



Big Data Driven Agriculture: Big Data Analytics in Plant Breeding, Genomics, and the Use of Remote Sensing Technologies to Advance Crop Productivity

Nadia Shakoor,* Daniel Northrup, Seth Murray, and Todd C. Mockler

Core Ideas

- Interdisciplinary efforts in high-throughput field phenotyping
- Linking proximal and remote field phenotyping
- Cyberinfrastructure for high-throughput field phenotyping

Plant breeding and agronomy are labor-intensive sciences, and the success of these disciplines is critical to meet planetary challenges of food and water security for the world's growing population. Recent gains in sensor technology, remote sensing, robotics and autonomy, big data analytics, and genomics are being adopted by agricultural scientists for high-throughput phenotyping, precision agriculture, and crop-scouting platforms. These technological gains are ushering in an era of digital agriculture that should greatly enhance the capacity of plant breeders and agronomists. This report encompasses the priorities and recommendations that emerged from two USDA National Institute of Food and Agriculture (NIFA)-funded Big Data Driven Agriculture workshops held on 26–27 Feb. 2018 in Arlington, VA. The objectives of the workshops were to bring together diverse subject-matter experts in the represented disciplines of plant breeding, machine learning, remote sensing, and big data infrastructure and analytics to (i) explore how large and comprehensive datasets in plant breeding, genomics, remote sensing, and analytics will benefit agriculture; (ii) discuss strategies for creating a successful field phenotyping campaign and to determine protocols for the collection and analysis of agricultural big data; (iii) consider how to best engage the broader community of public and private plant breeders and agronomists to determine additional challenges, make wider use of the data, and ensure application of standardized methods to other datasets; and (iv) generate a report describing cross-cutting short- and long-term funding needs for continued success in this domain.

The USDA Food and Agriculture Cyberinformatics and Tools (FACT) initiative provides an excellent opportunity for supporting high-throughput, automated data collection platforms, remote sensing, and analytics to address public and private agricultural priorities. To ensure that agricultural datasets are used to their maximum potential, it is important to foster analytics expertise to derive biological insights for genetic associations, growth predictions, crop improvement, and precision agriculture. The FACT initiative also recognizes the need to develop communities and infrastructure to realize the potential of large agricultural datasets. The USDA is collaborating with NASA, the National Geospatial-Intelligence Agency, and other agencies to use satellite monitoring in agricultural applications such as water management, photosynthesis monitoring, and yield forecasting. New analytics methods are being developed that effectively upscale low-resolution, high-throughput data streams, such as satellite data, by using the higher resolution sensors being deployed in the field (e.g., USDOE TERRA platforms).

One compelling direction is to take advantage of new methods in machine learning to create tools that lower the bar for subject matter experts to use machine learning tools. A new analytics group funded by the Defense Advanced Research Projects Agency (DARPA), the Data Driven Discovery of Models, is working with agricultural experts on harvest prediction, weather, trade, and immigration to understand regional food

N. Shakoor and T.C. Mockler, Donald Danforth Plant Science Center, Saint Louis, MO 63132; D. Northrup, Booz Allen Hamilton, Inc., McLean, VA 22102; S. Murray, Dep. of Soil and Crop Sciences, Texas A&M Univ., College Station, TX 77843.

© 2019 The Author(s). This is an open access article distributed under the terms of the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)
 Plant Phenome J. 2: 180009 (2019)
 doi:10.2135/tppj2018.12.0009

Received 10 Dec. 2018.

Accepted 1 Apr. 2019.

Supplemental material online.

*Corresponding author (nshakoor@danforthcenter.org).

Abbreviations: FACT, Food and Agriculture Cyberinformatics and Tools; NIFA, National Institute of Food and Agriculture; NSF, National Science Foundation; RFA, Request for Applications; UAV, unmanned aerial vehicle.

security and aims. This modeling effort represents one opportunity to leverage the nation's cyberinfrastructure and government investments in planetary science to advance agriculture.

