# Big Data and Computing

## Building a Vision for ARS Information Management

Workshop Summary

Feb. 5-6, 2013
USDA Agricultural Research Service

# TABLE OF CONTENTS

## INTRODUCTION

On February 5-7, 2013, scientific leaders from the Agricultural Research Service (ARS) held a workshop to identify scientific information management needs within the agency, and to find solutions to address those needs. Workshop participants represented scientists from all National Programs and each geographic area. Speakers from industry, academia, and federal agencies provided information about their experiences with Big Data. Participants were challenged to articulate a vision for ARS information management unconstrained by current ARS capabilities and to formulate a strategy to achieve that vision. This document outlines the recommendations arising from that workshop.

The ARS is a worldwide leader in agricultural research, providing a unique breadth and depth of scientific expertise to address problems of national and international importance. ARS scientists use leading-edge methods to solve problems and answer questions across a broad array of disciplines related to agriculture. The scope of research undertaken by the agency is highlighted by the ARS Mission Statement:

> *ARS conducts research to develop and transfer solutions to agricultural problems of high national priority and provide information access and dissemination to:*
> - *ensure high-quality, safe food, and other agricultural products*
> - *assess the nutritional needs of Americans*
> - *sustain a competitive agricultural economy*
> - *enhance the natural resource base and the environment, and*
> - *provide economic opportunities for rural citizens, communities, and society as a whole.*

However, the nature of the science supporting this mission is changing rapidly. In the past, scientific methods were often labor-intensive and consumed many person-hours to adequately address a single scientific question. Scientists are now generating vast amounts of high-quality data rapidly and relatively inexpensively. This fundamental change in the nature of science is presenting new challenges and demanding new approaches to maximize the value extracted from these large and complex datasets. This dramatic growth in data volume, variety, and velocity has come to be known as Big Data (Box 1).

**Box 1**

### The *Bigness* of Big Data

The phrase 'Big Data' obviously implies volume, but IT industry analysts and consultants have long acknowledged other characteristics of data that add commensurate challenges. The discussion remains vigorous, to the point of moving from the description of Big Data to that of Extreme Data.

*For the purposes of this white paper, Big Data is characterized as having extreme or variable values of one or more of the following characteristics:*

- *Volume*[1] *(size)*
- *Variety*[1] *(structure)*
- *Velocity*[1] *(acquisition rate)*
- *Veracity (uncertain quality or provenance)*
- *Variability*[2] *(in meaning)*
- *Complexity*[3] *(in relation-ships, sources, etc.)*

1. Doug Laney. 2001. 3-D Data Management: Controlling Data Volume, Velocity, and Variety.
2. Brian Hopkins. 2011. Blogging From the IBM Big Data Symposium - Big Is More Than Just Big. 2011. Brian Hopkins.
3. Valentin Sribar. 2011. 'Big Data' Is Only the Beginning of Extreme Information Management.

As a result of these changes, a new paradigm is emerging in science that is characterized by its data intensity. Previous methods for data collection, storage, and analysis are inadequate for handling the scale and complexity of this avalanche of new data. Consumer-oriented services such as Google Maps, Facebook, Wikipedia, and the National Oceanic and Atmospheric Administration (NOAA) weather forecasts, demonstrate the benefits that can be gained by aggregating data from multiple sources. These products have become essential resources intensively adopted by millions of people in an astonishingly brief time. Similarly, the vast data resources produced by ARS scientists could be used to describe the world in ways never imagined, but critical resources and coordination to maximize this impact are missing. A few examples of agricultural research that are enabled by, and even *necessitate*, such a paradigm shift in scientific information management include:

- **Increasing the resilience of production systems:** Deriving local or regional responses to environmental stressors requires defining the complex interactions among crops, animals, soil, water, weather, and climate. A better understanding of these interactions can be gleaned, even using current data or data collection strategies, but only with an improvement in how the vast amounts of data are integrated across locations.

- **Improving our understanding of genotype by environment interactions:** In all organisms, the environment affects phenotype, and variability in the environment can confound attempts to identify genes underlying important agronomic traits, and even human diseases. Traditional studies typically only identify genes with large effects, but many important traits are driven by multiple genes that are easily affected by the environment. Identifying these genes requires whole genome studies coupled with good environmental data for each individual. Such studies require high computing capacity for analyzing very large datasets.

- **Enhancing human and animal health:** Human and animal nutrition, health, and well-being can be better predicted by understanding the rich interactions among microbial communities through insights gleaned from the genomes of those community members.

