National Agricultural Research Data Network for Harmonized Data (NARDN-HD)



National Research Support Project (NRSP)
NRSP_TEMP11

University of Florida & Partners

Presented by Cheryl Porter

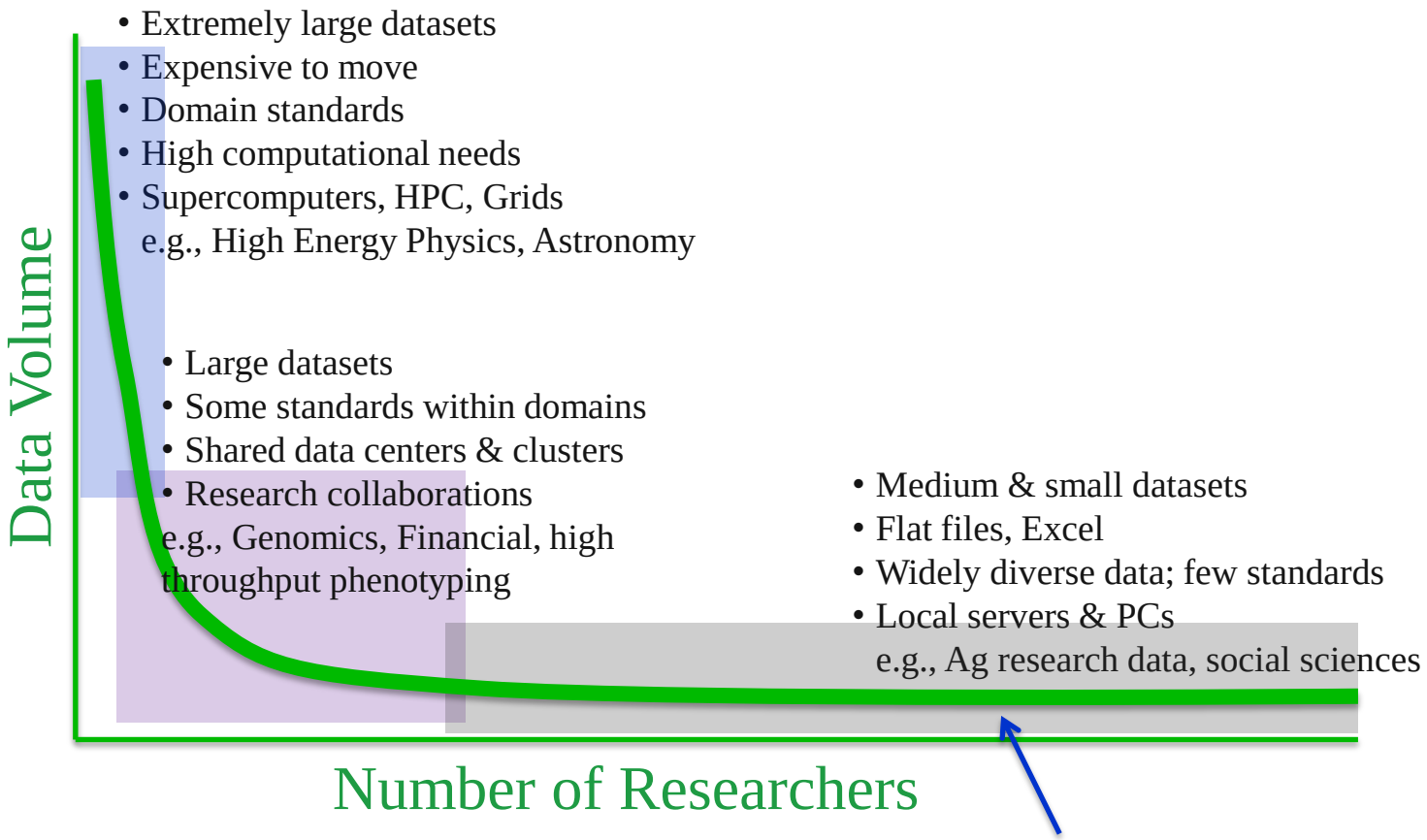
SAAESD Joint Spring Meeting
April 27, 2016, St. Thomas, VI

Outline

- Background & Need
- National Agricultural Research Data Network – Harmonized Data
  - Objectives
  - Structure, Characteristics & Components
  - Contributors & Milestones
- Questions

NARDN-HD

Data Intensive Scientific Discovery



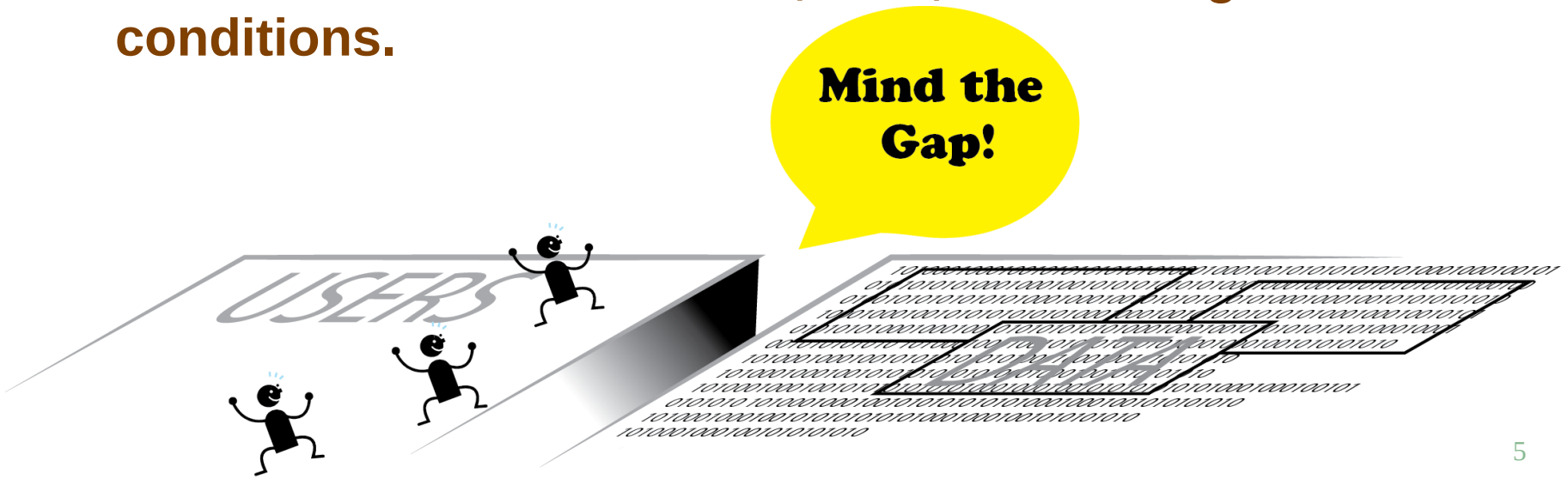
Tony Hey, 2016
http://www.slideshare.net/JISC/the-fourth-paradigm-data-intensive-scientific-discovery-jisc-digifest-2016/4

Background and Need

- ✓ **Research is essential to continually improve agricultural systems needed to meet the food, fuel, and fiber needs**
- ✓ **Experiment Station researchers are known for the quality of experiments and data that they collect and for providing science that keeps US agriculture the envy of other nations**
- ✓ **Many more benefits could be gained by making data available and usable across years and regions**

The Data Gap

- ✓ There is a major gap between the potential value of data collected in agricultural experiments and the value currently obtained through use of those data.
- ✓ Typically, data collected in experiments are used for the original research purpose only.
- ✓ Vastly greater value might be obtained if the data were combined across locations, time, and management conditions.



