Document updated on November 10, 2020

Fieldprint Platform Data Quality Analysis Tool

Description

Field to Market staff have developed an interactive, web-based Quality Analysis (QA) Tool to assist project administrators, project specialists, data entry personnel, and consultants with reviewing their data inputs and results obtained via the Fieldprint Platform from registered Field to Market projects. The QA Tool supports the ability to detect outliers, find errors, and gain insights from the project data and results. As with any modeling approach, project reporting output is only as good as the data inputs. A robust quality analysis by project administrators, specialists, or other personnel enhances confidence in the Fieldprint Platform data.

Version 3.0 of the Fieldprint Platform incorporates various mechanisms to improve data quality. Most are focused on preventing data input errors via the user interface. For example, set data input boundaries in the Platform prevent growers from entering excessive (out of range) values for inputs of high impact, such as maximum values for fertilizer amounts and crop yield. The Platform also includes warnings or prompts to alert growers when values might be incorrect. Where necessary, the Platform provides users with contextual help to address ambiguity. However, despite these efforts, it would be impractical to assume all data quality issues can be mitigated. Those charged with reviewing project reporting data should anticipate the need for data quality review.

Project administrators and specialists have access by permission to project reporting tools within the Fieldprint Platform. The primary project reporting product is the Microsoft Excel-based Comprehensive Data Output File. For each grower connected to the project, the report contains detailed, finalized grower input and output by crop year. The size of the Excel report is dynamic, it has as many rows as crop years, and the number of columns will vary according to the number of trips for the application of fertilizers and crop protectants, along with the number of harvest operations for crops such as alfalfa. This dynamic number of rows and columns could make it difficult to automate Excel processes that depend on columns staying in fixed positions. It can also be cumbersome to find columns of interest within a file with hundreds of columns.

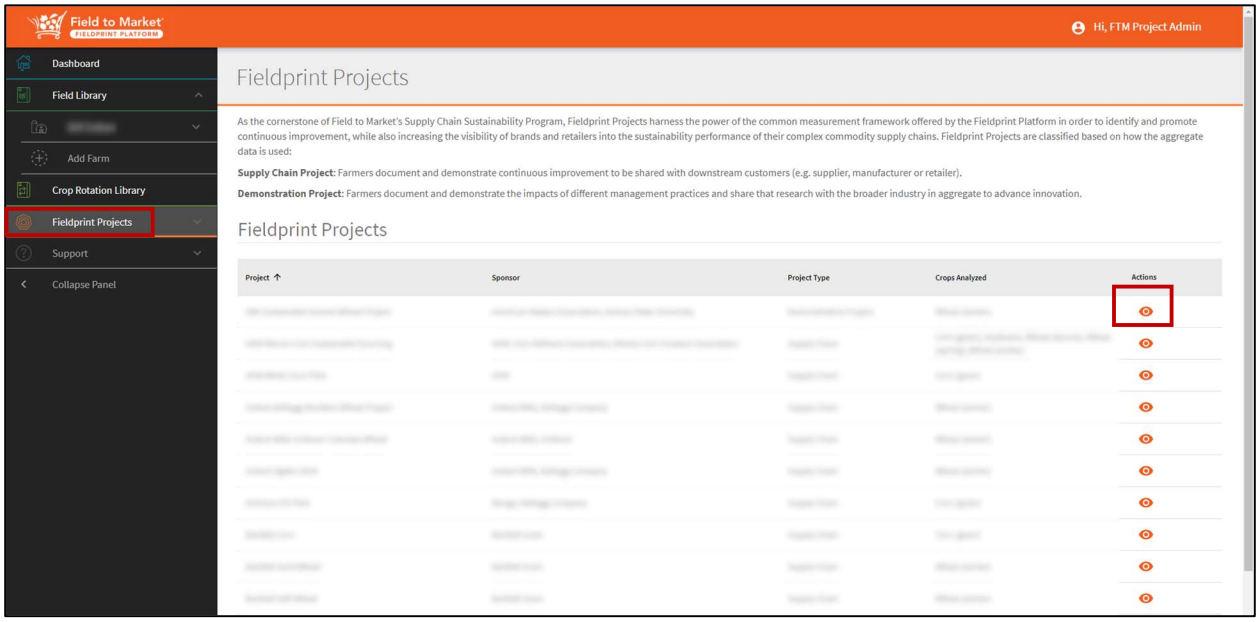
The goal of the QA Tool is to help project administrators conduct quality analysis in a consistent manner across all projects; however, the QA Tool is not intended to replace other data entry quality assurance mechanisms that project administrators implement with growers within projects. We highly recommend thoroughly reviewing the Comprehensive Data Output Report at the end of each growing season before using the data for storytelling, summaries, and official reports.

What do I need to use the QA Tool?