For a more extensive discussion of these and other examples, see Appendix 1 (p. 11).

**Enhancements in scientific computing in the ARS are both *critical* and *urgent*.** The ARS must upgrade computing and data management infrastructure to maintain its status as a world leader in the science of agriculture. The Big Data paradigm shift in science is taking place across disciplines, and action is necessary for ARS scientists to effectively perform key research missions. Enhancements to current systems are vital for maintaining the ARS as an effective, nationally-coordinated, multidisciplinary agency. Current research priorities target objectives that are unattainable without some broad, systematic improvements. ARS scientists cannot perform state of the art research without the appropriate tools, knowledge, and skills.

## BIG DATA AND COMPUTING

Participants at the Big Data Workshop expressed enthusiastic support of the worldwide leadership provided by the ARS in agricultural research and embraced the role of the agency to lead in the collection, storage, analysis, and distribution of scientific data related to agriculture (see Box 2). Workshop participants agreed that the **ARS is uniquely positioned to provide long-term, stabile, durable solutions for information management challenges involving Big Data**. Participants reported that urgent solutions are needed to address these technology issues to help meet scientific needs, needs that reach across agricultural research institutions and disciplines. This enhanced role for the ARS is consistent with the Big Data initiative recently announced by the White House Office of Science and Technology Policy (OSTP).

**Box 2**

### Big Data Vision Statement

*Scientists in the ARS and elsewhere are generating data at increasing rates. To maximize knowledge extracted from these large and diverse data sets, ARS should support scientific computing to efficiently combine disparate information for scientific discovery, and to enable the transfer of that knowledge quickly and efficiently to other scientists and to the public.*

The enhancements envisioned here will:

- enable or enhance data exchange and analysis
- create a more robust network of computer resources
- encourage data standardization, and
- help to develop a well-trained scientific and technical staff

These changes require support for ongoing learning by technical and scientific staff to facilitate adaptation to new data acquisition and computing methods. ARS should not only be able to quickly adapt to new methods, but be on the forefront of developing such resources for the US agricultural research community.

### Defining Research Needs

The first day of the workshop was dedicated to identifying the scientific needs surrounding information management within the ARS, and to determining IT and infrastructure issues associated with Big Data. In particular, participants were asked to identify some examples of scientific problems that could be newly addressed if better tools for data management and analysis were available. For the meeting agenda, see Appendix 2 (Pg. 13).

The workshop participants concluded that the ARS needs to make key investments in computing infrastructure now to establish and maintain leadership in agricultural scientific computing and to keep pace with the explosive growth in data and corresponding analyses. Data acquisition is

growing at exponential rates from nearly every source – including genomics, environmental monitoring and remote sensing, and in various other fields that impact agriculture. To leverage these advancements, workshop participants determined that the ARS needs:

- **a robust network of computer resources** to reduce redundancy in hardware purchases and increase computational efficiency across the agency;
- **an improved capacity to share data** across laboratories, institutions, and disciplines;
- **enhanced knowledge, skills and abilities within the ARS workforce** to enable the use of Big Data management and analysis strategies to solve agricultural research problems
- **standards for data collection, structure, and documentation**, and a means for developing and deploying these standards;
- **improved communication systems** that help scientists, IT experts, and the Office of National Programs (ONP) access and share new knowledge about data collection, including better communications regarding quality control, analysis, and interpretation;
- **new methods for dealing with both unstructured and complex data**.

## Steps to Achieve the Vision: Specific Recommendations

The second day of the workshop was dedicated to identifying how the agency might better meet the needs listed above (see Appendix 2, pg. 13). For the ARS to maintain its standing as a premiere scientific organization, deployment of state-of-the-art information management solutions is urgent and critical. The implementation of these improvements requires large-scale investments and substantial changes in human, computing, and financial resource management. Workshop participants articulated the following recommendations, which are also summarized in Table 1 (p. 9).