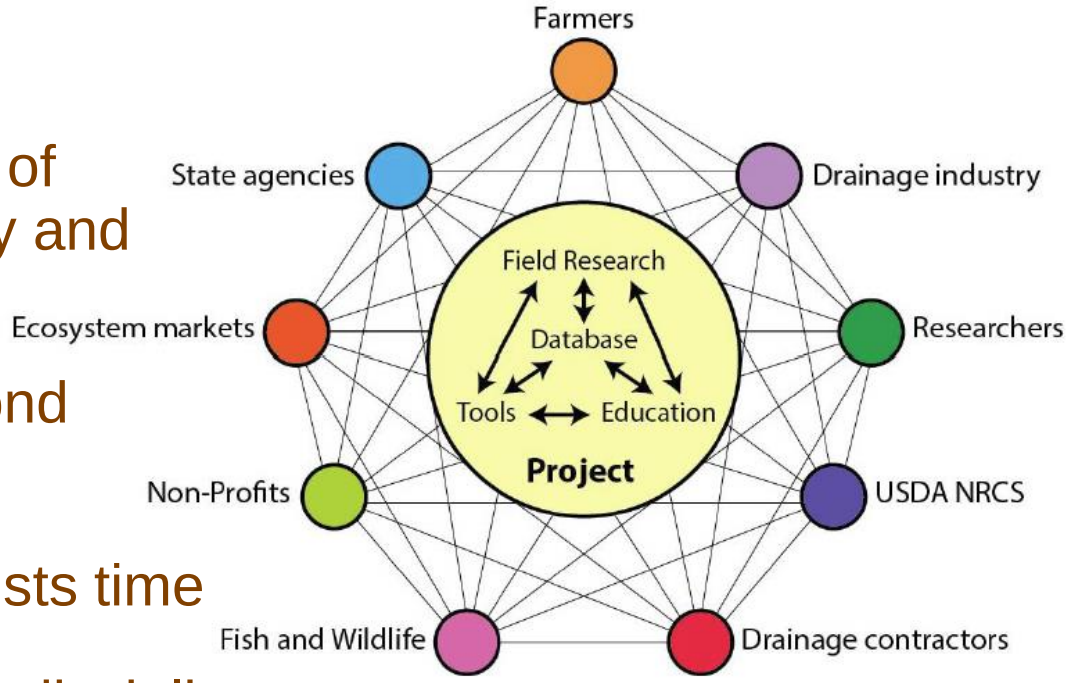
NARDN-HD

Examples of data intensive scientific discovery

- ✓ **Provide understanding of genetic, environment, and management (G * E * M) effects on production to further increase productivity and sustainability,**
- ✓ **Provide the science knowledge base for researchers to develop next generation models of agricultural systems and decision support systems, and statistical, visualization and other analytical tools to answer questions,**
- ✓ **Meta-analyses over many environments and management conditions to support evidence-based decision-making.**

Open Ag Data: The Carrots

- **Advancement of science**
- Refinement and expansion of research questions spatially and temporally
- Data available for use beyond original scope
- More efficient use of scientists time
- Collaboration in and across disciplines
- Improved transparency & reproducibility of findings to funders and other researchers

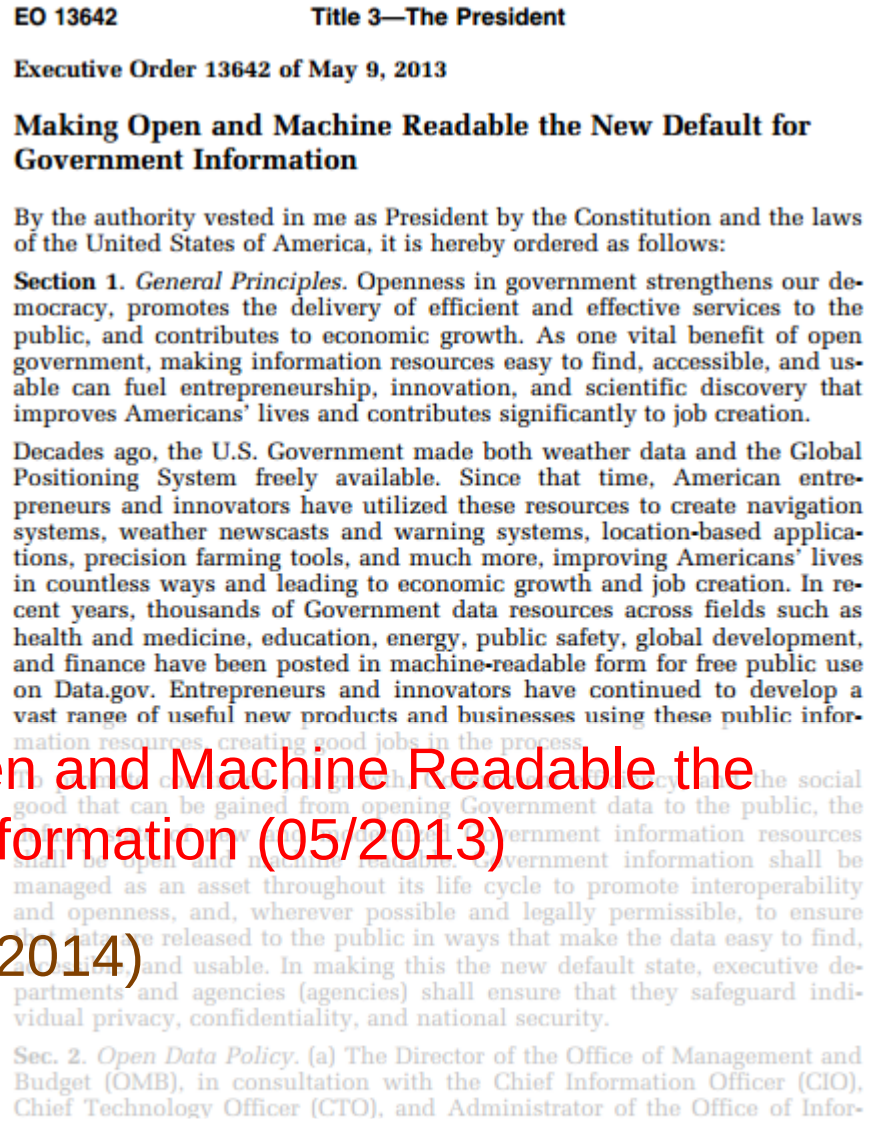


From L. Abendroth, Corn CAP Data PI, Sustainable Corn.org

NARDN-HD

Open Ag Data: The Sticks

- Mandates
  - America COMPETES Reauthorization Act (12/2010)
  - Office of Science & Technology Policy (OSTP) Public Access Memo (02/2013)
  - Executive Order – Making Open and Machine Readable the New Default for Government Information (05/2013)
  - US Open Data Action Plan (05/2014)



