

National Institute of Food and Agriculture Summit on Data Science in Agriculture

Changing the Face, Place, and Space of Agriculture

EXECUTIVE SUMMARY

The National Institute of Food and Agriculture (NIFA) collaborated with the National Science Foundation-supported Midwest Big Data Hub at University of Illinois Urbana-Champaign and with the College of Agriculture and Life Sciences at Iowa State University to convene a summit on data science in agriculture. The purpose of this data summit was to achieve three broadly defined goals: (1) identify the current state of data in agriculture and capitalize on the potential of unprecedented opportunities created by the availability of big data to address major challenges in food and agricultural sciences; (2) formulate priority areas for future workshops that focus in detail on the needs for generating, managing, and integrating big data in the food and agricultural system; and (3) launch the Food and Agriculture Cyberinformatics and Tools (FACT) Initiative and highlight upcoming opportunities for workshops in FY 2017 for the stakeholder community. More than 85 distinguished leaders in the fields of data science and agriculture across the academic, industry, and government spheres attended the summit in person. In addition, more than 150 people joined the conversation remotely through social media outlets and through live streaming of the event.

NIFA Director Dr. Sonny Ramaswamy opened the event by announcing NIFA's new FACT initiative to support data science in agriculture. The subsequent keynote address and invited talks by several thought leaders in food and agriculture disciplines presented visions for the next phase of growth in big data and its use in the agricultural enterprise. Speakers articulated ideas that included: (1) agri-business collaboration and data exchange facilities; (2) anticipated data benefits and opportunities of data science; (3) predictive modeling for big data and genomics; (4) smart and connected communities: innovating cities of the future; and (5) modeling climate and environmental effects on food security. These talks set the stage for afternoon discussions in breakout groups organized around specific thematic areas.

Several common themes emerged from the breakout sessions and discussions throughout the summit:

- Cyberinfrastructure that includes open data, privacy, and sharing.
- Workforce/education/training - increased personnel and data competency for the public, farmers, extension professionals, consumers, scientists/researchers, etc.
- Standardization (metadata) across all agriculture-related data.
- Leadership by USDA and federal data platforms.
- F.A.I.R. Principles: Findable, Accessible, Interoperable, Reusable.

- Empowering community members to understand and to use data for making better decisions.
- Balancing federal oversight with creativity in big data.
- Quality and accuracy in big data.
- Incentives for safe data sharing. Identification of groups and what motivates them.
- Innovative engagement of youth in data topics - new types of programs that lead to jobs.

Summary by breakout sessions

The summaries below present key issues identified by attendees in six facilitated, thematic breakout sessions held during the summit. These summaries were derived from facilitator records, breakout group leader synopses, and note-taker discussion drafts. Ideas also are presented below that were generated through an on-line real-time Ideas Engine (Codigital, London UK), which was available for public input during several weeks immediately preceding and during the summit.

Thematic area: agricultural systems

What are the most promising opportunities for data-driven advances in agriculture and the food-production system?

Stakeholders discussed big data issues relevant to the thematic area of food production, food processing, food safety, and public health. Among the needs expressed repeatedly was the strong desire for improvements in data accessibility, data mining tools, data infrastructure, standardization of data, human resources, and transdisciplinary public/private interactions. Stakeholders seek to enhance the above aspects of big data to improve the predictive capabilities in the field and inform management decisions, ultimately leading to a more productive, sustainable, and economically viable agriculture system. Participants emphasized the need for resilient agricultural systems in the face of increased climate variability. Concerning data accessibility, stakeholders desire a shared, open, facile data infrastructure that allows multiple user types across disciplines (consumer, producer, farmer, etc.) to interact and to utilize data mining tools to conduct analyses that provide actionable findings is desired. Standardization of definitions, formats, and tools and greater transparency in data quality are also of interest. Finally, training, trust, and developing public-private partnerships were major themes.

Almost all of the top ideas from the ideas engine focused on the need for a platform that enables data sharing. Other relevant ideas from the ideas engine included:

- F.A.I.R. principles
- Adopting infrastructure from data mining powerhouses like Microsoft, Apple, Oracle, Google, Amazon etc.
- Federal funders need to work together and use a common proposal format
- Web based soil test calibration database
- Monitor plant and animal diseases to predict spread
- Need for reward system for data generation and sharing

Thematic area: data applications across sectors

What are the most promising opportunities for enhancing cross-sector advances in data applications?

To enhance cross-sector advances in data applications, stakeholders expressed that all collaborators must benefit from data sharing; a common vocabulary is needed; interoperability and a strong infrastructure are critical; a balance of privacy and regulation is required; and subject matter experts in agriculture fields require training to become adept in data management. Stakeholders discussed considerations related to private-public partnerships and data transfer and expressed the need for a governing body to oversee all of the proposed data systems changes and encourage incentives for safe sharing of data. Two recurrent human-focused themes were the need to establish and to maintain trust among all data users (farmers, scientists, consumers, etc.) and to educate existing and future users. For cross-sector, interdisciplinary work, researchers should hold collaborative learning (or co-learning) workshops in which team members can share their expertise and work together to generate new ideas and creative solutions.