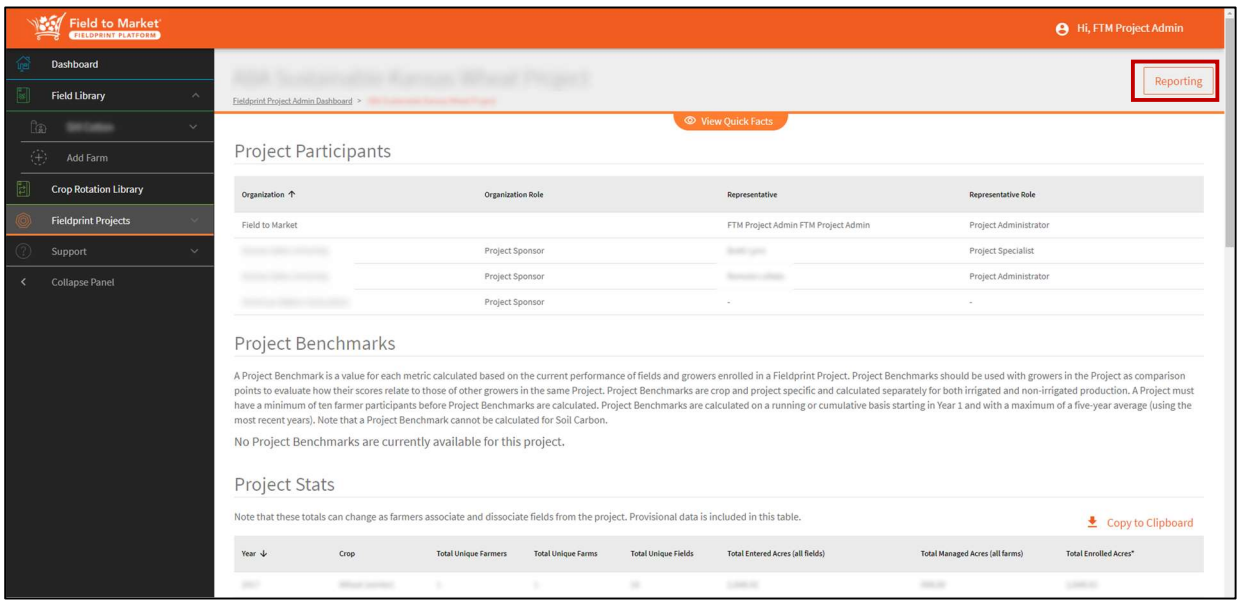
Project administrators and specialists need to export the Microsoft Excel-based **Comprehensive Data Output File** using the project reporting functionality in the Fieldprint Platform. For each grower connected to the project, the report contains detailed, finalized grower input and output by crop year.

How do I export the Comprehensive Data Output File for my project?

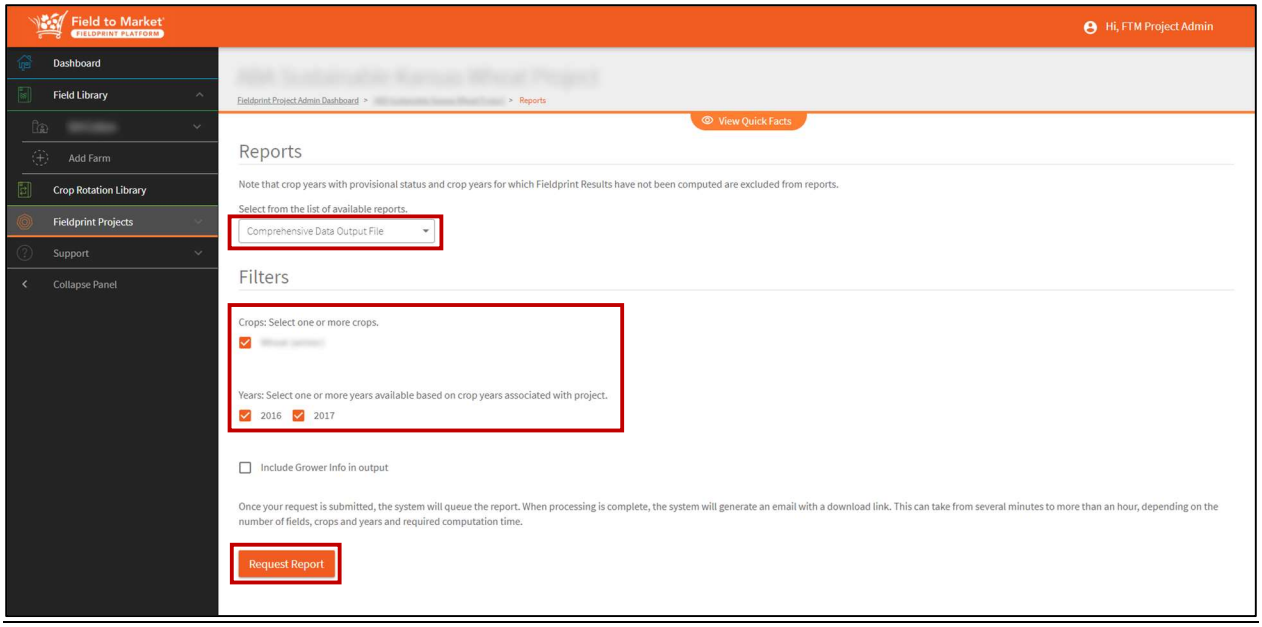
Step 1. Login to the Fieldprint Platform. Find your project under the section Fieldprint Projects and click on the View icon