EO 13642 Title 3—The President

Executive Order 13642 of May 9, 2013

Making Open and Machine Readable the New Default for Government Information

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. *General Principles.* Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

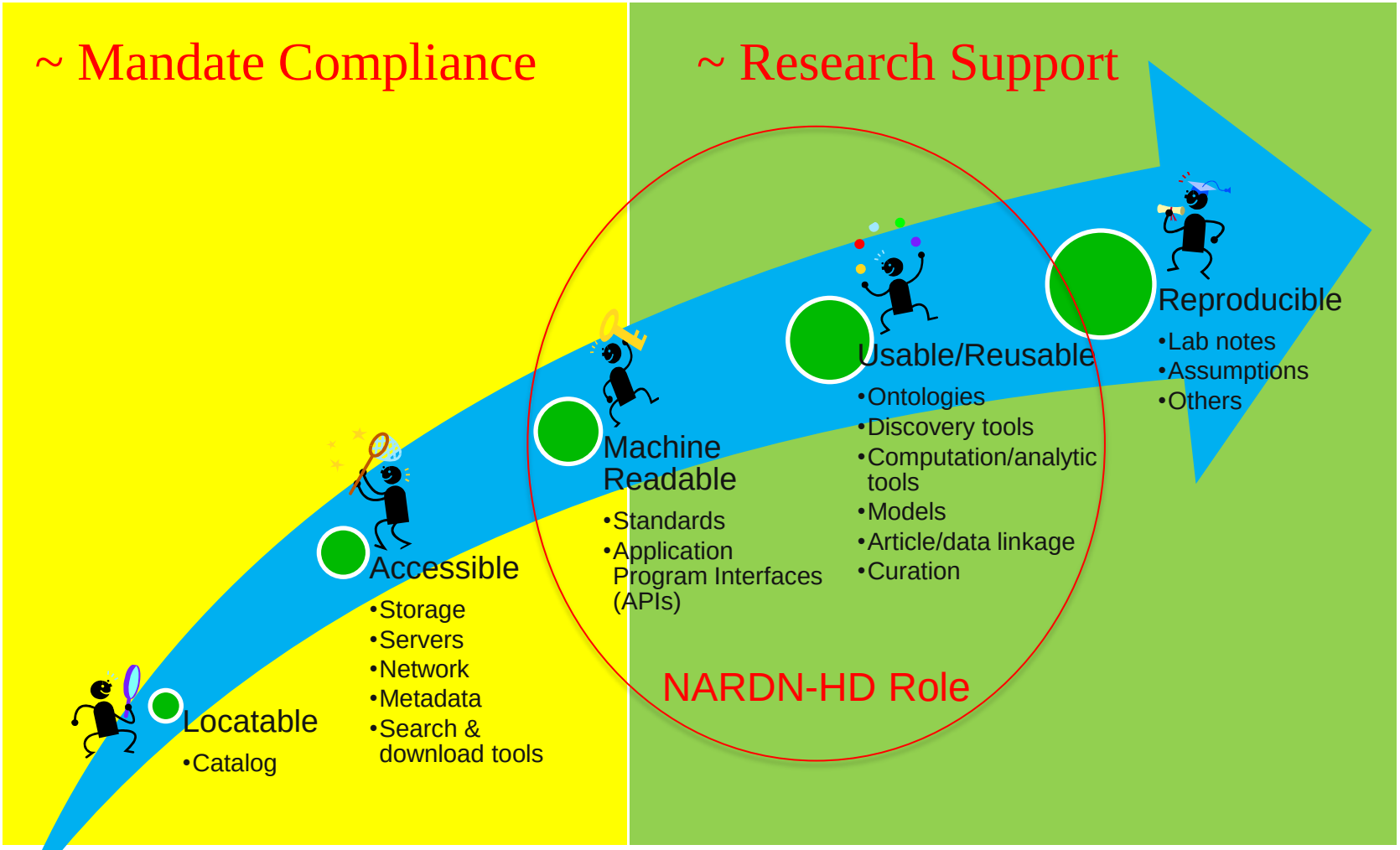
To promote continued innovation and economic opportunity, and the social good that can be gained from opening Government data to the public, the default state of new and modified Government information resources shall be open and machine readable. Government information shall be managed as an asset throughout its life cycle to promote interoperability and openness, and, wherever possible and legally permissible, to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable. In making this the new default state, executive departments and agencies (agencies) shall ensure that they safeguard individual privacy, confidentiality, and national security.

Sec. 2. *Open Data Policy.* (a) The Director of the Office of Management and Budget (OMB), in consultation with the Chief Information Officer (CIO), Chief Technology Officer (CTO), and Administrator of the Office of Infor-

NARDN-HD NRSP

- ✓ National effort is needed to allow researchers to comply with these mandates for federally-funded projects to make their data open, accessible and interoperable.
- ✓ More importantly, it will open up opportunities for new scientific discoveries via use of big data and analytics that are increasingly being used across sectors
- ✓ Opportunity for creating a virtual research laboratory for creating next generation models, analytical tools, and decision support systems