Ideas from the ideas engine included:

- F.A.I.R. principles
- Need for reproducibility
- Revamping the reward system
- Data standards
- Demonstrated case studies that show value of data
- Addressing risk vs incentive for different types of data and different types of data producers

Thematic area: consumers and data

What are the most promising opportunities for data-driven advances to address societal well-being and consumer demands?

Stakeholders discussed big data issues relevant to the thematic area of consumer data. Among the themes that emerged were data provenance and ethics, data quality/accuracy, consumer access, communication, F.A.I.R. principles, and people-driven data. Transparency is essential, and stakeholders discussed the importance of data accessibility to lay users (for example, via an app) and the need to contextualize data. Stakeholders focused on the value of communicating to the community the impact of big data, including helping farmers and end-users make informed data-driven decisions. The issue of trust among all parties was repeated and led to a focus on collecting and assuring high quality and accurate data. Extension professionals, who conduct outreach, would benefit from increased training on best practices in data management, collection, analysis, and interpretation. To successfully address societal well-being, stakeholders articulated the need to develop a people-driven system that involves community members in all aspects of big data processes. The land-grant university and Cooperative Extension System exist as valuable entities for developing, teaching, and translating data needs, tools, and solutions.

Relevant ideas from the ideas engine included:

- Include two-way communication with the public
- Include “small data”
- Some type of organized oversight is needed to examine and address failures of “data-driven” decisions and make improvements over time

Thematic area: data management

What are the most promising opportunities to address challenges of various facets of data management and application?

On the topic of big data management, stakeholders expressed the need for open data, transparency, decision-support tools, cybersecurity, data integration, workforce development, standardization of vocabulary, F.A.I.R. data principles, development of shared goals, building university capacity to access big data, and creating automated systems for data analytics. Stakeholders cautioned that a shared, integrated system to reduce data redundancy is needed. Requiring the standardization of metadata across all agricultural disciplines were also proposed. Stakeholders expressed the need for oversight that strikes a balance between sharing of public and private information, reducing the risk of leaking personally identifiable information. They also discussed funding to fill future data workforce gaps and a desire to train professionals who understand the limitations of big data. Lastly, stakeholders acknowledged the need to incentivize many of these best practices in data management and application. Proper incentives might require changes to the tenure and promotion structure within academia and profit structures within the private sector. By collectively addressing these needs and challenges, stakeholders indicated that the agricultural enterprise has the opportunity to become more efficient, sustainable, and profitable for all parties.

Ideas from the idea engine not captured above include:

- Incentivizing the capture of unpublished data
- Automated annotation as data is collected

Thematic area: future workforce

What are the most promising opportunities to ensure future generations of data expertise?

When considering the thematic area of the future data workforce, stakeholders expressed a desire for well-connected, high-level, big data systems that are useful in furthering research, practice, and policy. Finding agreement across disciplines in protocols, data analysis knowledge, and skills is important. Linking resources across disciplines will support people-driven approaches to creating data-driven knowledge and insights. Stakeholders recommended that the agriculture enterprise use technology to increase collaboration, inclusion, knowledge exchange, and stakeholder engagement. Leveraging experience from both government and private sectors is important. Increasing stakeholder capacity around computational thinking is suggested; this would require supplemental funding. A common theme throughout the breakouts was educating the public to make good, data-driven decisions. The breakout group participants concluded by saying, “Together, the resources we bring to bear to address Director Ramaswamy’s charge, will indeed take us where no one has gone before in building the technological infrastructure and in creating the relationships, collaborations, and connections to harness the promise of big data for a better future.”

Some ideas from the ideas engine are directly captured here:

“Develop an experiential learning system for public-private partnerships that will match undergraduate students with teams of industry and/or community practitioners and university faculty to find data-driven solutions to real world problems.”

“A new type of AgSchool graduate / program is needed, one that combines agro-ecosystem knowledge with analytical skills in areas such as: machine learning and data mining; statistical and quantitative Analysis; data visualization and problem Solving.”

“Request all proposals in the field of big data development and application to have a significant extension and outreach component. This increases public awareness and promotes grass-root activities toward incorporating big data in teaching curricula.”

“Engage biology high school teachers & students in data sciences using free, online resources (e.g., NCBI, UniProt) to introduce fundamental principles of data science at an early stage.”

“Engage biology undergraduates in data management courses & projects, in addition to more traditional bio-informatics & computer science courses.”

“Show students in middle school and high school that data acquisition, sharing and management is exciting and useful and that there will be jobs available in systems building and application, then make it happen at universities and companies.”

Thematic area: communication, ethical and legal

What are the most promising opportunities for big data in communication, property rights, and communities?

Stakeholders discussed big data issues relevant to the thematic area of the communication and legal issues. Central themes included capacity building of extension professionals, community-collaborative science, F.A.I.R. principles, cyberinfrastructure, property rights and privacy. Sharing best practices that connect farmers, researchers, producers, consumers, and others is important. Determining data use and ownership are critical. Involve communities early on, and constantly engage. Creating strong guidelines for the sharing of private information is suggested. AgGateway was mentioned as a possible collaborator.

One idea from the ideas engine is directly captured here:

“Access to local data and information through public systems can increase individual stakeholder awareness of land use, climate patterns, water use, or other common topics. Access can help to break down knowledge barriers and serve communities.”