RESEARCH LETTER

Development of a soil test correlation and calibration database for the USA

Sarah E. Lyons1 | Dan K. Arthur2 | Nathan A. Slaton3 | Austin W. Pearce1 | John T. Spargo4 | Deanna L. Osmond1 | Peter J. A. Kleinman5

1 Dep. of Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC 27695, USA
2 USDA-ARS, University Park, PA 16802, USA
3 Dep. of Crop, Soil, and Environmental Sciences, Univ. of Arkansas System Division of Agriculture, Fayetteville, AR 72701, USA
4 Agricultural Analytical Services Lab., Pennsylvania State Univ., University Park, PA 16802, USA
5 USDA-ARS, Soil Management and Sugarbeet Research Unit, Fort Collins, CO 80526, USA

Correspondence
Sarah E. Lyons, Dep. of Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC 27695, USA.
Email: selyons@ncsu.edu

Assigned to Associate Editor Randy Kolka.

Funding information
USDA Agricultural Research Service,
Grant/Award Number: 58-8070-8-016;
USDA Natural Resources Conservation Service, Grant/Award Numbers: 69-3A75-17-45, NR203A7500010C00C

Abstract
As part of the Fertilizer Recommendation Support Tool (FRST) project, the FRST database was developed to consolidate and preserve U.S. soil test correlation and calibration data. Legacy phosphorus (P) and potassium (K) soil test data that met a minimum requirement were included in the database. The FRST database initially included over 1,200 individual trials from a range of years, cropping systems, geographic regions, and management practices. The FRST database is being migrated from a Microsoft Excel spreadsheet to a relational database format housed within the USDA-ARS Agricultural Collaborative Research Outcomes System (AgCROS) to be accessed via the online FRST decision support tool. Data will be continually added to the FRST database through an online submission form following peer review by the FRST team. The FRST database and associated decision support tool will aid researchers, extension associates, consultants, and farmers in improving fertilizer recommendations for crops across the United States.

INTRODUCTION

Fertilizer recommendations in the United States have been developed and maintained on a state-by-state basis since the early 1900s. Soil test recommendations based on political boundaries rather than regional characteristics or cropping systems have led to inconsistencies among states in soil
collection, research practices, and resulting fertilizer recommendations (Voss, 1998; Spargo et al. unpublished data, 2021; Zhang et al., 2021). A 2020 survey conducted as part of the Fertilizer Recommendation Support Tool (FRST) project found that only 12% of land grant universities had either recalibrated or validated their soil test phosphorus (P) calibration for corn (Zea mays L.) in the last 10 yr, and nearly 80% of calibrations were either over 20 yr old or had unknown origins (Spargo et al. unpublished data, 2021). Fertilizer recommendations that are up to date, accurate, and science based are needed for more efficient crop production and soil fertility management in the United States.

The FRST effort was initiated in 2018 to advance the accuracy of soil-test-based fertilizer recommendations through a foundational database and decision support tool from which recommendations can be scientifically developed and defended as best management practices (Lyons et al., 2020). The FRST project was modeled after a similar initiative, Better Fertiliser Decisions for Cropping Systems (BFDC) in Australia, a nationwide effort to deliver “specific interpretation guidelines suited to regional farming systems, crop types, and soils” (Watmuff et al., 2013, p. 425). The BFDC national database represents a variety of crops, nutrients, soil testing methods, and management practices and is accessed by the BFDC Interrogator, a web application for searching the online database and developing relationships between relative grain yield and soil test values. As a result of broad outreach and training efforts, 80% of all soil samples tested in Australia derive from decision support systems that use BFDC, and growers have become more confident in critical soil test ranges (Fixen et al., 2019). While the FRST database and decision support tool will be based on U.S. farming and research needs, the successful BFDC has provided useful insight for the development of FRST.

The FRST project involves multiple components, including the following:

1. Survey on current soil fertility practices and recommendations
2. Minimum dataset (MDS) for future correlation and calibration trials (Slaton et al., 2021)
3. Development of the FRST database to preserve and use correlation and calibration data
4. Relative yield calculation selection for the decision support tool
5. State-level soil test correlation and calibration trials funded by FRST
6. Multistate analysis of soil sampling depth influence on soil test outcomes
7. Modeling soil test correlation data
8. User-friendly decision support tool

The objective of this paper is to describe the development of the FRST database.

