PARTNERSHIPS FOR DATA INNOVATION (PDI) COTTON – TOOLS AND DASHBOARDS

C.D. Delhom
H.A. King
USDA-ARS
New Orleans, LA
C.B. Armijo
USDA-ARS
Mesilla Park, NM
M. Ulloa
J.R. Mahan
P. Payton
USDA-ARS
Lubbock, TX
T.G. Teague
UADA-ASU
Jonesboro, AR
E.F. Hequet
Texas Tech University
Lubbock, TX
E.M. Barnes
N. Kothari
R.W. Kurtz
G. Morgan
Cotton Incorporated
Cary, NC
B. Vandenberg
L. Messinger
G. Zoch
USDA-ARS
Ft. Collins, CO
M.D. Buser
USDA-ARS
Beltsville, MD

Abstract

Partnerships for Data Innovation (PDI) is a collaborative effort sponsored by the U.S. Department of Agriculture, Agricultural Research Service (ARS) to develop customized data management solutions for agricultural researchers. The effort relies on achieving data standardization, integration, and automation through innovation and public-private partnerships between ARS and universities, commercial partners, and industry stakeholders, such as Cotton Incorporated. PDI Cotton efforts have focused on standardizing cotton data terminology, developing data collection tools for use in the field by researchers, and developing dashboards and data repositories to store and visualize cotton trial data. Presentations related to the PDI Cotton initiative were given in numerous conferences across the Beltwide Cotton Conferences, including Agronomy, Engineering, Improvements, Insect Research, and Quality Measurements. Each presentation focused on the tools and efforts specific to that audience. The PDI effort is being designed by researchers from the bottom up with support from programmers and GIS specialists. Data collected with PDI tools will be stored within the Agricultural Research Outcomes System (AgCROS) and allow for integration with historic cotton data and other databases such as soil and weather information. The researcher controls access, and customized dashboards allow for the rapid visualization of data while also allowing for creating public-facing dashboards for the rapid dissemination of data to stakeholders.

Introduction

Agricultural research, and cotton research, in particular, is often attempting to solve problems that are highly complex with many interactions which confound results. Developing cotton cultivars based on genomics that are more resistant to biotic and abiotic stresses is important for a competitive and sustainable cotton industry; however, the fiber quality

is also impacted by harvesting and ginning, which are not always readily quantified in experimental data. PDI's backbone is the Agricultural Research Outcomes System (AgCROS), a network of networks and based on geospatial solutions, data visualization, standardized data dictionaries, and standardized data entry templates. The use of a geoinformatics approach to AgCROS allows for a "one-stop-shop" for data storage, dissemination, model development, and collaborations (Figure 1).