A Logical Journey

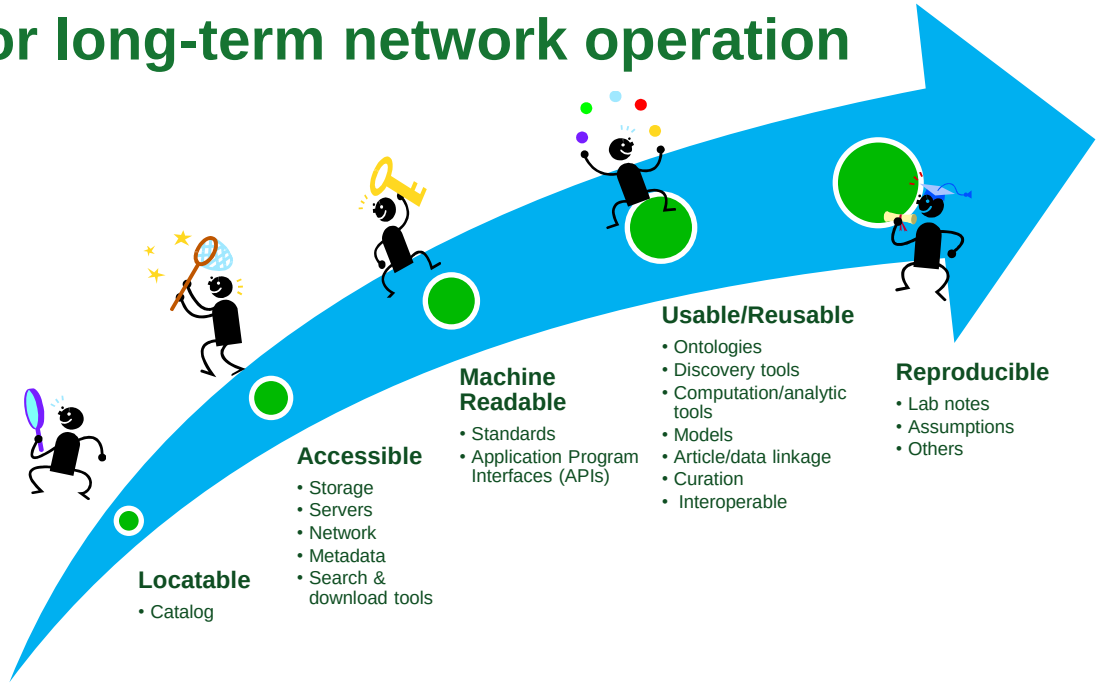


From Simon Liu, USDA/ARS May 2015

NARDN-HD

NARDN-HD: Objectives

1. Create distributed network for harmonized crop & livestock data
2. Devise common metadata for those systems
3. Develop tools for discovering, accessing, and using the data
4. Develop tools & procedures for researchers to contribute data
5. Develop plan for long-term network operation