2 | BUILDING THE FRST DATABASE

2.1 | FRST is a collaborative effort

The most important component of the FRST project is the group of more than 80 scientists located across the United States and Canada (the FRST team) that have collaborated in the development and population of the FRST database. The majority of trials included in the database were provided by the FRST team.

2.2 | Building the FRST database

The FRST database was built with Microsoft Excel by FRST team project leaders and included a data dictionary tab, five data tabs, and four supporting information tabs (Table 1). The FRST database was developed for soil-test P and K correlation and calibration data, but with enough flexibility to allow for future changes and additions including metadata fields, nutrients, and production systems. Building the FRST database was an iterative process; changes and additions were made as data were entered, alongside MDS development, and during regular monthly FRST team meetings. These team meetings were used to provide project updates and discuss project components including database development.

The FRST database is hosted by the USDA-ARS Agricultural Collaborative Research Outcomes System (AgCROS) and cataloged in the USDA National Agricultural Library Ag Data Commons (https://data.nal.usda.gov; Lyons et al., 2020). The AgCROS data dictionary and National Agricultural Library thesaurus were used extensively when building the FRST database. When the FRST database included a field
that was not in the AgCROS data dictionary, additions to the AgCROS data dictionary were requested.

Following initial data collection and review, a legacy minimum dataset (LMDS) was defined for dataset inclusion. Whereas the MDS, intended to guide future research, includes more than 100 required and recommended factors (Slaton et al., 2021), the LMDS is less stringent to preserve as much historical data as possible. The LMDS includes trial year, location, replicated fertilizer P or K treatment rates and yield responses, and soil sample depth, analytical method, and pre-fertilizer test P or K concentrations. If a trial did not meet this standard, the information was archived and deemed incomplete and not added to the FRST database.

### 2.3 Obtaining and entering data

Data were collected from many different sources, including but not limited to journal articles (68%), extension and research bulletins (16%), conference proceedings (3%), dissertations and theses (11%), raw data in spreadsheets (1%), and summarized data in word-processing documents (1%). Both raw and summarized data in either electric or hard-copy formats were accepted if the LMDS information was complete. Data were obtained via general search engines and literature searches online or provided by the FRST team or other soil fertility faculty.

A legacy data collection guide was developed to aid collaborators with identifying data appropriate for the FRST database (http://www.soiltestfrst.org). All resources considered for the FRST database were scanned (if necessary) and archived. To encourage provision of unpublished data, the FRST team assisted with data submissions to Ag Data Commons for publication with accompanying citations including digital object identifiers. The FRST team also offered to embargo data following submission if researchers preferred to publish their data before making the data available in the FRST database.

Data entry began in July 2019. Before entry, a trial was examined to ensure it contained all information in the LMDS. If data were only presented in figures or important information was missing, including information beyond the LMDS, the FRST team contacted authors to complete the dataset requirements. When numerical datasets were not available for data presented in figures, an online extraction tool was used (WebPlotDigitizer; https://automeris.io/
WebPlotDigitizer/) and data were flagged in the database as extracted data. If it was impossible to obtain critical information about a trial, then data from the trial were archived and not included in the FRST database. Each trial entered was given a unique FRST trial ID number. A trial was defined as an individual site year for single-year trials, and for multi-year trials as an individual location where the same treatments were applied to the same plots annually. As much information as possible from the original data source was included in each entry, including original units. While FRST will allow the end user to request results in alternative units, only metric units will be formally accessed by FRST. If data were not originally in metric units, conversion equations were used and converted data was entered into separate columns. Additionally, each data tab in the FRST database included a “notes” column where unique or specific details not already captured by the existing fields could be included. Data were regularly checked for errors and inconsistencies. Finally, an instruction manual was developed to complement one-on-one training for data entry as personnel enter and exit the project.