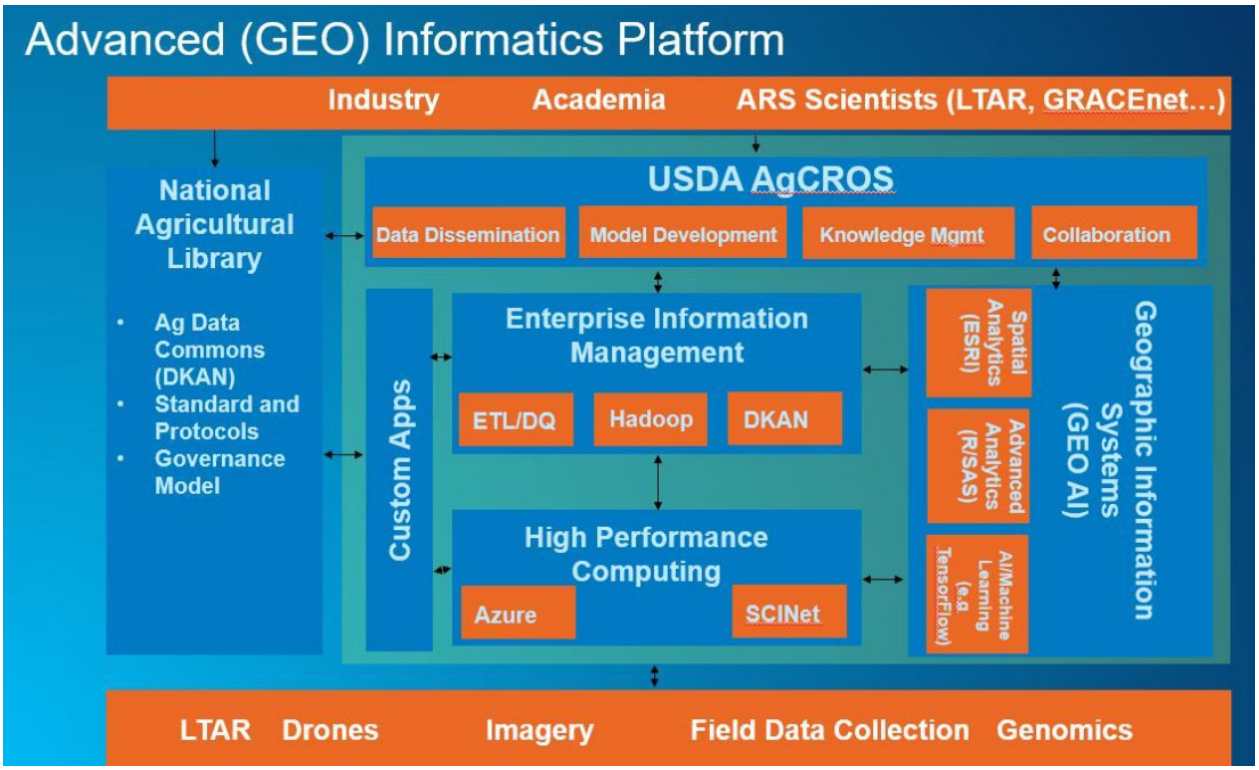

Figure 1. AgCROS geoinformatics approach (courtesy of ArcGIS, 2021)

The PDI Cotton efforts began in mid-2020 with four inter-related efforts: 1) Data Standardization, 2) Variety Trial Data Collection Tool, 3) Agronomic Trials Data Collection Tool, and 4) Cotton Dashboard Development. The four efforts were spearheaded by teams representing industry stakeholders, university researchers, and ARS scientists, along with the support of dedicated GIS specialists and programmers. The efforts are designed by researchers and built out by specialists, which frees the researchers from handling the work's information technology side.

A guiding principle of PDI is to assist researchers by improving efficiency in the research process. Improving the efficiency of research is being addressed by eliminating data silos and reducing non-productive data touches, which scientists often perform. In a traditional scenario, data observations are written down in the field and later transcribed into electronic form. However, with PDI data collection tools, field observations and laboratory data will be seamlessly brought together with minimal effort on the investigator's part. Likewise, traditionally a researcher must spend countless hours searching for historical results or datasets such as soil and weather for the growing locations, which may be needed to analyze the results correctly, or when made easily available, found to be valuable co-factors in the statistical analysis of the experiment. PDI will seamlessly pull these existing datasets together, allowing for faster and more informed analysis of experiments.

Data Standardization

For data to be housed within AgCROS, the data must be standardized. Data standardization includes developing a list of standard variable names, table assignments, and data variables unique to cotton. Data stored within AgCROS will be stored in SI units but can be collected or reported in various predetermined systems of measurement. The critical components of the data standardization efforts are the Cotton Data Dictionary and the Data Entry Template.

Cotton Data Dictionary

The Cotton Data Dictionary is a standard set of terminology. A diverse group of stakeholders is developing the Cotton Data Dictionary to ensure that data across various cotton-related projects can be collated and compared. Previous efforts, such as the Cotton Management Expert System Software, COTMAN, (Oosterhuis and Bourland, 2008), CottonGen (Yu, J. et al., 2014), and various state and national variety trials are being used as the starting point for developing the necessary terminology. The Cotton Data Dictionary includes descriptions of each parameter, relevant units, and where it belongs within AgCROS and provides data validation. As part of the dictionary, the data validation rules allow real-time quality assurance and quality control (QA/QC) to minimize data errors. The data dictionary provides for parameters to be defined by aliases and "also known as" to map common variable relationships across various research fields.