To provide insight into these various initiatives and agencies, two Big Data Driven Agriculture workshops, focused on big data analytics in plant breeding and genomics, satellite data and modeling, and the use of machine learning and other remote sensing technologies to advance crop productivity, were organized by the Donald Danforth Plant Science Center. The first workshop began with opening remarks by the Director of USDA–NIFA, Dr. Sonny Ramaswamy, and the second workshop was opened by the Administrator of the USDA–ARS, Dr. Chavonda Jacobs-Young. The opening remarks were followed by a series of presentations designed to provide the workshop participants with a status report on state-of-the-art research and applied work in related disciplines. Each group of presentations was followed by a facilitated panel discussion allowing the opportunity for questions, answers, and productive discussion. In the afternoon of each workshop day, participants interacted in smaller group discussions and “hackathons.” Key outputs of the presentations and discussion sessions are presented here.

To encourage graduate student and postdoctoral interactions with geneticists, breeders, and remote sensing experts and to promote interdisciplinary career perspectives in agriculture, abstracts were invited from graduate students and postdoctoral scholars. Ten travel awards to the workshop were given to selected student and postdoctoral applicants. Four students and postdoctoral researchers were selected from the abstract submissions to give 15-min presentations on their research.

This meeting brought together participants from USDA, the National Science Foundation (NSF), the Advanced Research Projects Agency–Energy (ARPA-E), DARPA, and NASA with a community of scientists and engineers to develop a road map for the delivery of immediately applicable algorithms and best practices and a strategic plan for future success in this domain through managed standards, data repositories, and interdisciplinary engagement. The meeting participants were solicited by a diverse organizing committee (see Supplement 1), and are hereafter referred to as the “Big Data Driven Agriculture community” (Fig. 1).

Background and Significance Interdisciplinary Efforts in High-Throughput Field Phenotyping

High-throughput field phenotyping is a relatively new but rapidly growing research area, and it will remain a top agricultural research priority in the next decade. The NIFA FACT Initiative provides a timely opportunity to develop a cross-disciplinary research agenda, bringing together plant breeding and analytics with phenotyping data and modeling. Remote sensing technologies, proximal sensors, deployment platforms such as unmanned aerial vehicles (UAVs) and ground vehicles, and statistical analytics are being rapidly customized and deployed for high-throughput

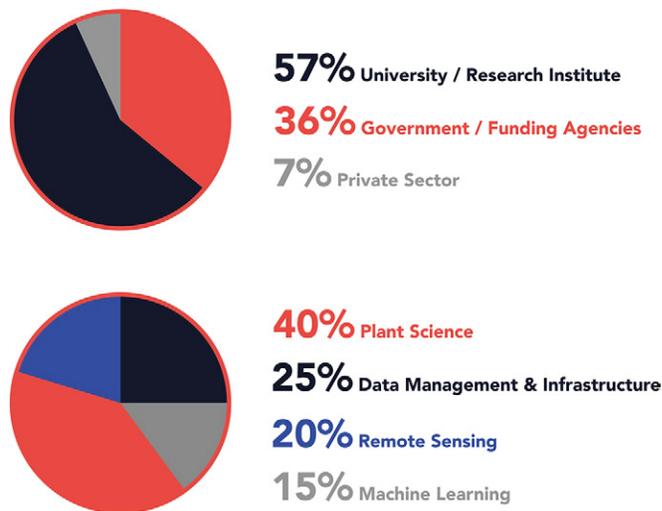


Fig. 1. Breakdown of attendees by background.

phenotyping and use as plant performance measurement tools for crop improvement and breeding and precision agriculture platforms for agronomy, soil science, and farm management. Currently, the most important challenge is to ensure that the plant science communities and data analytics communities know how to use these data to deliver actionable results for scientists and farmers. Coordination and interaction among the key disciplines of breeding, agronomy, computer science, data science, engineering, and genomics is needed so that these high-throughput phenotyping tools are accessible for broad and applied agricultural use.

Linking Proximal and Remote Field Phenotyping

The coordinated collection of high-resolution proximal in-field datasets and satellite monitoring has rarely been done, and it may be possible to create even greater insights by fusing these methodologies. The collection of high-resolution proximal data is critical for many applications, but in many cases insight and recommendations will be acceptable at far lower resolution and can be generated from high data volume satellite resources. For example, weather patterns, environmental determinations, and assessments of target crop populations may be achieved with very-high-throughput satellite platforms.

In many remote-sensing applications, high temporal resolution is the most important feature, and the goal is to have more “revisits,” ultimately reaching the potential for daily imaging. However, a breeder may be willing to sacrifice temporal resolution for higher image resolution from proximal sensors to ensure data collection at the plot, plant, or leaf level. This is a nascent and rapidly developing field, and research is needed to understand the inherent tradeoffs between temporal and spatial resolutions. The optimal solution depends on the crop and research objectives.

