Ball (becky.ball@asu.edu)

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DUE-1431848/1431661. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.