Data Entry Template

The Data Entry Template (DET) allows the automated upload and storage of data using the data dictionary to ensure data goes into the proper tables within AgCROS. Although the data collection tools are the primary method for uploading data to AgCROS, the DET allows for the large-scale importation of historic data or data collected on standardized forms and spreadsheets. Programming support within the PDI team assists with creating a DET where that is the optimum method of importing data.

Cotton Data Collection Tools

The PDI effort is based on geoinformatics, and thus all data is related to a geospatial frame of reference, such as the location in which the cotton was grown. Software from ArcGIS underpins most of the data collection and visualization tools. Survey123 is a form-centric solution for capturing data through survey-like tools (Esri, 2021b). The PDI team can rapidly develop custom Suvey123 applications to collect data via mobile devices or desktop computers. Data can be collected on a mobile device while in the field, even without internet connectivity, and is stored locally until connected to the internet, where it is securely stored directly into the PDI cotton data repository. The GPS sensor in the mobile device can automatically identify the plot location for field data and automatically use other device settings to record the data collection time and date of the data collection. Observational data usually collected via pen and paper can be directly collected with the data collection tools. The information is stored locally in real-time, so no data is lost if the device loses power.

The cotton surveys allow the user to set up individual experiments, including selecting which system of measurement the user desires for specific experiments and assigning a unique identification code for each experiment. Contact data and general metadata is configured so that the investigator and experiment are correctly identified. The user can then select what type of data will be entered during each use of the survey. At the beginning of the experiment, the experimental treatment information is set up. During the experiment, the user or their authorized delegate(s) can choose to update field management or in-season measurements and observations. During this initial setup, the baseline field and planting conditions can be entered to include information such as tillage, row spacing, seeding rate, and plant spacing. Images have been incorporated into various parts of the survey to ensure consistency. For example, when entering tillage information, pictures of different tillage types can be displayed if the user is unsure of the proper answer. Users can also take photos with their mobile devices and include the images with the data.

Variety Trial Data Collection Tool

One of the first data collection tools developed by the PDI Cotton team was to collect variety trial data. Variety trials are generally simple in design, with the variety and growing location as the most likely variables, while field treatment is consistent across the plots within a location. This survey is designed to provide a streamlined interface that minimizes the amount of interaction the researcher must have with the survey tool. The harvest and ginning method can be recorded in the survey and applied to all varieties in a given location, easily capturing data that often goes unreported in various trials.

Agronomic Trial Data Collection Tool

Agronomic trials can be much more diverse than variety trials and represent a more significant challenge for the PDI team. Agronomic trials may be investigating a single variable or multiple variables as part of a complex study. They may examine one or multiple varieties including other factors such as irrigation, fertilization, pest control, tillage, or other production or protection issues. These trials may be small-scale or large-scale and may include various harvest and ginning methodologies as well. The interface of the Agronomic Trial Data Collection Tool is similar to the

Variety Trial Data Collection Tool but consists of a broader range of options, configurable by the end-user, to collect data in a more complex trial. In all cases, the user can configure the tool to display only relevant data fields.

Cotton Data Dashboards

The cotton data is readily conveyed to researchers and stakeholders via customized dashboards using ArcGIS Dashboards by ESRI (ESRI, 2021a). These dashboards allow for location-based analysis and trend visualization. The PDI team can readily customize dashboards to meet specific needs. Example cotton dashboards are under development using various historical data such as a public-facing dashboard containing data from the National Cotton Variety Tests (NCVT) (Figure 2).