Multiple US agencies and other government services are using satellite data for estimating crop area, yield, and condition, including USDA–National Agricultural Statistical Service (NASS), USDA–Foreign Agriculture Service (FAS), Group on Earth

Observations—Global Agricultural Monitoring Implementation Team (GEOGLAM), Famine Early Warning Systems Network (FEWS NET), and for weather modeling (NOAA and others). In the private sector, many firms are deploying low-cost and high-return-rate satellite fleets to develop decision support systems for crop producers. These firms are leveraging models developed by plant physiologists and are investing heavily to deliver actionable data for farmers and plant breeders.

Cyberinfrastructure for High-Throughput Field Phenotyping

The Big Data Driven Agriculture community has need for additional tools from the national cyberinfrastructure toolbox, including knowledge frameworks that organize the work of a diverse community in a coherent manner. A highly promising method is to develop models that provide recommendations for plant breeding, management, or policy decisions. Gramene's Plant Reactome (plantreactome.gramene.org) is an example repository where researchers contribute findings to maps of protein networks and pathways. This site presents large amounts of work in a digestible way and guides new research predicting emergent cellular behaviors. Likewise, an effort funded by the Foundation for Food and Agriculture Research (FFAR), Crops In Silico, is creating models that link genomic insights, molecular networks, and plant phenotypes, providing a method for researchers working at any scale to contribute to shared results. At a macro scale, a modeling community organized by DARPA, through the program World Modelers, is charged with using similar methods to combine crop yield predictions, weather, trade, and immigration to predict regional food security challenges. These large-scale models represent another opportunity to leverage the nation's cyberinfrastructure and government investments to advance agriculture. Once the methods are developed for food security prediction, they will be broadly applicable for other complex modeling exercises that require multiple data stream applications like farm management and crop improvement. The Big Data Driven Agriculture meeting, as summarized here, aimed to link diverse fields and to create a research community that can develop and deploy a national cyberinfrastructure network that supports plant breeding, food security, and other USDA missions.

Key Takeaways

Specific Recommendations for NIFA FACT and Related Initiatives and Agencies

1. The Big Data Driven Agriculture community gives a strong recommendation for longer term funding or formal grant extension and “plus up” opportunities to support breeding and genomic selection projects. The breeding cycle for most annual crops can take 7 to 10 yr from an initial cross to commercialization, with perennial and tree crops taking longer. Incorporating high-throughput field phenotyping data will probably increase the number of genotypes that can be screened and improve selection accuracy once technologies and tools for breeders are available and accessible.
2. The Big Data Driven Agriculture community requests funding opportunities or a specific initiative that supports a sustainable data repository system with tools for analysis (Fig. 2). This is a vital undertaking for continued success in this interdisciplinary effort, and long-term federal agency support is necessary for it to succeed. While some infrastructure and tool development can be conducted through competitive grants, ultimately a permanent repository (e.g., the National Institutes of Health's National Center for Biotechnology Information [NCBI], USDA–NASS) is needed for long-term stability and to ensure sufficient maintenance. An alternative model would be a centralized clearinghouse that provides a seamless interface to permanent institutional repositories such as libraries. This approach could serve a similar purpose while being more robust and inclusive but will require more advanced technologies and coordination.
3. Many members of the community recommend that NIFA should create funding opportunities to invest in existing infrastructure rather than initiatives to develop new infrastructure for high-throughput field phenotyping technologies. Interdisciplinary grants related to data-driven agriculture generally have two main components: (i) development of the infrastructure or core technology followed by (ii) hypothesis-driven experimentation using the newly developed infrastructure or technology. There is a sense that the current timeline of a typical USDA–NIFA grant (3 yr) realistically results in achieving only the first component of the project, which is successful infrastructure development. Within a typical funding period, there is not sufficient time to implement key improvements to the infrastructure and/or technology or to demonstrate application of a developed system. Either grant terms need to be extended to allow time for implementation or funding should be made available specifically for productive and impactful applications of scientific research.
4. To facilitate sustained interdisciplinary interaction over the next decade, the Big Data Driven Agriculture community recommends that agencies such as NIFA fund interdisciplinary training programs for principal investigators as well as students and postdoctoral researchers. For example, can NIFA package fellowships in which graduate students from multiple disciplines work in multiple labs? Similar to the NSF-funded Predictive Plant Phenomics program at Iowa State University or the NIFA National Needs scholarships, a program that cross-trains students should be prioritized. The community strongly believes that the next generation of grant applicants, Ph.D. students, and postdoctoral researchers will benefit from training and exposure to diverse fields including data science, biology, engineering, and math (Fig. 3).
5. The Big Data Driven Agriculture community requests funding opportunities that are problem oriented without being overly narrow. Many constituents are interested in developing