2.4 Overview of data in the FRST database

By the end of the initial phase of the project (June 2021), the FRST database contained 1,295 trials from 1949 to 2018 representing 17 crops (alfalfa [Medicago sativa L.], camelina [Camelina sativa (L.) Crantz], chickpea [Cicer arietinum L.], cool season grasses, corn grain, corn silage, cotton [Gossypium hirsutum L.], flax [Linum usitatissimum L.], lentil [Lens culinaris Medik.], pea [Pisum sativum L.], peanut [Arachis hypogaea L.], rice [Oryza sativa L.], sorghum [Sorghum bicolor (L.) Moench], soybean [Glycine max (L.) Merr.], sugarcane [Saccharum spp.], sweet potato [Ipomoea batatas (L.)], and wheat [Triticum aestivum L.]) and 34 states (Alabama, Arkansas, Colorado, Connecticut, Delaware, Florida, Georgia, Iowa, Kansas, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, North Carolina, North Dakota, Nebraska, New Hampshire, New Jersey, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, Washington, West Virginia, and Wisconsin). Trials used a range of soil test P and K methods, including Mehlich-1 and -3, Bray-1 and -2, Olsen, Morgan, Modified Morgan, Lancaster (Mississippi Soil Test), Haney, Truong, ammonium acetate, acetic acid, nitric acid, resin, water, iron hydroxide, oxalate, sodium acetate, K saturation, rate of K release, and tetraphenylboron. Data were from 48 published journal articles, 11 extension and research farm bulletins, seven dissertations, two conference proceedings, and one unpublished dataset provided by a member of the FRST team. The types of data in the FRST database were largely dependent on either work published and available online or the cooperation from researchers and collaborators; we suspect that the majority of historical correlation and calibration data is unpublished, either located in filing cabinets or discarded as soil fertility researchers have retired. The database will continue to grow over time; over 100 additional data sources that contain correlation and calibration information have been collected to be vetted and added to the database, and there is likely more data located online and with collaborators that we have not yet accessed.

Dataset completeness was an obstacle when building the FRST database, as there were often inconsistencies in what data and metadata, and the format they were reported in, were included in publications or reports. These inconsistencies made it particularly challenging to build a database for both historic and future data and resulted in many fields being left blank for the older trials. In some cases it was possible to obtain important information from authors to supplement a dataset; however, this was often impossible for older studies. While the LMDS was instrumental in selecting appropriate data, it allowed for some trials with very little metadata to be included in the database. The FRST project is attempting to preserve as much historical correlation and calibration data as possible, and so while a lack of metadata will make it difficult to confidently interpret and use some data, the decision tool will allow users to search by year and other metadata to filter out any historical data that may not be relevant or complete enough for developing modern recommendations.

Regardless of a study’s publication date, certain pieces of information were more common than others. For example, 93.5% of trials reported soil pH, 51.3% reported county information, 28.3% reported crop planting date, and only 9.7% reported a measure of variability for yield. We hope that with the guidance of the MDS (Slaton et al., 2021), soil test correlation and calibration publications will report more complete datasets going forward.

3 LESSONS LEARNED AND NEXT STEPS

One of the most challenging aspects of building the FRST database was the regular need for changes to accommodate both minimum dataset development and different types of data being entered. Although this was largely unavoidable due to the iterative nature of the project and the varying data formats in the literature, it was important to have the database layout and data dictionary as close to the end product as possible. Data organization, consistency, and recordkeeping of activities were very important as the project grew and evolved.

Dataset completeness presented multiple challenges. Scanning through documents ensuring that a dataset met the LMDS, as well as contacting researchers requesting additional information, was often a slow and unsuccessful process. We estimate that approximately 30% of the obtained data...
was incomplete. In Australia, it was estimated that only 30% of correlation and calibration trial data was included in the BFDC, possibly due to underreporting of nonresponsive trials (Fixen et al., 2019). Extracting data from figures when necessary was also tedious and potentially imprecise.

The input and contributions provided by the FRST team were instrumental in building the FRST database. Having representatives with regional connections and expertise was especially helpful for collecting relevant data from a broad geographic area. Transparency and inclusivity through regular updates and meetings were worthwhile for building trust and supporting communication throughout the project.

The FRST team worked with Partnerships in Data Innovations, a collaborative data management infrastructure development and modernization initiative between USDA-ARS, other federal agencies, and nongovernmental organizations, to ensure that data ontology and database structure were properly aligned for migration into the relational database housed in AgCROS. New data will be entered into the relational database using an online submission form and data upload. All submissions will go through peer review by members of the FRST team before being incorporated into the FRST database. Any new raw data submitted will have the opportunity to receive a publication digital object identifier (DOI) through Ag Data Commons, further encouraging data sharing and publication. The FRST database will only be as useful as the data it includes; researchers and funding agencies will be encouraged to recognize it as a preferred data destination in data management plans for relevant soil fertility research.