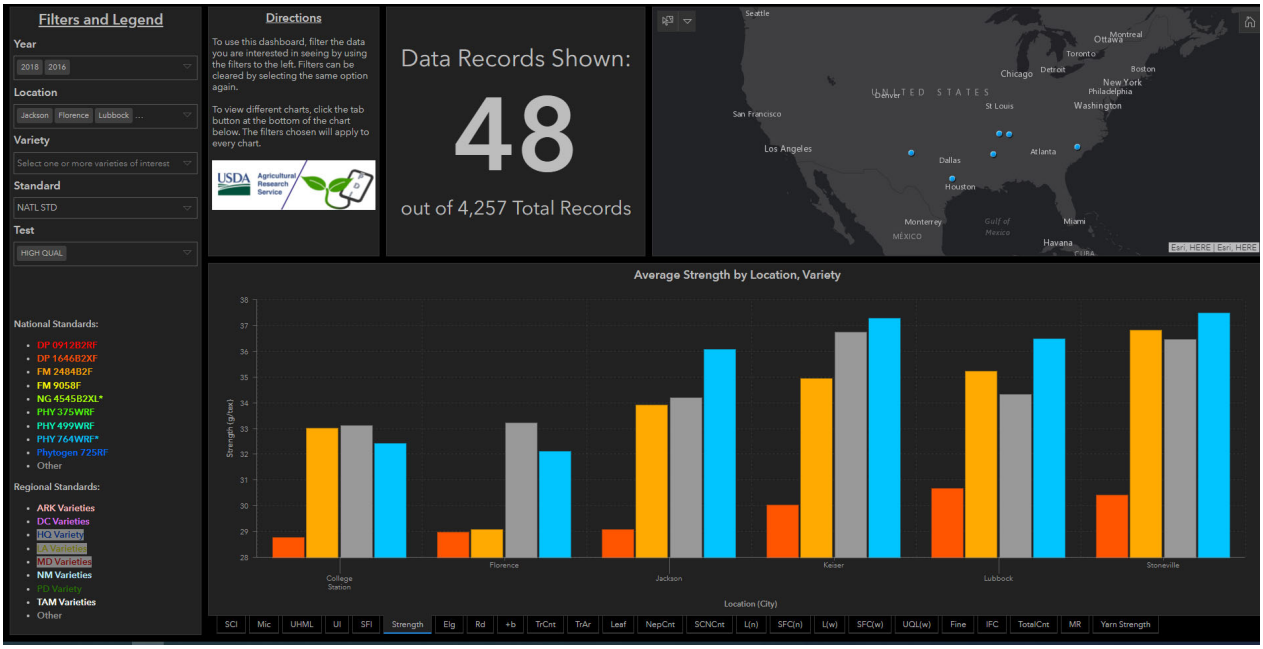

Figure 2. Screenshot of the NCVT Cotton Dashboard

The dashboard has several regions of interest. On the upper left is a series of drop-down menus, allowing a user to filter the data by crop year, location, variety, and other factors. In this example, data is being filtered by crop year (2018 and 2016) in six locations (College Station, TX, Florence, SC, Jackson, TN, Keiser, AR, Lubbock, TX, and Stoneville, MS) and only for the national standards grown for the NCVT. The number of records found meeting the criteria is updated in real-time (48), and the map highlights the locations where data meeting the query has been found. Fiber quality traits from the NCVT are plotted at the bottom of the screen, and the user can select the traits they are interested in observing from the various tabs on the screen.

In this way, the user can readily observe trends across locations and variety. The user can hover their cursor over a specific bar in the plot, and the screen will display the location, variety, and specific value of the fiber property. Figure 3 shows a closeup of the fiber quality plot for fiber strength from Figure 2 and reveals that fiber strength for PHY 764WRF (light blue) ranges from a low of 32.1 g/tex in Florence, SC, to a high of 37.5 g/tex in Stoneville, MS while DP 1646B2XF (orange) ranges from a low of 28.8 g/tex in College Station, TX to a high of 30.6 g/tex in Lubbock, TX.