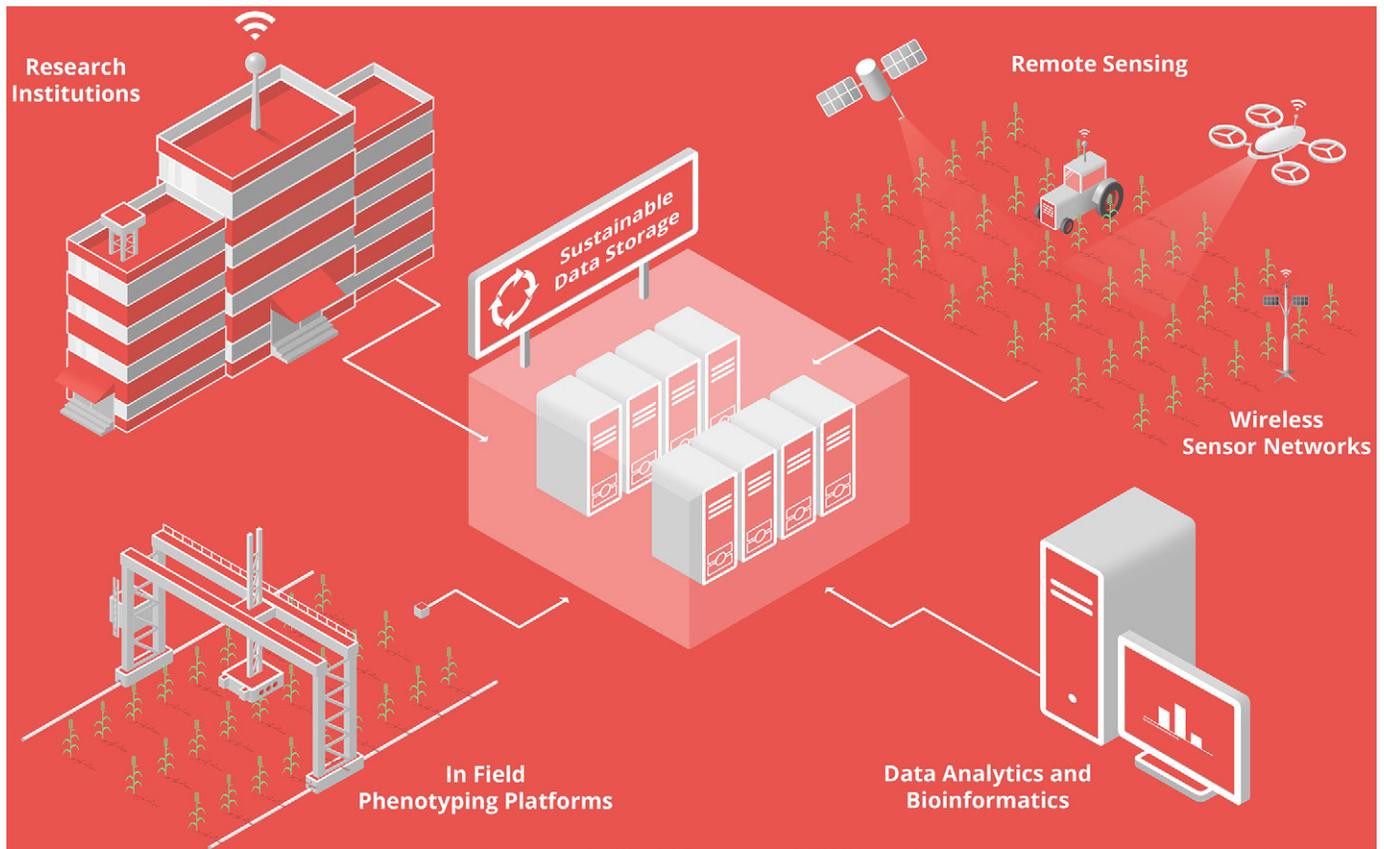


Fig. 2. Sustainable data repository system.

and using phenology-focused tools that are specifically designed to solve problems for agricultural stakeholders. As an example from the private sector, Planet Labs is working with Farmers Edge to analyze data to determine crop-cycle changes. These funding opportunities could initiate research with a small amount of money, with the opportunity for larger amounts with progress as suggested in Recommendation 1 above. Further, research problems should be generated from an agricultural perspective to focus on a current and impactful problem and then bring in other disciplines to figure out how to solve the problem. With each proposal call, we recommend an explicit statement from both the agency and the applicant addressing the applied objectives of big data collection and analysis.