- **Establish an Advisory Committee for scientific computing, led by a Chief Scientific Information Officer** (CSIO) to develop strategic plans and policies for scientific computing, specifically for the purpose of advancing the scientific missions. No current entity in the ARS sufficiently represents the breadth of ARS research and informatics expertise to house this office. The mission of the CSIO should cut across all National Programs and serve as a bridge between the ONP, the ARS Chief Information Officer (CIO) and scientists. The Advisory Committee should be responsible for making recommendations regarding IT services, infrastructure, and training needed to support the use of Big Data. The CSIO should have a budget to help implement these recommendations, with sufficient initial funding to enable the assessment described below, and to provide travel for organizing the working group. Future budget needs would be contingent on the assessment and recommendations of the Advisory Committee. Targeting infrastructure and resource improvements to high-impact research areas is important, but priority should also be given to ensuring that those improved resources are made accessible to all potential users. The CSIO should be established at a level within the agency that reflects the critical nature of these responsibilities.

  Initially, an interim CSIO could be detailed into the position, if necessary to fill the position quickly. However, the position is not only urgent, but will also remain critical over the long

term. The CSIO should work with the Advisory Committee to gain maximal input and represent the ARS with regard to the diversity of scientific disciplines, IT expertise, and geographic distribution. Together, the CSIO and Advisory Committee would be responsible for (1) developing computing and data policies in support of the scientific mission, (2) establishing agency priorities that enhance scientific capabilities associated with Big Data, (3) facilitating communication within the agency, with other federal agencies, universities, domestic and international research partners, and (4) assisting in decisions pertaining to scientific information management. The governance and structure should reflect scientific needs in the ARS and incorporate protocols that have proven to be successful in similar bodies at other federal or academic facilities[*]. The intention behind the creation of the CSIO and Advisory Committee is to facilitate communication across the agency and to assist with formulating and making recommendations, particularly regarding resource availability and utilization.

- **Conduct an assessment of agency resources** associated with scientific computing, such as human resources, computation hardware, data storage, and network capacity—including resources utilized by, but not necessarily maintained or owned by the ARS. Following the model and advice of representatives for a similar assessment conducted by the National Institutes of Health (NIH), the assessment could be done by a third-party firm or contractor (as reasonable to enable a quick turn-around and in keeping with available resources). The assessment is foundational, and thus is a top priority. This review will provide critical information regarding current resources and future needs.

- **Develop computing and data policies in support of the scientific mission**, perhaps at the REE level. Current administrative data security policies substantially hinder researchers' abilities to work in collaborative teams on data intensive problems. Rapid deployment of a scientific data policy, one that distinguishes it from the administrative personal identifiable information (PII) and financial data, would greatly facilitate sharing of data and analytic tools among researchers and other customers within and outside the agency. A clear distinction between computers warehousing sensitive data and those that do not could ease access and reduce administrative overhead on non-sensitive systems.

- **Invest in high-priority enhancements in scientific IT capabilities**. Improve data storage, computational resources, and network infrastructure. Identify, deploy, and if needed, develop data standards. Speed up the full rollout of ARSnet 2.0 and immediately investigate ways to continue to improve bandwidth to all ARS locations.

- **Develop human resource policies that keep the ARS on the scientific forefront.** Recruit and retain personnel with expertise in state-of-the-art information management skills. Develop new position descriptions and classifications that better reflect the work conducted in the computer science-related positions, e.g. database administrators, geographical information systems specialists, computational biologists and image analysts. To be

[*] For example: Riding The Wave, How Europe Can Gain From the Rising Tide of Scientific Data. 2010. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf; or NIH Working Group on Data and Informatics, 2013. http://acd.od.nih.gov/diwg.htm

successful, these positions often require strong skills across multiple disciplines. The ARS needs better options to successfully recruit and retain candidates with these skill sets, including development of appropriate promotion options.

Train and educate current ARS personnel to deal with Big Data challenges. The ARS should systematically assess its scientific Big Data needs in computational methods, data management, and data visualization. The agency should use this information to identify appropriate training resources and commit the time and financial resources to bridge knowledge gaps within the agency.

Recognize that many critical scientific outputs are products that can be evaluated independently of the standard scientific journal article, including software, datasets and models. These outputs can have substantial impact for stakeholders and end users, and in many cases their impact eclipses that of peer reviewed publications. Scientific staff need recognition for creating and sharing these products.

- **Promote documentation and sharing of scientific data.** Better mechanisms for publicly sharing data must be developed to overcome current deficits and future needs. Of particular concern is the need for a broad adoption of relevant national and international standards for data format, transfer, transformation, and interpretation. Communities of interest and working groups should be established to help develop these standards, and experts outside of the ARS should be included when appropriate. The National Agricultural Library (NAL) could play a major role in these activities, especially in serving as a repository of standards, recommendations, best practices, computer code, and example data sets.