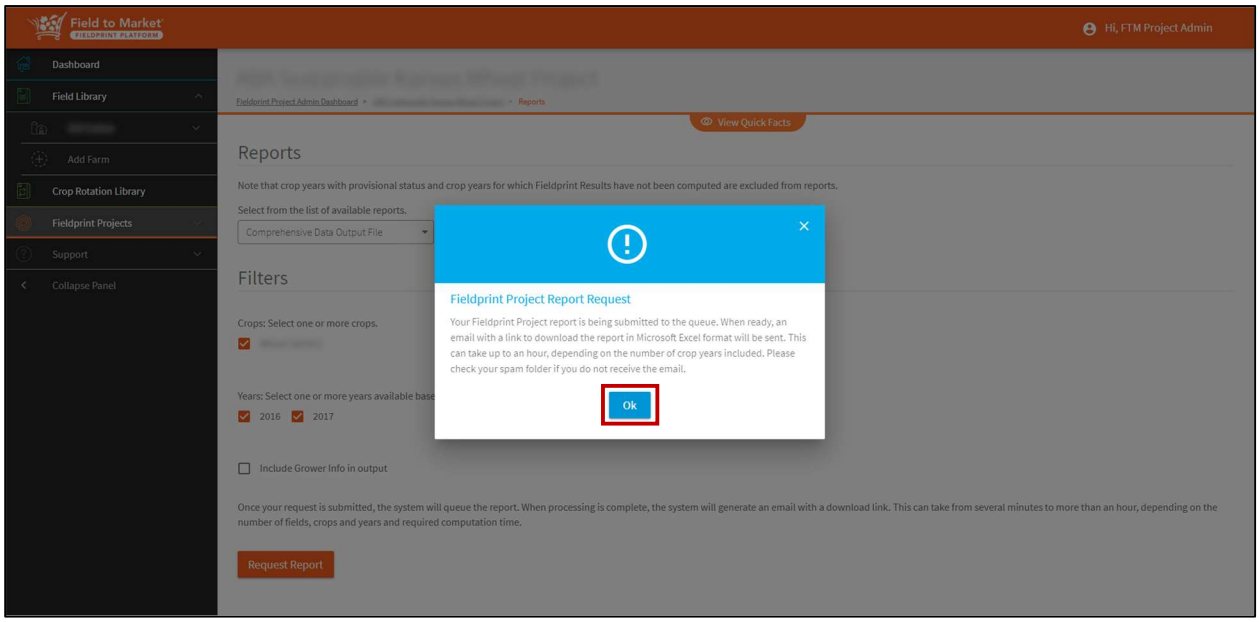
Step 2. Click on Reporting on the top right



Step 3. Select Comprehensive Data Output File for the Crops and Years of interest and click on Request Report at the bottom



Step 4. Click OK. The request was submitted, and a download link will be sent to the email on file



QA Tool details

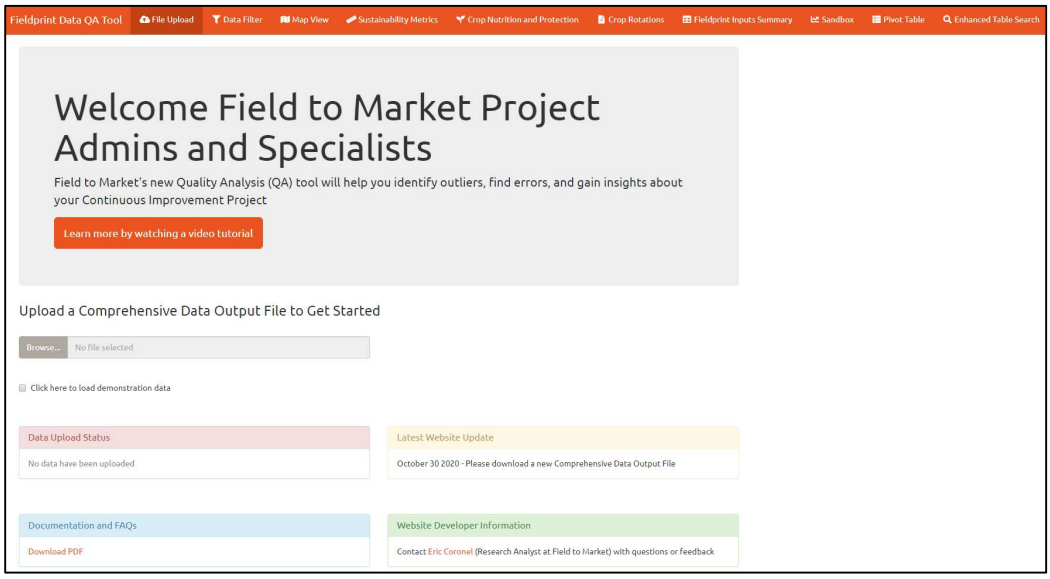
What is the QA Tool?