Programming of the decision support tool (FRST) will be an iterative process involving FRST team input and beta testing by researchers, extension associates, and producers. The usability and accessibility of FRST and the FRST database are critical for achieving the FRST project goals: to increase soil testing transparency by promoting clear and consistent interpretations of fertilizer recommendations and provide the best possible science to enhance end-user adoption of nutrient management recommendations.

ACKNOWLEDGMENTS
In addition to the more than 80 members of the FRST team, we would also like to thank the USDA-NRCS (Grants 69-3A75-17-45 and NR203A7500010C00C) and the USDA-ARS National Programs for Natural Resources (Grant 58-8070-8-016) for funding the project.

AUTHOR CONTRIBUTIONS
Sarah E. Lyons: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft. Dan K. Arthur: Conceptualization; Investigation; Methodology; Project administration; Software; Supervision; Visualization; Writing – review & editing. Nathan A. Slaton: Conceptualization; Data curation; Methodology; Project administration; Resources; Supervision; Visualization; Writing – review & editing. Austin W. Pearce: Data curation; Validation; Visualization; Writing – review & editing. John T. Spargo: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Supervision; Visualization; Writing – review & editing. Deanna L. Osmond: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Visualization; Writing – review & editing. Peter J. A. Kleinman: Conceptualization; Project administration; Resources; Visualization; Writing – review & editing.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ORCID
Sarah E. Lyons https://orcid.org/0000-0001-5091-8926
Dan K. Arthur https://orcid.org/0000-0001-6471-0149
Nathan A. Slaton https://orcid.org/0000-0002-0015-2034
Austin W. Pearce https://orcid.org/0000-0002-2541-896X
John T. Spargo https://orcid.org/0000-0002-7400-4739
Deanna L. Osmond https://orcid.org/0000-0002-6336-8318
Peter J. A. Kleinman https://orcid.org/0000-0002-5093-3736

REFERENCES
Fixen, P. E., Norton, R., Peverill, K., & Watmuff, G. (2019). BFDC: An Australian soil test calibration database success [Paper presentation]. 2019 ASA-CSSA-SSSA International Annual Meeting, San Antonio, TX. https://scisoc.confex.com/scisoc/2019am/meetingapp.cgi/Paper/118110
Lyons, S. E., Osmond, D. L., Slaton, N. A., Spargo, J. T., Kleinman, P. J. A., Arthur, D. K., & McGrath, J. M. (2020). FRST: A national soil testing database to improve fertility recommendations. Agricultural & Environmental Letters, 5(1), e20008. https://doi.org/10.1002/ael2.20008
Slaton, N. A., Lyons, S. E., Osmond, D. L., Brouder, S. M., Culman, S. W., Gerson, D., Gathoni, L. C., Hoben, J., Kleinman, P. J. A., McGrath, J. M., Miller, R. O., Pearce, A., Shoher, A. L., Spargo, J. T., & Volencic, J. J. (2021). Minimum dataset and metadata guidelines for soil-test correlation and calibration research. Soil Science Society of America Journal. https://doi.org/10.1002/saj2.20338
Voss, R. (1998). Fertility recommendations: Past and present. Communications in Soil Science and Plant Analysis, 29(11–14), 1429–1440. https://doi.org/10.1080/00103629809370040
Watmuff, G., Reuter, D. J., & Speirs, S. D. (2013). Methodologies for assembling and interrogating N, P, K, and S soil test calibrations for Australian cereal, oilseed and pulse crops. Crop & Pasture Science, 64, 424–434. https://doi.org/10.1071/CP12424
Zhang, H. J. A., Grove, J. H., Osmond, D. L., Alford, S., Florence, R. J., Huluka, G., Hardy, D. H., Lessl, J. T., Maguire, R. O., Mulavarapu,
R. S., Oldham, L., Pena-Yewtukhiw, E. M., Provin, T. L., Slaton, N. A., Sonon, L. S., Sotomayor, D., & Wang, J. J. (2021). Soil test based P and K rate recommendations across the Southeast: Similarities and differences, opportunities and challenges. *Soil Science Society of America Journal*, https://doi.org/10.1002/saj2.20280

**How to cite this article:** Lyons, S. E., Arthur, D. K., Slaton, N. A., Pearce, A. W., Spargo, J. T., Osmond, D. L., & Kleinman, P. J. A. (2021). Development of a soil test correlation and calibration database for the USA. *Agric Environ Lett*, 6, e20008. https://doi.org/10.1002/ael2.20058