6. The collective Big Data Driven Agriculture community requests that funding agencies promote and support initiative-wide standard operating practices (SOPs), data standards, and data formats. With research grants heavy in experimentation and development, scientists generally do not take the time to learn and teach SOPs. Standardization of phenomics-related variables is long overdue; however, without incentive or repercussion, funded proposals can often be biased toward individual research objectives over transdisciplinary research, and meaningful translation across phenomics platforms and transdisciplinary research efforts is difficult, if not impossible. Standard procedures

and formats would allow transformative research and accelerated discovery through integration of data across species, time, and location.

Concerns and Additional Recommendations from the Big Data Driven Agriculture Community Facing this Interdisciplinary Effort Real World Metrics for Success

The key goals for plant breeders are to predict phenotype (preferably in untested genotypes grown in untested environments) and increase genetic gain; however, deep data collection

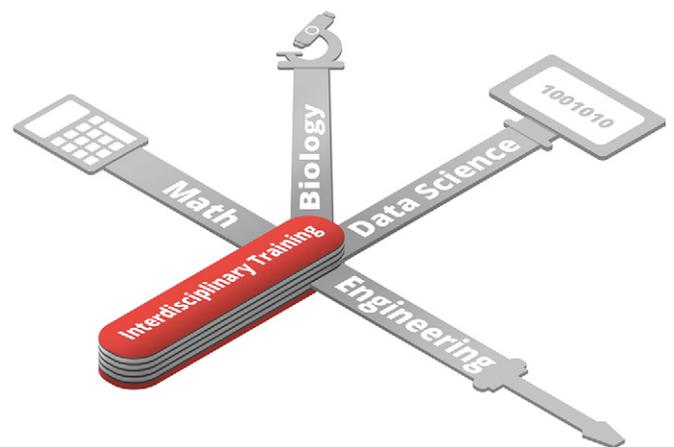


Fig. 3. Diversifying tool sets through interdisciplinary training.

and analysis are needed to support this objective. The concern with precision agriculture is the lack of rapid data analysis that provides actionable guidance that farmers can trust. Farmers require near-real-time data to effectively adjust management strategies to optimize yield. However, it is unclear the extent to which sensing has improved crop production. Further, of the many technologies currently available, it is unclear to the community which tools are being effectively used for crop improvement. How can funding agencies work to measure the impact of diverse research groups? Real world metrics and milestones for cross-disciplinary projects are underdeveloped and should be formally considered by NIFA and other funding agencies as Requests for Applications (RFAs) are drafted.

Trade-offs of High-Throughput Sensing

Another concern about the use of current high-throughput sensing and machine-learning-based prediction efforts is that it is unlikely that the rare “unicorn” genotype that has the potential to make large step-function crop improvement advances (an outlier almost by definition) will be detected. The current model of driving steady but incremental genetic gain favors exclusion of possible outliers. To partially address this concern, integration of multiple data layers, both remote and proximal, is needed to dramatically improve the phenotype prediction equation.

Emphasis on Data Quality and Standards

E.O. Wilson’s comment, “We are drowning in information, while starving for wisdom,” was quoted by Dr. Sonny Ramaswamy in his opening remarks, and this statement captures one of the biggest challenges facing the Big Data Driven Agriculture community. In the race to publish and be awarded grants, there is a concern about the lack of emphasis placed on data quality by principal investigators. There is a consensus that “garbage in” in terms of primary data quality results in “garbage out” of final data quality (Fig. 4) and that collecting and quality checking data can take a long time—too long for the current 3-yr lifespan of a typical grant. Insights from current machine learning models are only as good as the data used. Further, there are