The Fieldprint Platform Data QA Tool is a stand-alone interactive website. A user needs to upload the Comprehensive Data Output Report to the website and content is automatically generated to explore the project data. The administrator would then be able to explore the location and boundary for each field, the Fieldprint scores for each metric, the amount of fertilizers and number of crop protectants applied, the operations within crop rotation templates, and summary tables for items such as crop yield.

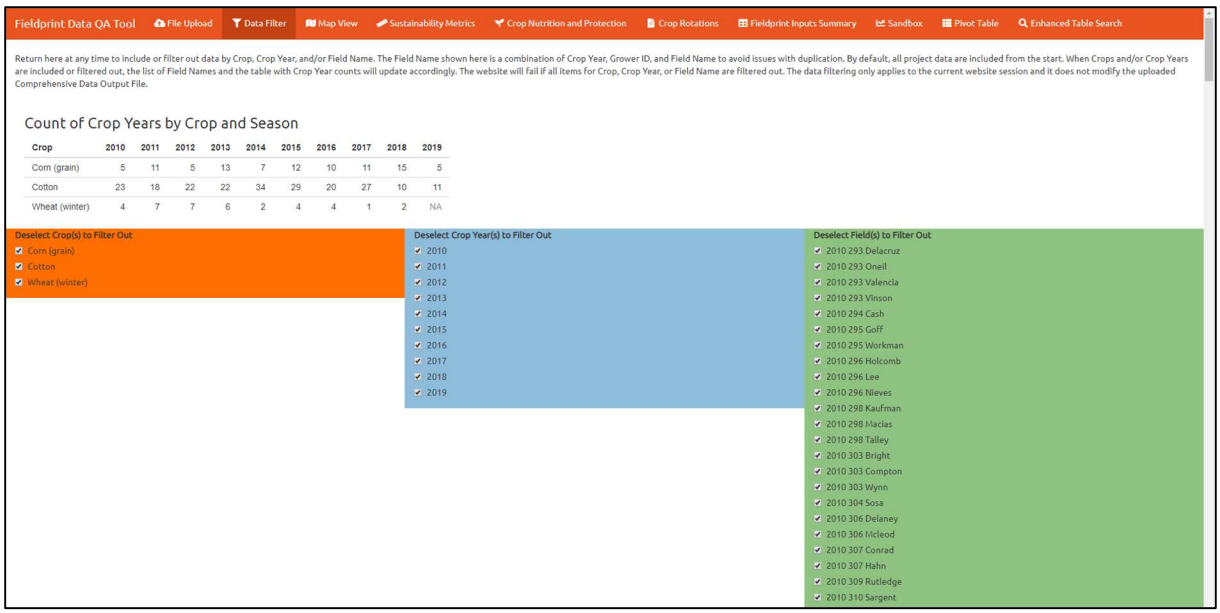
The tool provides scatter plots and boxplots that allow users to compare results for all fields and crop years. These simple graphs are effective for quick detection of outliers. Outliers are data points that fall outside the regular pattern and might indicate an error in data entry. There is also an option to create customized scatter plots and pivot tables to explore the data in a personalized manner.