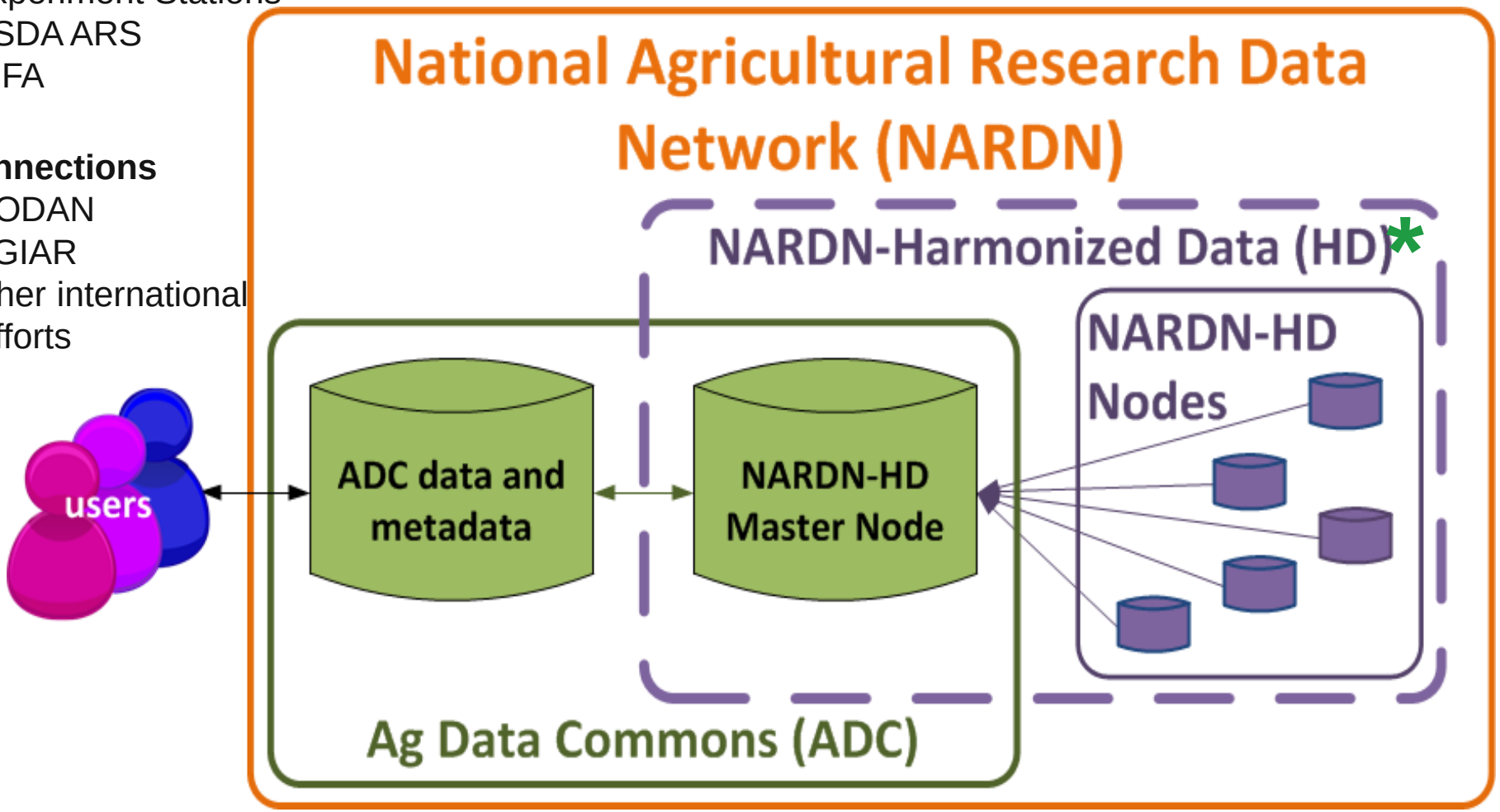


NARDN-HD

NARDN-HD Structure

Partners
- National Agricultural Library
- Experiment Stations
- USDA ARS
- NIFA

Connections
- GODAN
- CGIAR
- -other international efforts



* *Translated into a common set of variable names, units, and formats*

GODAN

Global Open Data for Agriculture and Nutrition

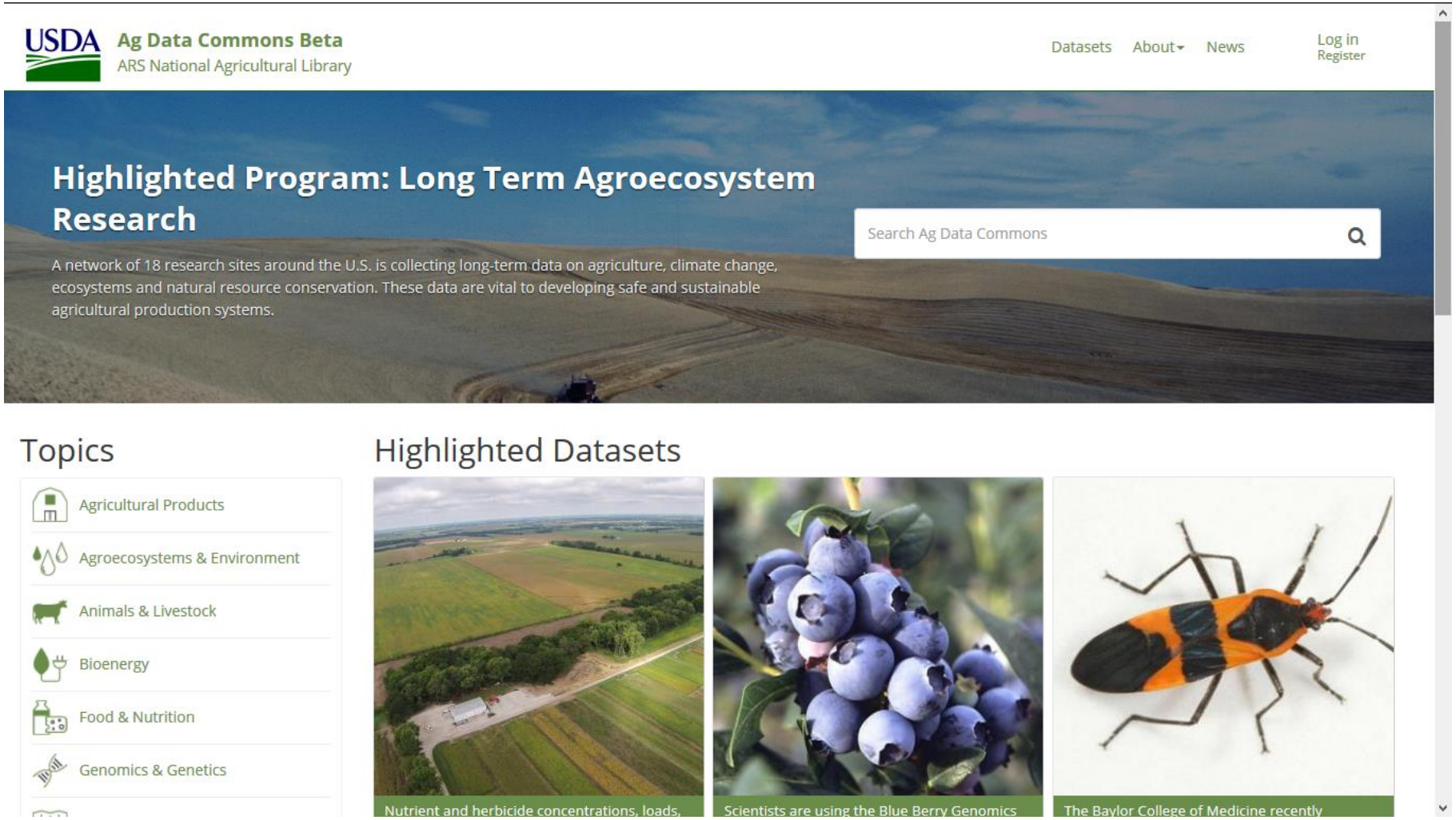
Home | What is GODAN? | News | Partners | Events | Publications | GODAN Summit 2016 | Contact us

What is GODAN?

GODAN supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible and usable to deal with the urgent challenge of ensuring world food security. It is a rapidly growing group, currently with 169 partners from national governments, non-governmental, international and private sector organisations that have committed to a joint Statement of Purpose.

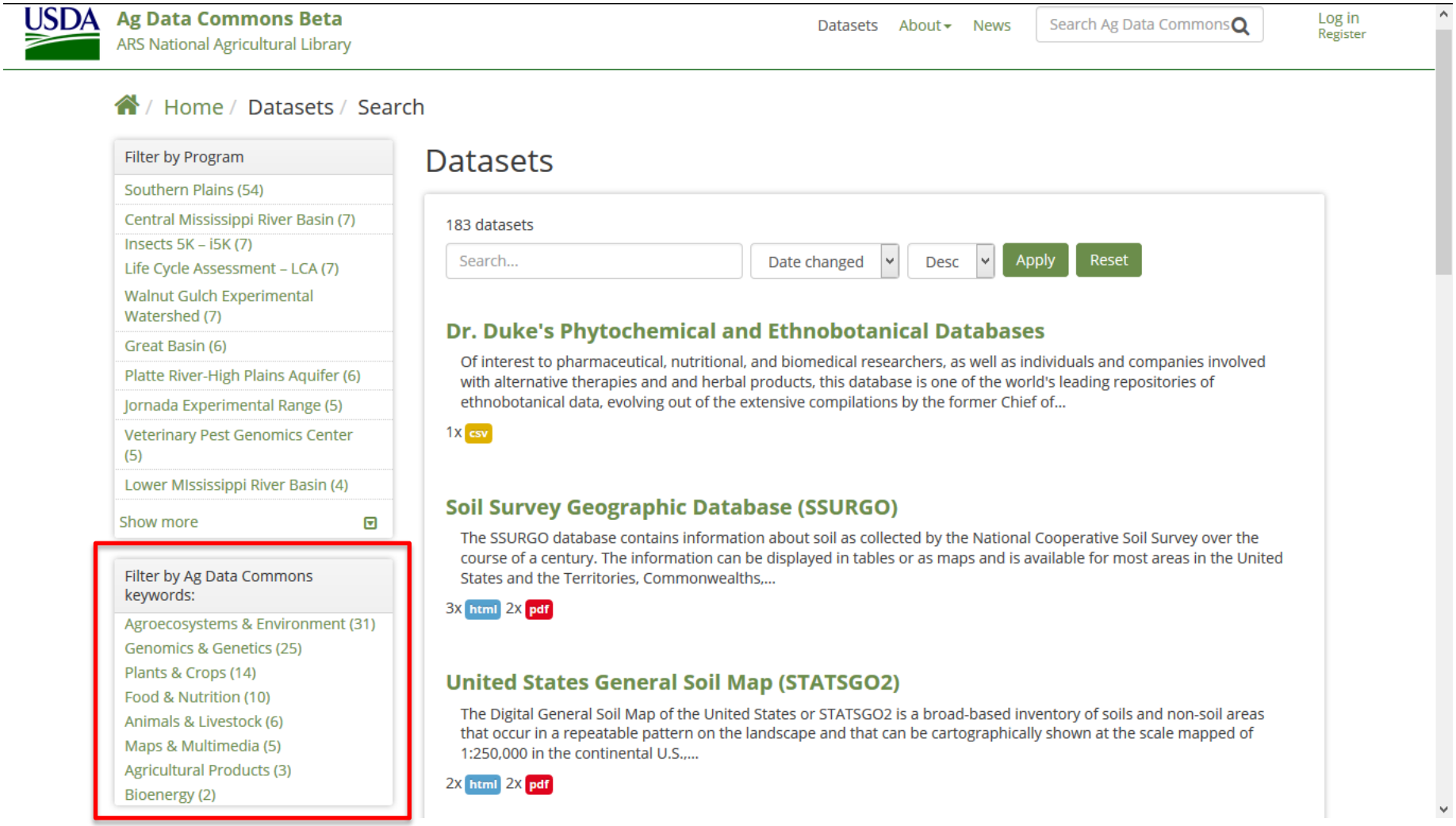
The initiative focuses on building high-level support among governments, policymakers, international organizations and business. GODAN promotes collaboration to harness the growing volume of data generated by new technologies to solve long-standing problems and to benefit farmers and the health of consumers. We encourage collaboration and cooperation between stakeholders in the sector.

