**A. Expected Data Types**

Data generated from this project will consist of quantitative phenotypic data on yeast-bacterial and yeast fungal interaction (Objective 1). It will also include quantitative phenotypic data on IAA production, ACC deaminase activity, and phosphatase activity (Objective 2). Finally, it will include genomic DNA sequence data and associated phylogenetic and biosynthetic gene cluster prediction data (Objective 3).

***Common data types*** *to be captured include:*

* *Interaction data with yeasts:* Data will include growth performance with and without bacteria and fungi. It will also include growth performance of bacteria and fungi in the presence and absence of yeasts.
* *Plant growth promotion phenotypes:* Data will include numerical data on the quantity of IAA produced, the amount of ACC deaminase and phosphatase activity.
* *Genomic data:* Data will include raw sequence data and assembled genomic data. Additionally, genomic data will include phylogenies, and the number and types of biosynthetic gene clusters present in yeasts.

**B. Data Formats and Standards**

***Genetic*** consist of raw unedited genomic sequences in .fastq format. The project’s Data Manager will then compile, format, and save material as **.fastq (genomes) files** so they can be more easily stored and accessed on a national repository. Supporting quarterly contextual reports will be initially saved and converted to **.TXT files**. ***Phenotypic data*** will be recorded as colony or population measurements and will be recorded in **Excel**. Content will then be formatted and saved as .**CSV files**. The .CSV files will then be uploaded to a shared lab database and stored on external hard drives. The progress made in each study will be recorded at similar time “capture points” across the 24-month investigative period. The Data Manager (PI Noel) will review content to ensure that timelines from each study lab are similar. Results will be compiled into an **Excel file**, then converted to a .**CSV file**.

**C. Data Storage and Preservation**

Auburn University has contracted with ***Box***, a commercial data storage vendor, to provide dedicated storages space specific to Auburn-based research activities. Additional cloud-based storage will be contracted to allow all research sites to access a common project data storage account. Raw sequence data and time series viability data will be backed up on external hard drives as part of routine weekly backups. Long-term information storage will be archived on Auburn’s ***AUrora University Scholarly Repository***, which is part of AU’s general access library system. This project's duplicate data files will also be stored on secure servers maintained by the ***Alabama Cooperative Extension System*** at Auburn University. A dedicated (full-time) **Data Manager** will be hired explicitly for this project and manage and prepare all collected data for storage and subsequent public sharing. AU’s IT staff will assist in unloading and compiling larger sets of data material. AU’s IT staff will also maintain this data for five years beyond the project’s completion date.

**D. Data Sharing, Protection, and Public Access**

Raw data will be secured on a password-protected cloud-based server (***Box***). Data will only be shared among investigators and their staff as necessary to facilitate project completion to protect the confidentiality and proprietary interests. Shared data will include provisions for appropriate protection of intellectual property or other rights of Auburn University and collaboration institutions. At the project’s conclusion, data will be disseminated widely through academic publications. ***Ag Data Commons* will be the national public repository to post and share time-series viability data generated from this project. This** USDA-sponsored repository uses customized open-source DKAN software and complies with all US Project Open Data standards for federal agencies. Additionally, once materials are successfully uploaded, reviewed, and accepted for posting on *Ag Data Commons*, the service automatically generates a *Digital Object Identifier* (DOI) number to support the digital tracking and citation of contributed content. ***NCBI’s GenBank and the Sequence Read Archive*** will be the national public repository used to post and share genetic and genomic sequencing data generated from this project once the data has been analyzed for publication. This will facilitate public dissemination of the data and analysis by other investigators.

**E. Roles and Responsibilities**

**Dr. Zachary Noel (Principal Investigator). Dr. Noel will be responsible for ensuring that project data and supporting content are transferred promptly to both AU’s cloud-based project storage site, onto *Ag Data Commons* national repository, and to *NCBI’s Genbank and Sequence Read Archive (SRA)*. A designated Project Data Manager will assist Dr. Noel hired for this project. They will compile, catalog, and format all project data for upload onto Ag Data Commons, Genbank, and SRA. Dr. Noel’s genomic and metagenomic analysis expertise will guide the project and data management to success.**

***Additional data collection support will be provided by:***

**Bernardo Chaves-Cordoba (Director, Statistics Teaching & Consulting, College of Agriculture) will serve as a project consultant and assist in preparing data for public sharing. Finally, Ali Krzton (AU’s Research Data Management Librarian) will provide data management training to all research team members and project-affiliated graduate assistants. Ms. Krzton will also conduct annual audits over the project’s 2-year length to ensure data content is preserved correctly and made publicly accessible.**