The structure of the QA Tool is as follows:

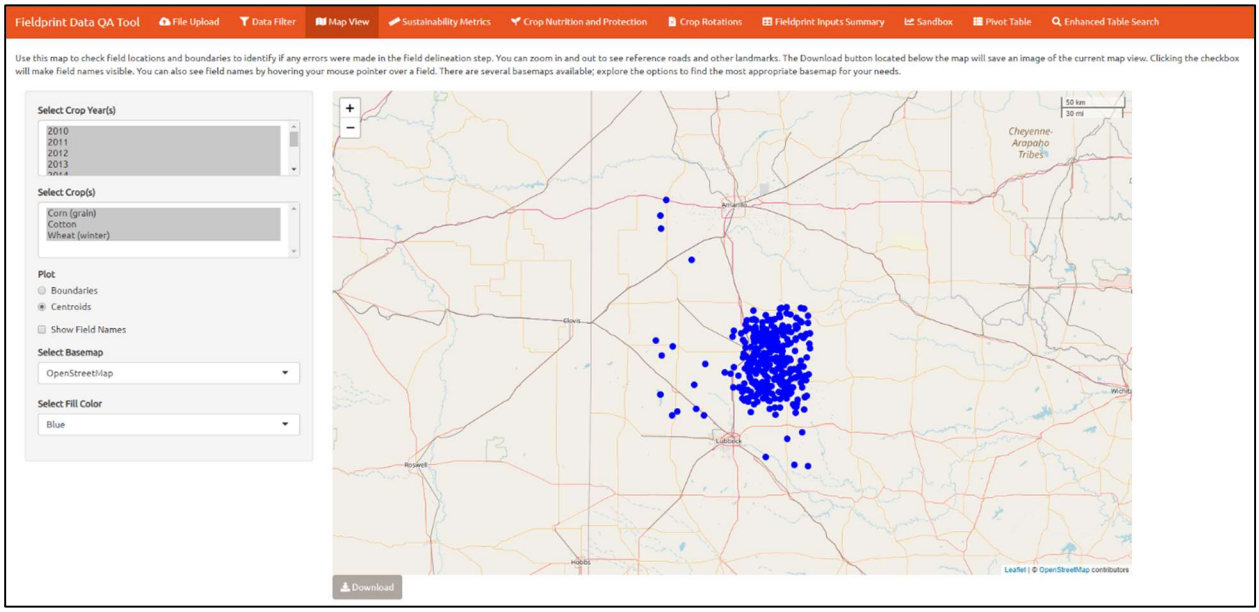
Landing page with file upload input, documentation, and contact information. Click Browse to select the Comprehensive Data Output File from your local storage



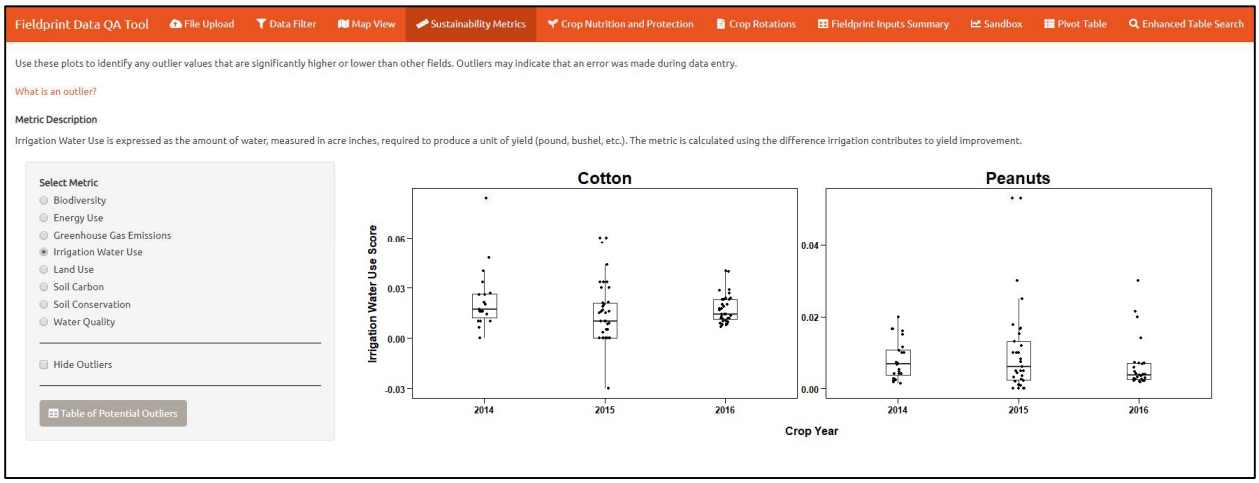
Data filter to temporarily include or filter out entire years, crops, or individual fields to focus on the data of interest