NAL – Ag Data Commons



NARDN-HD

NAL – Ag Data Commons



NARDN-HD

Characteristics of Proposed Project

- Emphasis on core sets of data, defined by research community
- Uses ICASA/AgMIP Data Standards for crops (~30 years experience)
- Development of a data dictionary and for livestock core data
- Includes crop, soil, weather, and management details
- Data harmonization based on proven methods developed by AgMIP and demonstrated in a proof of concept workshop in 2015 at the National Agricultural Library
- Demonstrated to work for several different families of crop models
- Approach also allows for storage of additional (non-harmonized) data from experiments in addition to harmonized core data

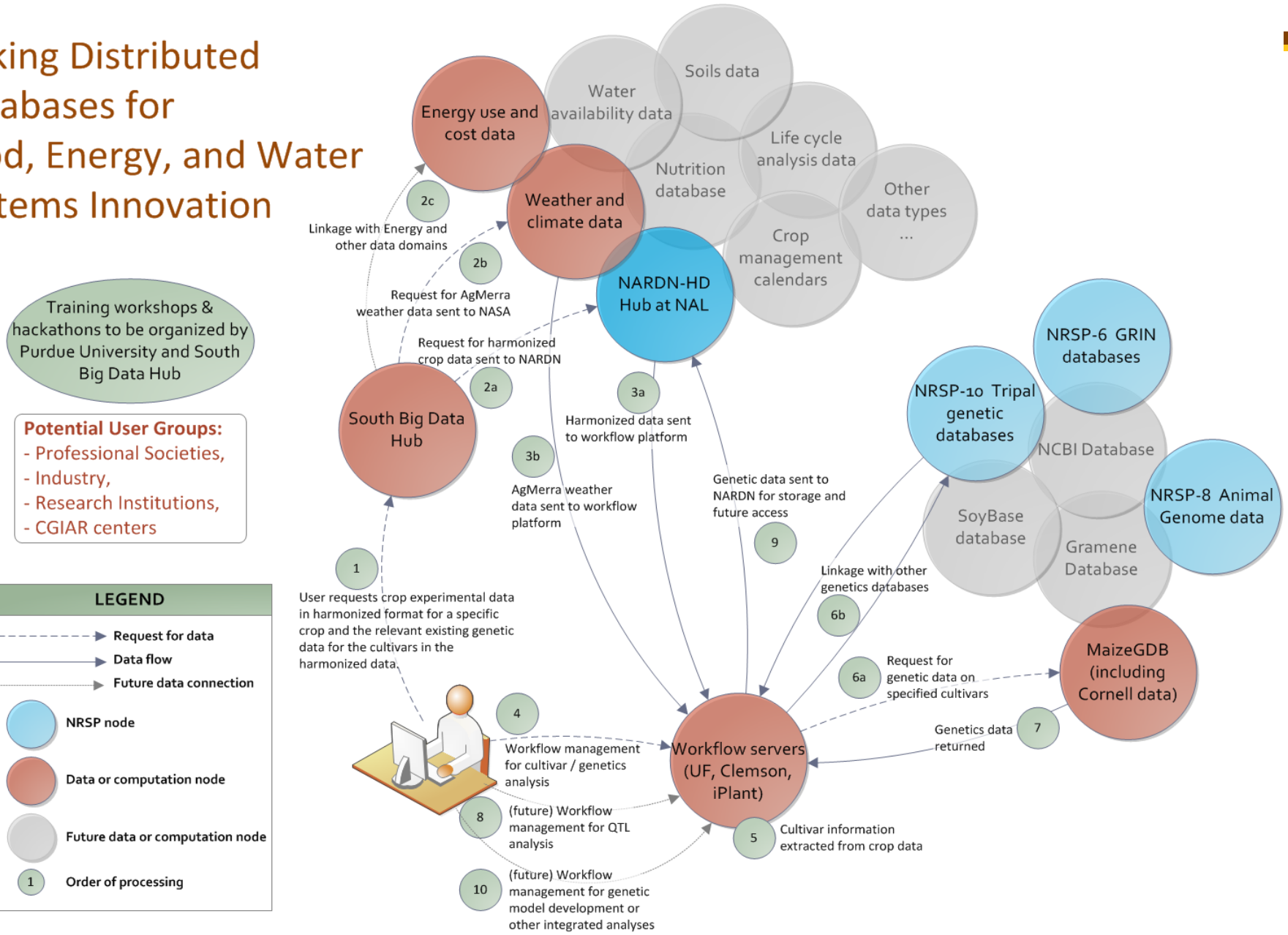
Characteristics of Proposed Project

- Active contributions by researchers, initially in 13 core states included in the proposal
- Open to participation by all states, including all workshops
- ARS endorsement, participation and support for data portal at the National Agricultural Library (letter)
- Multi-state research projects are supportive; letter from S-1032 project (25 states), recent interest by SC-33 project
- Endorsed by international data initiatives and private sector collaborators
- Interest by broader scientific community (e.g., Network of Networks for addressing Food, Energy and Water research issues)

NARDN-HD

Vision of Network of Networks

Linking Distributed Databases for Food, Energy, and Water Systems Innovation



NARDN-HD Components

- Metadata – Description of the datasets available in harmonized format anywhere in the network
- AgMIP common data format (crops) – flexible and extensible
  - Weather
  - Soil
  - Management
  - Crop/soil responses
- Data dictionary – variables and units (upload, access, use)
- Data translators
- Web portal and interface

NARDN-HD: Initial Contributors

1. University of Florida
2. Columbia University
3. Cornell University
4. Iowa State University
5. Kansas State University
6. Michigan State University
7. North Carolina State University
8. Purdue University
9. University of Wisconsin
10. National Agricultural Library
11. USDA-ARS
12. University of Georgia
13. Texas A&M University
14. University of Idaho
15. Washington State University
16. University of California-Davis