In the future, plans are for researchers to be able to export selected data from the dashboards for further analysis. Additionally, the dashboards will be integrated with additional data from both AgCROS and other datasets, allowing investigators to explore the harvest and ginning methods and weather and soils data.

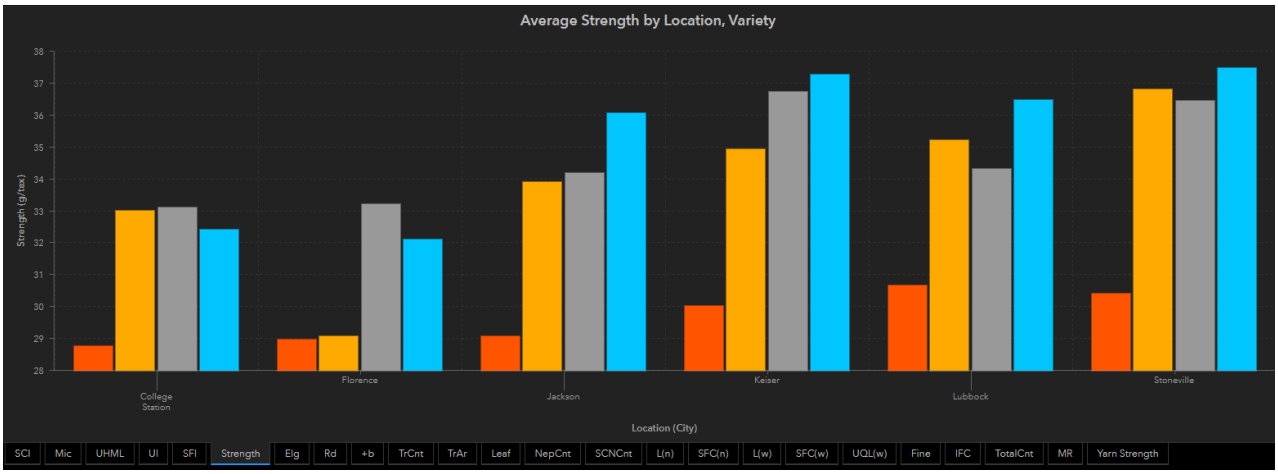

Figure 3. Close up of fiber strength plot for four varieties in six locations from Figure 2.

Future Plans

The PDI Cotton effort began in the second quarter of 2020 and is developing quickly. The Survey123 cotton data collection tools for variety and agronomic trials have undergone multiple revisions, and a beta version will be deployed to several researchers for the 2021 crop year. Currently, six years (2013-2018) of fiber quality and yarn data from the NCVT is available on a public-facing dashboard. Seed and yield data is being added for those years, and the dashboard will be updated as the data is entered. Upon completion of the 2019 and 2020 NCVT data, which is delayed due to COVID-19, that data will be added to the dashboard. Additional historical data from the NCVT will be added to the dashboard as it is digitized and formatted.

The beta version data collection tools and dashboards may be accessed via the PDI Cotton effort storymap on the USDA AgCROS exploration page (https://agcros-usdaars.opendata.arcgis.com/).

References

ArcGIS. 2021. USDA ARS Collaboration network – A story map. https://www.arcgis.com/apps/MapJournal/index.html?appid=b3a1780ffda04dc2a4e0395a491f5bf3 [Accessed 19FEB21].

ESRI. 2021a. ArcGIS Dashboards. https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview [Accessed 21FEB21].

ESRI. 2021b. ArcGIS Survey123. https://www.esri.com/en-us/arcgis/products/arcgis-survey123/overview [Accessed 19FEB21].

Oosterhuis D.M. and F.M. Bourland. 2008. (eds.) COTMAN Crop Management System, University of Arkansas Agricultural Experiment Station, Fayetteville, AR. pp. 107. http://www.cotman.org

Yu, J., Jung, S., Cheng, C-H., Fickling, S.P., Lee, T., Zheng, P., Jones, D., Percy, R.G., and Main, D. 2014. CotonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 42, D1229-D1236.