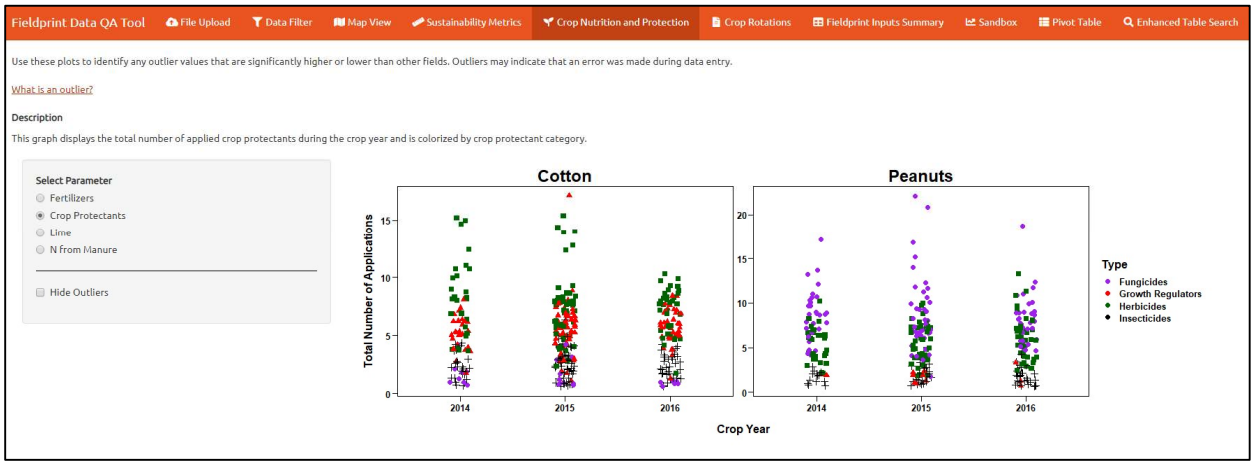
Map view to visualize field locations and boundaries. Please do not print or share maps with visible field names.



Scatter plots and boxplots to visualize sustainability metric scores. Clicking Table of Potential Outliers on the bottom left will open a popup table with suspected scores.

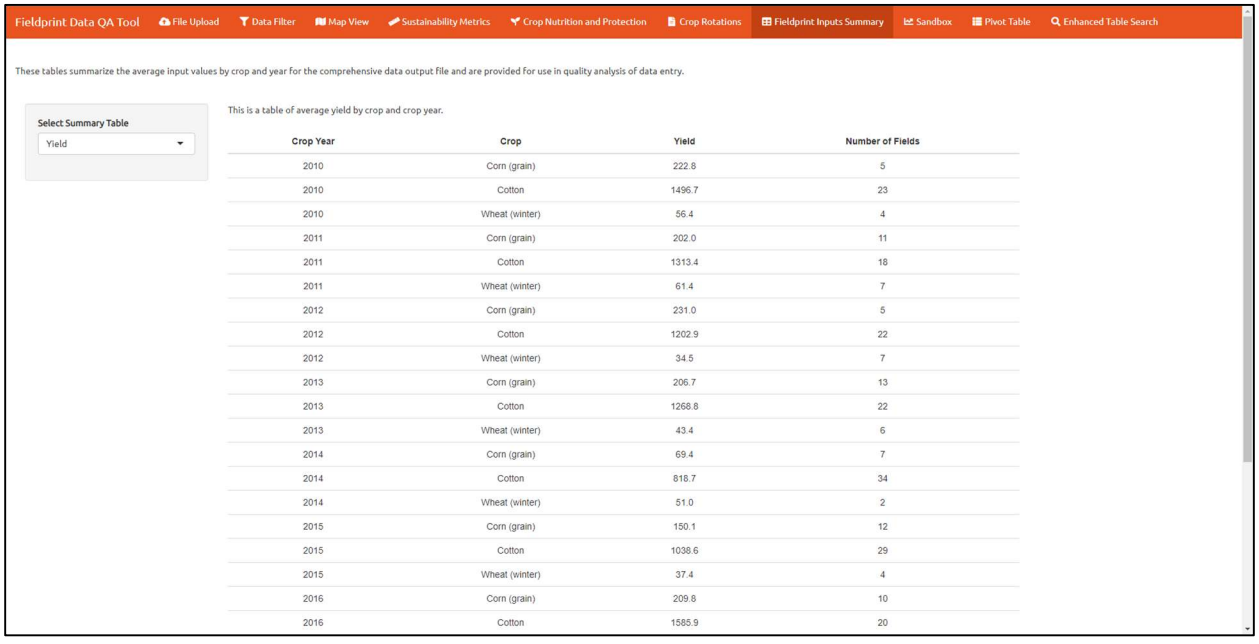


Scatter plots and boxplots to visualize fertilizers and crop protectants. Move through the various options to spot outliers for a given crop or year.