## CONCLUSIONS

Improvements are needed within the ARS to increase scientific capacity and keep pace with new developments in computer technologies that support data acquisition and analysis. Enhancements in computing power and IT infrastructure are needed to provide scientists better access to high performance computing, networking, and greater data storage capacity. However, improving scientific capacity as it relates to Big Data involves more than just hardware and software. Elements of the ARS's organization also limit the agency's ability to achieve maximal scientific impact. Information management within the ARS should give high priority to developing computing, data security, and human resource policies that advance the scientific mission. New methods that enable the use of Big Data are needed, both to facilitate sharing datasets and to make ARS data and software products more easily accessible to customers. The Big Data workshop and this document provide a snapshot "view from the field" and are the first steps in defining how the ARS should improve its use of Big Data to address agricultural problems, allowing scientists to gain new knowledge and address critical, recalcitrant questions. Enabling broad access to data through such systems could eliminate the bottleneck that is beginning to constrain agricultural advancement, opening a world where research is limited only by human creativity, which is boundless.

## Table 1. Summary of Recommendations.

| Action | Whom Takes Leadership |
|---|---|
| *Needed immediately* | |
| Create a Chief Scientific Information Officer (CSIO) position (perhaps initially a detail) and fill the position | Office of Administrator, ONP |
| Seek nominations for and appoint a CSIO Advisory Committee for Big Data issues | CSIO, ONP and Office of Administrator |
| Assess resources currently used by the ARS for scientific information management | CSIO, Advisory Committee, and OCIO |
| *Early, short-term adoption needs* | |
| Develop scientific data policies that distinguish scientific data from that which requires higher levels of security, such as PII and financial data | CSIO, Advisory Committee, and OCIO |
| Develop/revise position descriptions and classifications that better define more data-science related positions | ONP, Human Resources, CSIO, Advisory Committee |
| Establish communities of interest and working groups to enhance scientist-to-scientist communication | ONP, CSIO, OCIO |
| Assess training needs, identify appropriate training resources for current staff associated with Big Data | CSIO, Advisory Committee, Area Offices |
| Deploy ARSnet 2.0 fully and improve bandwidth at ARS locations | OCIO |
| Revise SY evaluation performance plans and evaluation policies to include publication of programs, datasets, and models in addition to peer-review publications | Area Offices |
| *Intermediate development* | |
| Develop systems that segregate administrative PII and financial data from scientific data | OCIO, CSIO, Advisory Committee |
| Establish a repository of standards, best practices, computer code, etc. to enhance scientific data sharing | NAL, CSIO, Advisory Committee, ONP |
| Where needed and when lacking, develop agricultural specific computational methods and data standards | CSIO, Advisory Committee |
| *Long term development* | |
| Improve data storage, computational resources, and network infrastructure | CSIO, Advisory Committee, OCIO |

## Acknowledgements

**This document was prepared as a collaborative effort among ARS scientists:**

| | | |
|---|---|---|
| Laj Ahuja | Ivan Baxter | Steven Cannon |
| Rosalind James | Carolyn Lawrence | Joan Lunney |
| Sharon Papiernik | Caird Rexroad III | John Sadler |
| Thad Stanton | Charles Stephensen | Curt Van Tassell |

**Special thanks to the workshop speakers who informed and inspired workshop participants and contributed greatly to its outcomes:**

| | |
|---|---|
| Sean Davis | National institutes of Health |
| Jay Evans | ARS |
| Derrick Fouts | J. Craig Venter Institute |
| Paul Gibson | ARS |
| John Helly | San Diego Supercomputer Center |
| Edward Knipling | ARS |
| Bill Kustas | ARS |
| Simon Liu | ARS-NAL |
| Laxmi Parida | IBM Research |
| Jim Reecy | Iowa State University |
| Caird Rexroad, Jr. | ARS |
| Jeff Silverstein | ARS |
| Matthew Vaughn | Texas Advanced Computing Center |
| Doreen Ware | ARS |
| George Wiggans | ARS |
| Hai Zhu | DuPont de Nemours & Co. |