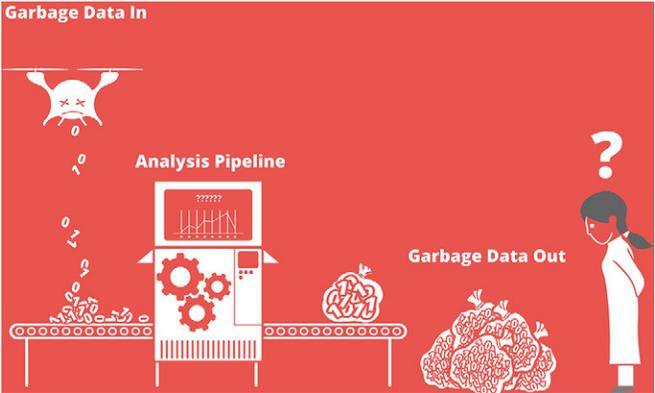


Fig. 4. Data quality and standards: “garbage in” data, “garbage out” results.

no real standards or protocols for sensor precision and calibration of instruments. There are individual efforts to address this from universities and large research programs; however, there is a need for one or more organizing bodies or programs to oversee standard or protocol implementation, evaluate the quality of research, and develop standards for future research programs.

Diversity in Grant Applicant Groups

Due to a number of varied reasons, investigators generally apply for grants within their networks of institutional and first degree colleagues. These circles of established applicant groups can be hard to join, particularly for new faculty. To address this concern and benefit the interdisciplinary FACT initiative, we suggest that NIFA and other agencies consider “playing matchmaker” and pairing grant proposals toward a common goal. We recognize this is not a traditional role for NIFA, but it could be a very impactful one. A pilot project might be a beneficial first step.

Q&A with the Big Data Driven Agriculture Community

Question 1: How can large and comprehensive datasets on plant breeding, genomics, remote sensing, and analytics benefit agriculture (Fig. 5)?

1. These large datasets can significantly contribute to cultivar development. Data fusion from multiple sensors can be used to make cultivar selections, as breeding programs often deploy multiple sensors to measure unique physiological or architectural attributes to make informed breeding decisions.
2. The use of sensor datasets that have relationships with target traits (e.g., yield, drought tolerance) can be effectively used during the breeding season to assist selection decisions.

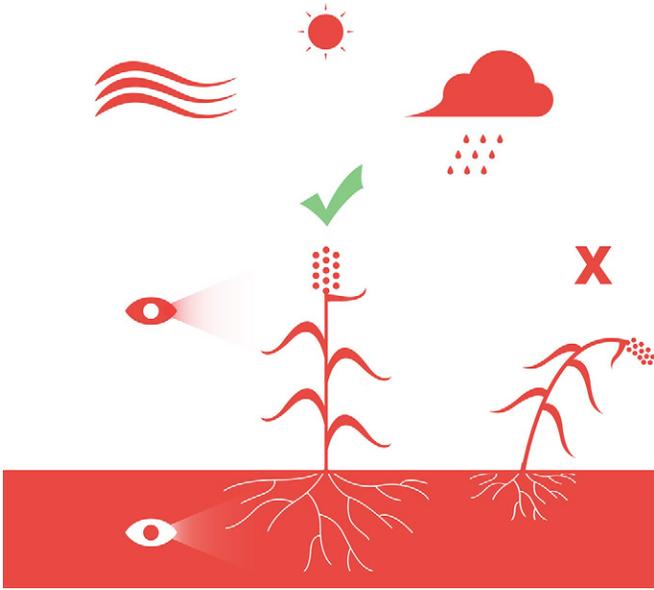


Fig. 5. Question 1: Using field sensing and large data sets to inform cultivar development.

3. These large datasets can inform genomic selection and machine learning models for breeding and crop modeling.
4. Results, knowledge, and ideas from big data initiatives in agriculture need to formally integrate university extension services. Extension bridges basic and applied research, and extension scientists are uniquely positioned and skilled to translate knowledge and technology applications and deliver it to the farmer or producer.

Question 2: What methods could be used to create a successful field phenotyping campaign?

1. Well-tested and documented sensor calibration is important to collect reproducible and biologically relevant data. Protocols for sensor calibration should be published with the research outcomes.
2. Appropriate adjustment of sensor data resolution to the field campaign and experimental design is necessary. Phenotyping speed is generally inversely correlated to sensor spatial resolution, and the right balance should be struck to achieve the field campaign and project goals.
3. Phenotyping campaigns need to have clearly defined strategies to prevent unnecessary and time-consuming data collection. Data collection for field campaigns should measure what is important to the project goals, not what is simply easy to measure.
4. Precision and accuracy are often unknown in a field phenotyping effort. Measures of the environment need to be standardized to account for variation in the sensor phenotypes that are observed.