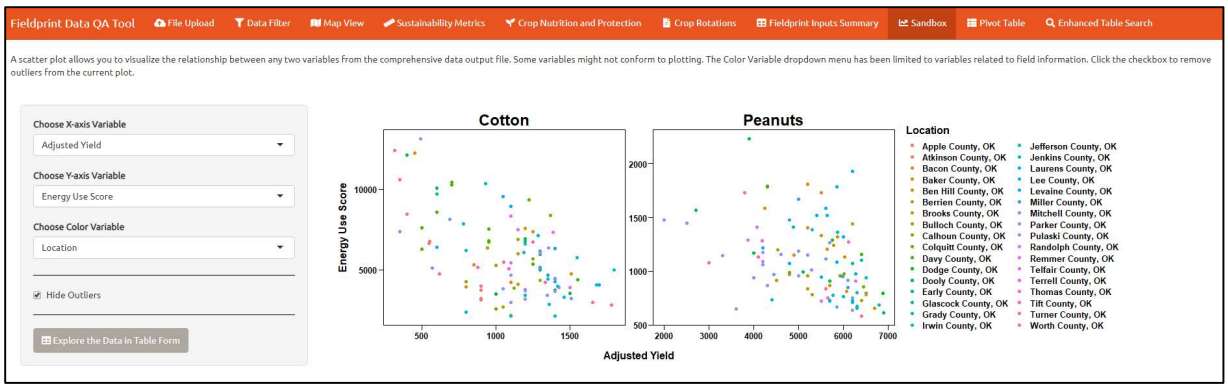


Visualize timeline of rotation operations in graph and table form, see rotation template frequency count, and find which crop years used each rotation

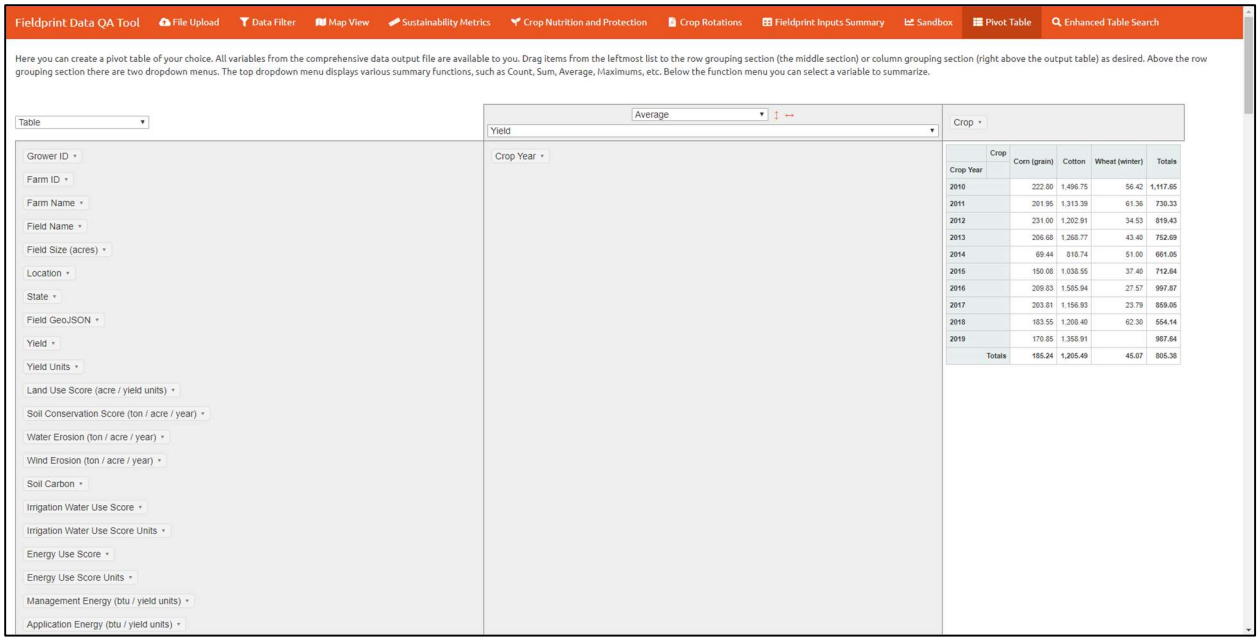
Find project-wide summaries for yield, fertilizers, crop protectants, field cropping history, and grower participation through the years



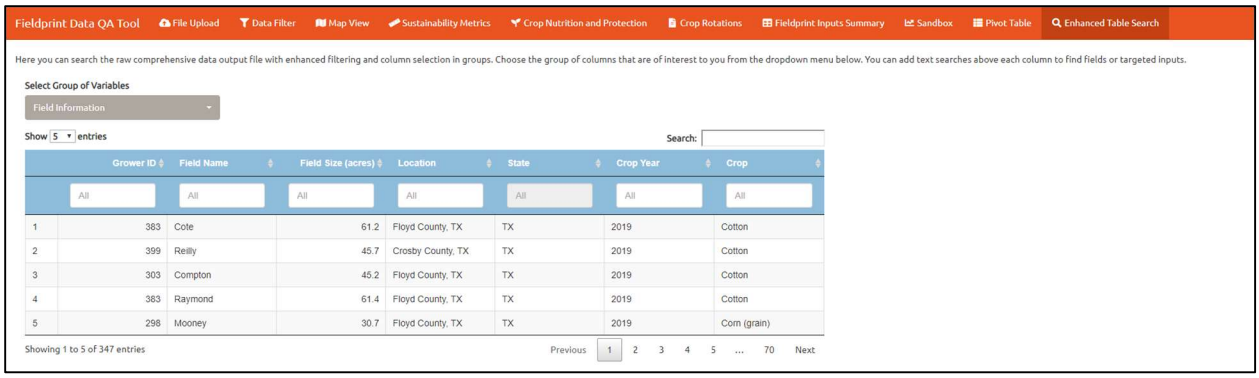
Create customized scatter plots of any two variables