Open to all states involved in federally-funded agricultural research

NARDN-HD: Milestones

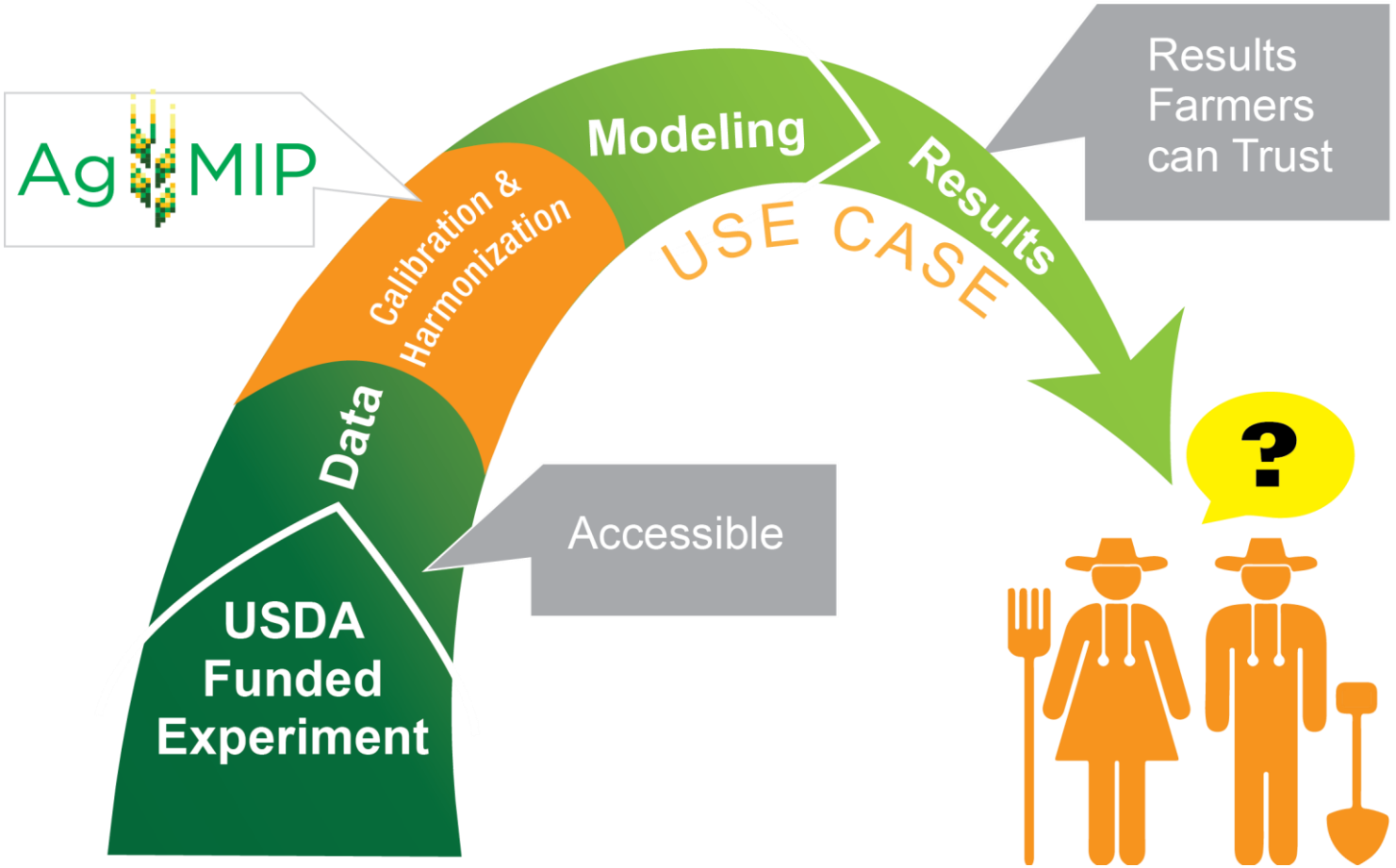
1. Annual workshops, development sprints
2. Submit additional proposals (e.g., NSF)
3. Year 1 – Implement basic structure at NAL
4. Year 1 – Upload first set of crop data
5. Year 2 – Data dictionaries for livestock draft for review, revision
6. Year 2 – Links in place to other databases (i.e., genomics, NSF BD hubs, CGIAR AgTrials, etc.)
7. Year 3 – Translators in use for crop and livestock data; more than 10,000 crop/livestock “treatments”
8. Year 3 – Spinoff research demonstrating value of NARDN-HD
9. Year 5 – More than 50,000 crop/livestock records
10. Year 5 – Global connectivity, more spinoffs
11. Year 5 - Plan implemented for sustaining the NARDN-HD

Opportunities

- Identify, access, and use quantitative data to develop and evaluate agricultural systems models (statistical, dynamic, meta-analysis)
- Perform meta-analyses across space and time
- Better understand genotype, environment, and management interactions

