AG DATA SERVICES
NEW FACULTY RESEARCH
DATA SUCCESS PROGRAM

Ag Data Services is a team of data professionals that provide specialized expertise, tools, methodologies and services to faculty, staff, and graduate students who work in Agriculture research. ADS is continuously compiling and developing best practices for data stewardship and analysis, as well as working with a variety of IT partners and other service providers to establish reusable infrastructure components. ADS is part of College of Agriculture Information Technology department and sponsored by Agricultural Research and Graduate Education.

How will ADS help? An Ag Data Services (ADS) data professional will work closely with a faculty member and their research team to think strategically about their data, guide them down the path to being good stewards of their data, and show them how to leverage campus resources to process data using best practices. Through this partnership, ADS will use informal methods to teach key data-related skills and best practices that are most relevant to the faculty member's research program. These efforts will give the faculty member a jump start on their early career research, and set them up with foundational research data best practices for greater long term success.

agdatateam@groups.purdue.edu

ag.purdue.edu/arge/Pages/Ag-Data-Services.aspx

Each faculty engagement will be tailed to their specific circumstance and needs. Some examples of deliverables that can be completed during this engagement with ADS include the following:

1. Strategic Data Plan (SDP) -
Early in the faculty's first year, ADS will spend 20 hours meeting with the faculty member and their research team to understand their research goals and develop a holistic five-year SDP for optimizing data production, storage, and analysis. The document will capture anticipated opportunities and challenges, and include strategies for addressing them that include partnering with various research support groups and leveraging campus cyberinfrastructure. The goal of this brief document is to describe key research workflows so that they can be optimized across the lab for reusability and repeatability, and to implemented practices early on that will scale as the research program grows. The result will be a shared understanding of high-level goals for handling data throughout its lifecycle, along with the best people and computational resources on campus to support their research goals.

2. Data Stewardship Plan (DSP) -
In the first half of the year, as the data strategy plan is being finalized, ADS will spend 30 hours working with the lab on creating and documenting a data stewardship plan. In addition to policies and procedures commonly found in a data management plan, the DSP addresses the human side of the things by explaining why the policies and procedures are in place, and defining roles and responsibilities for data stewardship. ADS will explore the workflows of individual researchers to identify existing best practices that can be shared across the lab, work with the group to create a shared plan for data stewardship, and develop best practices for implementing the plan. The goal of this effort is to produce a light-weight, practical document that explains what's expected, how to do it, and why it's being done. The outcome of this process will be a research lab that creates high quality and integrity data that is findable, accessible, interoperable, and reusable (FAIR).
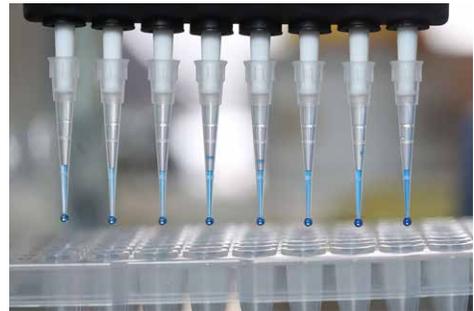
3. Data pipeline proof of concept (POC) -
In the second half of the year, as the data stewardship plan is taking shape, ADS will spend 80 hours working side-by-side with a researcher to create a data pipeline proof of concept. Data pipelines consist of a series of modules that manipulate data to accomplish a large portion of a research workflow such as data ingestion or analysis. The researcher will work with an ADS data professional to develop interoperable scripts that are well documented, human readable, and versioned. The goal of this effort is to script several modules that handle real research data and accomplish key pieces of the research workflow. In addition to providing an opportunity to test and refine the data stewardship plan, this effort will accelerate the faculty's capability to effectively and efficiently use their data to answer research questions.