Create customized summaries with interactive pivot table



Interactive table filtering and sorting to visualize raw project data



Frequently Asked Questions

What are appropriate uses of the QA Tool?

- Identify outliers
- Find errors
- Gain insights for internal uses within your organization, such as presentations, reports, and storytelling

What uses of the QA Tool would require approval?

Data insights cannot be used for claims without prior approval from Field to Market (more information about claims further down).

What are the most common data quality issues?

- Excessive N-P-K inputs, along with large N from manure values
- Mistaking fertilizer units of lb/acre and gal/acre
- Expressing lime application rates in units of lb/acre instead of short ton/acre
- Incomplete crop rotations

How do I assess and correct data quality issues?

If a user suspects that there is an error for a crop year input, for example if the total nitrogen applied to a field is twice as much as any other crop year in the project, the project administrator or specialist could go back to the source to verify the data entry and make any necessary corrections. Not all outliers or suspected errors will be incorrect data, but they represent an input that should be checked. If any corrections in the Platform are made, the Fieldprint metrics need to be recalculated for that crop year, and a new Comprehensive Data Output Report needs to be exported. It should not be assumed that checking project data via the QA Tool will result in clearing the project data as accurate, nor that the QA Tool will be able to identify all errors.

Are project administrators or specialists required to use the QA Tool?

There is no requirement to use the QA Tool. However, an important role of the project administrator or specialist is to ensure that growers are entering accurate data into the Fieldprint Platform. Visualizations and tables like those shown by the QA Tool could be created with Microsoft Excel, Python, R, SAS, Tableau, and many other analysis and visualization tools. This tool is also not designed to replace other quality assurance mechanisms that projects may use at various points within the project lifecycle.

Is the QA Tool synchronized with the Fieldprint Platform?

The QA Tool is a stand-alone website and not synchronized with the Fieldprint Platform; it has been developed to explore the data from a Comprehensive Data Output Report exported from the Fieldprint Platform. When an error is found, the data entry needs to be corrected within the Fieldprint Platform, and metrics need to be recalculated. Then, the project administrator or specialist would have to generate a new Comprehensive Data Output Report.

Can the output and summaries given by the QA Tool be used for claims?

The purpose of the QA Tool is to conduct quality analysis of Fieldprint project data to find errors before the data are reported elsewhere. Insights could be extracted using the QA Tool for internal uses, such as presentations, reports, and storytelling, but the data cannot be used for claims without prior approval from Field to Market. If a project wishes to use insights gained from the QA Tool, please contact claims@fieldtomarket.org.

Does the QA Tool website store any data?

The website does not store any data uploaded by users. The Comprehensive Data Output Report is temporarily uploaded to the server while a user is actively interacting with the QA Tool; once the QA Tool website is closed the data are erased.

Are the features in the QA Tool going to be added into the Fieldprint Platform?

Field to Market has included several quality assurance features within the Fieldprint Platform Version 3.0. There are no immediate plans to integrate these additional QA Tool features into the Fieldprint Platform, but there may be opportunities to do so in the future once we learn more about how projects are using the QA Tool.

If our project data is collected and reported on through a QDMP, can we still utilize the QA Tool?

In its current release, the QA Tool has been designed to work only with the Comprehensive Data Output Report generated from the Fieldprint Calculator. Qualified Data Management Partners (QDMP) may offer the option to export project data in a similar format so exported data can be used in the QA Tool. In addition, Field to Market will make available on request to QDMPs the underlying source code of the QA Tool should they elect to build and deploy a version within their own systems.

Will any QA Tool functionality be made available to QDMP via the Fieldprint Platform API?

There are no current plans to offer QA Tool functionalities via the Fieldprint Platform API. However, the backend code for the QA Tool will be made available on request. The code for the QA Tool is written in R, an open-source language, and the website is built using the Shiny from RStudio framework.

How can I give feedback, request features, or report errors in the QA Tool?

Please send feedback, error reports, or feature requests to Eric Coronel (ecoronel@fieldtomarket.org).