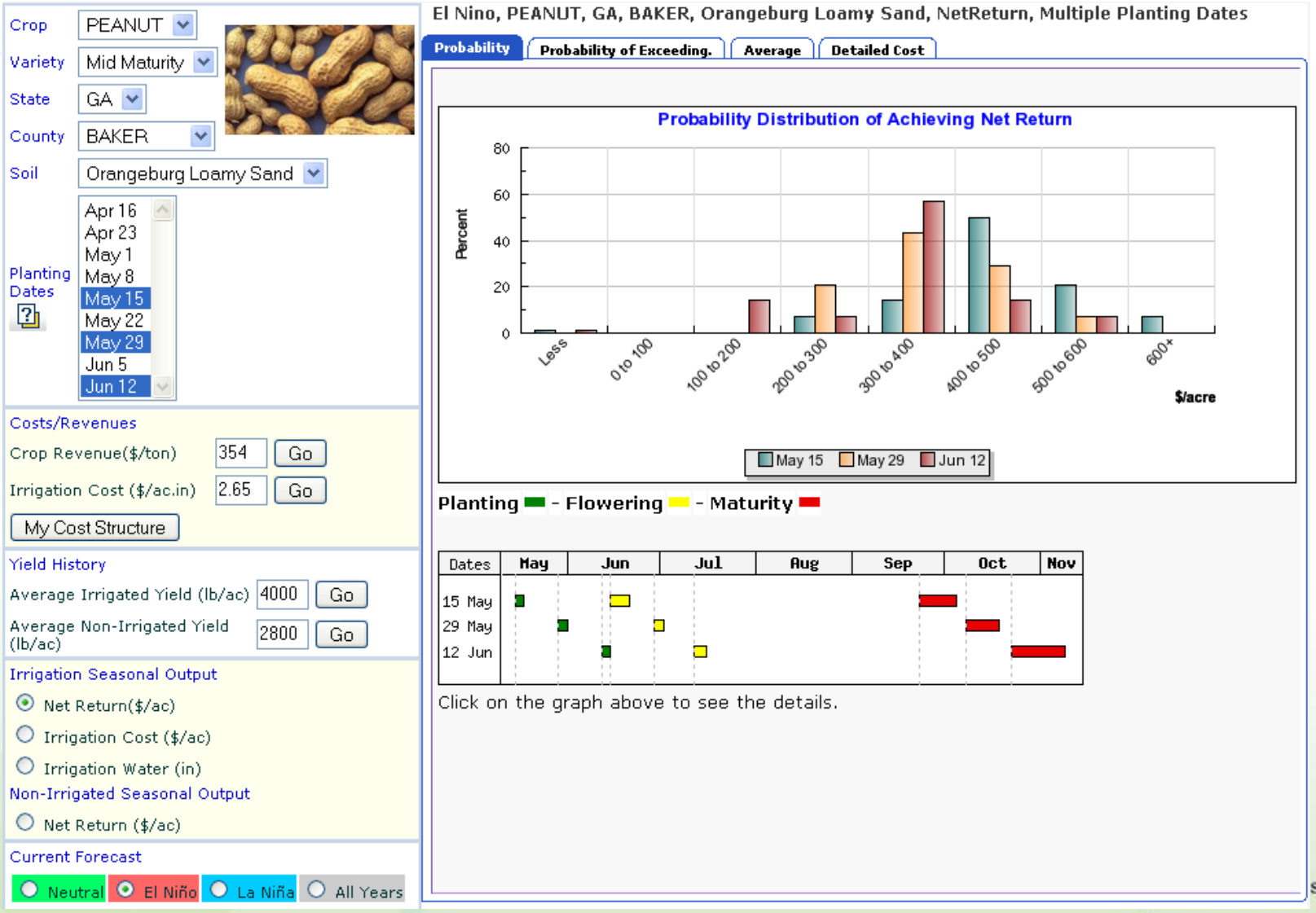
Initial Focus on Field Experiments and Variety Trials; > 50,000 crop-location-growing season records

Relevance to Extension



Crop Simulations: AgroClimate

Extension, Producers and Consultants



sortium

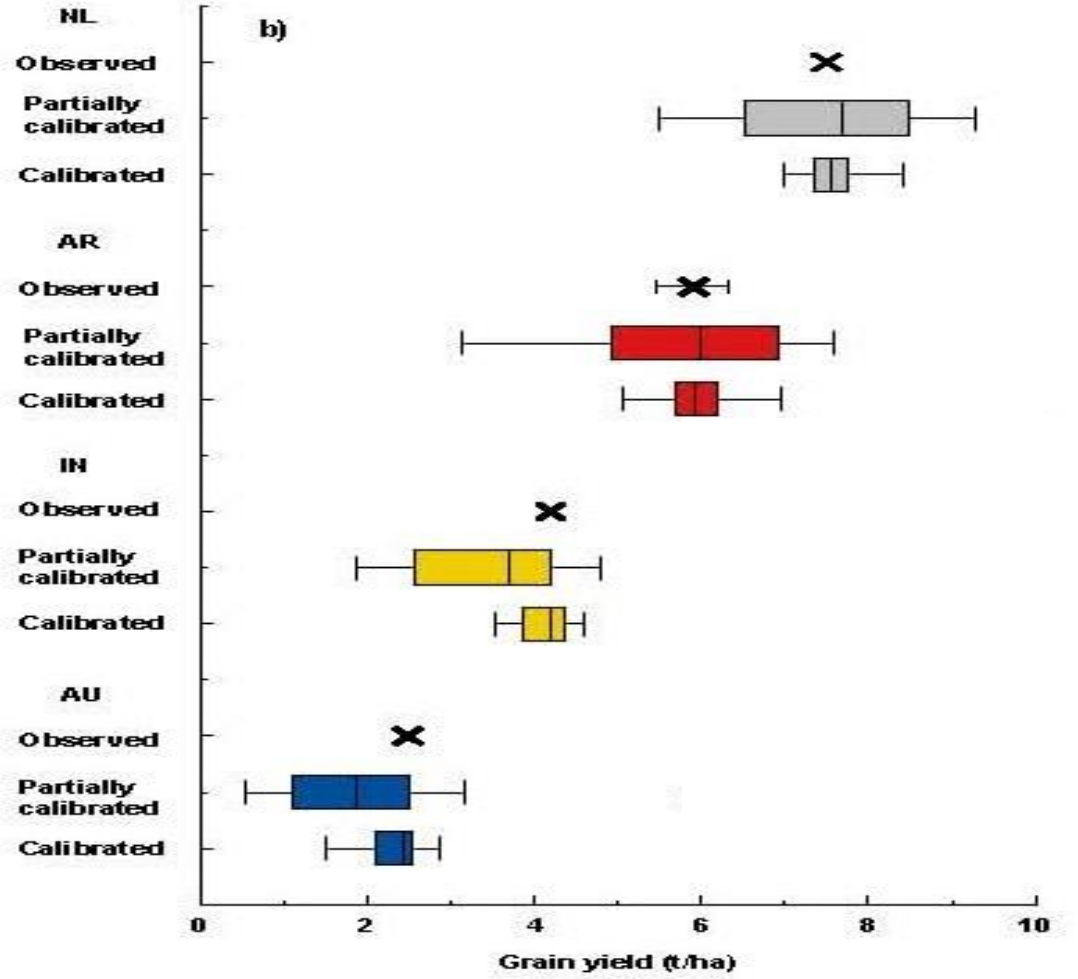
NARDN-HD

Next Generation Models, Data, Knowledge Systems: Use Cases

- Next generation agricultural models and decision support systems must be based on broader data
- Data are needed across environments, management, and genotypes in order to optimize systems for specific socioeconomic, climate, soil conditions
- Transdisciplinary efforts are needed, integrating agronomy, plant pathology, entomology, plant breeding, bioinformatics, socio-economics, policy, and stakeholders
- Data-driven models, data evidence, data for decision support, data for investment decisions, strategic foresight analyses, …
- Integrated farming systems models are needed, with crop, livestock, energy enterprises
- AgMIP has initiatives on next generation models, pest & disease models, economic models, and methodologies
- **Without a strong data foundation, scientific progress will be limited**

Simulations in Different Environments

27 wheat models



Uncertainty of model ensemble results much lower in well calibrated simulations

Asseng et al. 2013
Nature Climate Change

NARDN-HD

Final Thoughts

- NARDN-HD needs to be extended nationally and globally; already connecting with international networks through AgMIP and CGIAR
- Usable data required for coordinated national, regional, and global food security assessments for the US National Climate Assessment and IPCC AR6

Data Harmonization Essential!

Thank You!

Questions?