Question 3: How can we determine protocols for the collection and analysis of agricultural big data?

1. Newly established data collection and analysis protocols to be used in phenotyping should garner the input and support of professional societies.
2. The Big Data Driven Agriculture community is international. Where appropriate, the US-based research efforts should implement and apply standards commonly used and established in international programs.
3. Agencies like NIFA and the FACT initiative can support development and implementation of protocol standards.

Question 4: How can we most effectively address the need for a sustainable means of data storage and access?

1. The solution to this question needs to include discussion and buy-in from public and private industry, universities, and the government.
2. Financial support for a long-term data repository that maintains original copies is required, but uncertain, and should be addressed immediately.
3. The Big Data Driven Agriculture community proposes the development of a federated data storage system as a collaboration between private, public, and government agencies. Business

models for this concept will be needed at multiple levels to support collection and maintenance costs. The demand for data storage is growing at a rate faster than storage costs are decreasing, and long-term sustainability of the shared system is critical.

4. The recommended centralized platform is likely to attract other researchers who will bring even more data, thus increasing the storage demand.
5. Data collection often evolves over the course of a project and usually over-delivers types and amounts of data.
6. To ensure use of a federated data storage system, funding agencies might consider withholding funding until data are deposited into a central repository.

Question 5: What research engagement opportunities might cut across the represented disciplines of plant breeding, machine learning, remote sensing, and big data infrastructure and analytics?

1. Research challenges could initially be generated from the perspective of agricultural stakeholders (e.g., farmers, non-governmental organizations [NGOs], extension services), and subsequently bring in researchers in additional disciplines to address specific research challenges. Core disciplines of crop physiology, pathology, entomology, soil science, and in silico biology should not be overlooked.
2. Funding agency awards should support multiple, interdisciplinary principal investigators. More interdisciplinary teams of engineers, data specialists, and plant breeders are needed. The ARPA-E TERRA and ROOTS programs are potential models. It is difficult to coordinate these efforts without shared program planning, and greater results can be achieved through planned coordination.
3. The Big Data Driven Agricultural community is a highly interdisciplinary community, and few institutions have a full team to put all the pieces together. Funding agencies should take on “matchmaking” for specific initiatives, bringing together research groups and institutions that might not normally interact. This can include matching smaller, less well-funded research groups with larger institutions that may have greater resources.

Question 6: What cross-cutting short- and long-term funding needs can you identify for continued success in these domains?

1. Resources are needed for developing standards and best practices prior to the completion of the grant. There can be great value in the generation of template data sets for training and other learning opportunities.
2. When funding is granted, agencies should consider offering additional resources earmarked for curation, publishing, and promoting data. For example, in some cases NSF provides additional funding for computing resources for groups with NSF-funded grants, and the National Institutes of Health give credits for certain computational services and applications.

3. Principal investigator training across disciplines is needed to communicate current capabilities and state-of-the-art methodologies. This could be short courses, online modules, and webinars. Moreover, principal investigators should be exposed to stakeholders in agriculture (e.g., farmers, NGOs, extension specialists) to understand real world needs and challenges.
4. Interdisciplinary training opportunities for students and postdoctoral researchers are needed. Similar to the NSF Predictive Plant Phenomics program at Iowa State University or NIFA's National Needs scholarship, funding agencies should consider a package of fellowships in which graduate students from different disciplines work in multiple labs. Students and postdoctoral researchers should be cross-trained in the areas of data science, bioinformatics, engineering, and statistics.

Question 7: How can we incentivize cross-disciplinary and transdisciplinary work when discipline-specific discoveries are rewarded?

1. Agencies should consider direct funding support for student exchanges and support for multiple faculty across disciplines.
2. Agency RFAs should explicitly require a cross-discipline approach instead of making an implicit recommendation in the proposal guidelines.
3. Funding milestones and follow-up funding could reward cross-discipline discoveries.
4. Requests for Applications could incentivize research approaches that come from other disciplines for application to agricultural problems.
5. Mechanisms to highlight data, research, and code by individuals within a larger interdisciplinary program should be created so that individual contributions to overall projects are clear.