**We also give special thanks to all the other participants of the workshop.**

## Appendix 1: Examples of How Improved Big Data Capacity Will Increase ARS Scientific Capacity

### 1. Environmental Modeling

In the ARS, Natural Resources and Sustainable Agricultural Systems research is conducted at multiple locations characterized by variable agro-climatic conditions. These varying conditions present both challenges and opportunities. Studies that measure the impact of environmental conditions on agricultural systems often yield conflicting results at different locations because of complex interactions among the crop, soil, water, weather, climate, and management differences. For example, the practice of no-till has been shown to enhance precipitation storage in soil and lead to increased wheat and summer crop yields near Akron, CO. Alternatively, at Pendleton, OR, no-till management decreases soil moisture storage and wheat yields. Using Big Data approaches, results could be analyzed across locations and years to better understand and quantify the complex interactions that produce different results. These tools would allow new discoveries that enable researchers to:

- improve the definition of local best practices over long-term weather conditions;
- innovate, explore, and discover new management practices;
- extrapolate the results to other locations and agro-climatic conditions;
- integrate results over a region;
- evaluate future climate change effects; and
- develop theoretical frameworks and predictive tools of agricultural science.

The CGIAR Science Council[†] noted in their research priorities: "Modeling and the ability to combine data from different sources, promises to revolutionize understanding of processes affecting management of natural resources." Cross-location analyses will reduce duplication of research at multiple locations and lead to a more efficient use of allocated funds.

### 2. Complex Genetic Interactions

Methods that enable analysis of Big Data could also be used to improve breeding programs by allowing scientists to better predict genotype by environment interactions (GxE). The environment of an organism affects phenotype, and variability in the environment (which inevitably occurs) can confound attempts to identify genes that underlie important agronomic traits. Traditional attempts to identify genes in an agronomic context have circumvented this problem by limiting the search to loci that show a large effect across many environments. But, many agronomically important traits are driven by multiple genes, each with small effect, that have variable effects in different environments. To identify these genes, we need the ability to record and quantify environmental information associated with each individual, a process that is currently quite difficult. Improvements in ARS computing capacity would help to resolve this issue. Because many agricultural research questions depend on environmental conditions,

† The Consultative Group on International Agricultural Research Science Council. 2005. System Priorities for CGIAR Research 2005–2015. Science Council Secretariat: Rome, Italy.

integrating environmental data recording could be seamlessly incorporated into phenotyping pipelines. For example, individual plants and livestock can be represented by GPS coordinates. Once the organism has been localized in time and space, a wide variety of environmental data and projections can be added to the system. Examples include soil, weather, or satellite data, and models for parameters such as drought. Much of this data exists in databases and sites produced or supported by the ARS and other government agencies. Ideally, such a system would integrate these disparate data sources with locally collected environmental, genetic, and phenotypic data into a common framework. An example of such research would be to integrate weather and precipitation characteristics to predict variety performance in an arbitrary location with different climatic conditions.

**3. Genomic Data Analyses**

Major computational challenges face ARS laboratories engaged in genomic studies, including those researchers focused on human health (such as nutrition and aging) and livestock or plant breeding. A major challenge is termed “the big P small N problem,” or one of estimating many parameters (P) with relatively few observations (N). For example, a typical human genetic dataset might include several million genetic markers and a few thousand dietary/lifestyle measures and disease-related phenotypes from several thousand subjects. The computational challenge of trying to predict millions of marker effects for thousands of phenotypes based on several thousand observations requires fitting many complex statistical models. Data transfer and storage are ongoing issues. This challenge is only expected to grow given the onslaught of data being collected and generated. For example, the ARS Nutritional Genomics Laboratory at Tufts University studies human nutrition and its relationship to common diseases. One research objective is to characterize how dietary intake and lifestyle modulate the relationship between genetic factors and disease. Researchers in this lab would like to predict how nutrition affects human risks to disease. Analysis of genetic differences must be associated with dietary intakes, lifestyle data and disease status. Improvements in computational infrastructure would permit more rapid analysis and enhance the impact of such research.

**4. Understanding Microbial Communities**

Analyses of Big Data could provide new tools for better understanding microbial populations related to human, animal, and plant health. For example, the diverse microbial communities (microbiome) that exist in the intestinal tracts of animals and humans make essential contributions to the nutrition, digestion, immunity, organ physiology, disease resistance, and perhaps even the behavior of the animal hosts. Big Data analyses of the complexity and role of local microbial populations provide an unprecedented view into the key players in the intestinal or lung microbiome, their millions of genes and gene products, and their mostly unknown intercommunications with host tissues. Before high-throughput DNA sequencing, the tools to understand these populations were rudimentary; modern metagenomics and phylogenetics has transformed these analyses. Understanding these microbes and exploiting beneficial properties will enhance animal and human health and well-being through better diet formulations, enhanced early life immunity, and reduced food safety concerns. The explosion of metagenomic knowledge has only begun to be explored for our soil and water resources.

# Big Data & Computing: Building a Vision for ARS Information Management

## Tuesday, February 5, 2013

***Morning Session I, Rosalind James*** *moderating*

8:15 Welcome and Introduction to the Purpose of the Workshop

8:35 Comments from the ARS Administrator's Office
**Caird Rexroad**, ARS Associate Administrator

8:45 Bandwidth, and Other Big Data Issues in OCIO
**Paul Gibson**, ARS Chief Information Officer

9:00 BIGDATA, littledata and EvErYtHiNg in Between: Strategies for Scientific Data Management
**John Helly** (Plenary), San Diego Supercomputer Center
Scripps Institution of Oceanography/Climate, Atmospheric Science and Physical Oceanography

9:45 Making sense of Big Data
**Laxmi Parida**, IBM Research, NY

10:15 BREAK

***Morning Session II, Carolyn Lawrence*** *moderating*

10:30 U.S. and European Union Plant Biotechnology & Big Data
**Doreen Ware**, ARS Cold Spring Harbor, NY

10:50 High-performance Computing Needs in Animal Agriculture
**Jim Reecy**, Iowa State University, Ames, IA

11:10 Managing Large-Scale modeling and Remote Sensing Datasets for Agricultural Drought Monitoring
**Bill Kustas**, ARS Hydrology & Remote Sensing Laboratory, Beltsville, MD

11:30 Big data in support of genetic improvement of dairy cattle
**George Wiggans**, ARS Animal Improvements Program Laboratory, Beltsville, MD

11:50 Charge for Break-out Sessions, **Carolyn Lawrence**

12:00 LUNCH

12:30 Break-out sessions

15:30 Return to the Main Conf. Room. Coffee break and set up posters.

15:40 Break-out Session Reports, Poster-style
**Carolyn Lawrence** moderating

17:00 ADJOURN

## Wednesday, February, 6th

8:00 Welcome, & Charge for the Day **Rosalind James**

***Morning Session I, Curt Van Tassell*** *moderating*

8:10 Introduction to and Status of iPlant & iAnimal
**Matthew Vaughn**, iPlant Collaborative, Texas Advanced Computing Center, Austin, TX

8:45 High Performance Computing at the National Institutes of Health
**Sean Davis**, National Cancer Institute

9:15 Computational and Biological Challenges Facing Metagenomic Studies
**Derrick Fouts,** J. Craig Venter Institute Rockville, MD

9:45 BREAK

***Morning Session II, Sharon Papiernik*** *moderating*

10:00 How and Why DuPont Built a High Performance Cloud Computing System: Advice from Experience
**Hai Zhu** (Plenary), DuPont de Nemours & Co., DE

10:45 To Infinity and Beyond… Big Data and ARS
**Jeff Silverstein**, ARS National Program Leader, Animal Production and Protection

11:10 Digital Informatics Vision for the National Ag Library
**Simon Liu**, National Agricultural Library, Beltsville, MD

11:35 Final Comments from the ARS-CIO
**Paul Gibson**, ARS Chief Information Officer

11:50 Charge for Break-out Sessions, **Sharon Papiernik**

12:00 LUNCH

12:30 Break-out sessions

15:30 Return to the Main Conf Room. Coffee break and set up posters.

15:40 Break-out Session Reports, Poster-style
**Sharon Papiernik** moderating

16:50 Group Discussion and Workshop Summary
**Rosalind James** moderating

17:00 MEETING CLOSE OUT & ADJOURN

## Thursday, February 7th

8:00 Charge, writing begins **Rosalind James**

12:00 Lunch

13:00 Complete document together

17:00 ADJOURN

## Appendix 3. Acronyms Defined

| | |
|---|---|
| ARS | Agricultural Research Service (within USDA) |
| CGIAR | The Consultative Group on International Agricultural Research |
| CIO | Chief Information Officer |
| CSIO | Chief Scientific Information Officer |
| IBM | International Business Machines |
| IT | information technology |
| NAL | National Agricultural Library |
| NIH | National Institutes of Health |
| NOAA | National Oceanic and Atmospheric Administration |
| OCIO | Office of the Chief Information Officer |
| ONP | Office of National Programs (within ARS) |
| OSTP | White House Office of Science and Technology Policy |
| PII | personal identifiable information |