Question 8: What measurements are feasible with remote sensing, and when is in-field monitoring needed? How might you design experiments that incorporate ground sensing and remote sensing to leverage the capacity of both?

1. Traits like crop fraction cover, hyperspectral reflectance, leaf area index, and disease resistance are traits of interest. Radiometrically corrected data and surface reflectance and bidirectional reflectance distribution function are all feasible with remote sensing; however, noise in growth curves can be attributed to the plant or crop and also the atmosphere.
2. Enviro-typing campaigns would benefit from the assistance of remote sensing where several different types of data are needed.
3. Virtual constellations comprising different modes and scales of data collection is a challenging area (e.g., UAV to satellite data).
4. Calibration protocols for aerial platforms are different than for the satellite platforms. With surface reflectance, the atmosphere is modeled with the sun angle and reflectance. MODIS, Landsat, etc., all have been calibrated using surface reflectance,

and satellites have extra bands delegated to this correction. With the advantage of higher resolution, UAV systems do not have these standardized protocols for correction.

Question 9: Some groups are considering having shared UAV user facilities. How feasible would it be for a university to set up a core facility on analysis of geospatial data?

1. A shared research facility would be highly useful to bring in sufficient resources for all groups, and universities may have the infrastructure and resources to maintain such a facility.
2. Centralized locations have the potential to facilitate training in standardized operation and deployment of phenotyping technologies.
3. Data sharing policies have to be in place, as concerns related to proprietary data and licensing are likely.
4. Centralized facilities enable standardization of data products and methodologies as well as adoption and development of best practices.

Conclusions

The main outcomes of the Big Data Driven Agriculture workshops were (i) the current white paper with suggestions to NIFA and other interested funding agencies for future RFAs and (ii) connecting researchers from the various disciplines with each other and with the Departments of Agriculture, Defense, Energy, and other governmental departments for the discussion of adopting technologies and creating opportunities for agricultural research. New funding methods are needed to support innovation, and the Big Data Driven Agriculture community has six core recommendations for building a vibrant phenotyping community in the United States:

1. Provide phased, stage-gate funding up to a complete crop cycle (7–10 yr): Success of multifaceted systems projects require longer setup than traditional research programs. Funders should explore phased funding structures with stage gates and incremental increases in funding to allow successful teams the continuity to achieve large impacts.
2. Build a centralized data repository: Researchers need a centralized data repository to store, compare, and repurpose data. This resource could support a new team of data analysts who are available to researchers to assist in data preservation and reuse.
3. Invest in existing infrastructure and tools: The community needs opportunities to continue use of de-risked, existing phenotyping methods and equipment to achieve breeding or agronomic outcomes.
4. Provide interdisciplinary training opportunities for students: Expanded funding efforts are needed to train students and postdoctoral scientists in multiple disciplines, providing infrastructure and tools to support the next generation of

agricultural researchers who are equally comfortable on a keyboard and a combine.

5. Provide problem-focused funding: Phenotyping efforts can be accelerated by focusing teams on specific agricultural problems that allow comparison of algorithms and can serve to coordinate efforts at a program level.
6. Develop data standards and standard operating practices: Collaboration will be greatly enhanced by the development of standard, intercomparable data and software. This should include protocols for standardized data collection and calibration, gold standard datasets for algorithm validation, and common data exchange formats for interoperability. Coordination with organizations such as the National Institute of Standards and Technology (NIST) would be beneficial.

Acknowledgments

This workshop was sponsored by the USDA–NIFA FACT program via Grants no. 2018-67021-27483 and 2018-67013-27427. Any opinions, findings, conclusions, or recommendations expressed here are those of the workshop participants and do not necessarily represent the official views, opinions, or policy of the funding agency. The recommendations put forth in this report also do not necessarily reflect the opinions of all attendees of the workshop. We have summarized general consensus topics and suggestions that were documented by several note-takers and the authors during the meeting breakout sessions, panel discussions, and presentations. The names and affiliations of participants mentioned here were current at the time of the workshop and may have changed. The organizing committee would like to thank the speakers, moderators, and student note-takers. We also thank Kathleen Mackey and Bill Stutz from the Donald Danforth Plant Science Center for their assistance in organizing the workshop. Graphic design in this white paper is credited to Bill Kezele. Finally, we like to thank Dr. Stephen Thomson and Dr. Ed Kaleikau, national program leaders at USDA–NIFA for their perspective, support, and input.