Council for Agricultural Science and Technology

Kent G. Schescke
Executive Vice President
2019

#CASTreports2019

The Science Source for Food, Agricultural, and Environmental Issues
The CAST Mission

Through Our Network of Experts

We Assemble, Interpret, and Communicate

Credible, Science-based Information

To Policymakers, the Media, and the Public
A world where decision making related to agriculture and natural resources is based on credible information developed through reason, science, and consensus building.

“...what CAST does is very important to mankind.”
~Dr. Norman E. Borlaug
What Does CAST Do?

CAST disseminates science-based information through:

- Print materials
- Online sources
- Social media
- Videos
- Spanish and Chinese translations of select publications

#CASTreports2019
How Does CAST Do This?

With the help of many volunteer contributors:

**50 Board Members** representing scientific societies, companies, nonprofits, and universities

**160+ active task force members** working on CAST reports yet to be released

Volunteer scientific experts as authors and reviewers—**more than 750 volunteers** since 2007

<table>
<thead>
<tr>
<th>Sector</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia</td>
<td>65%</td>
</tr>
<tr>
<td>Government</td>
<td>15%</td>
</tr>
<tr>
<td>Companies</td>
<td>15%</td>
</tr>
<tr>
<td>Nonprofits</td>
<td>5%</td>
</tr>
</tbody>
</table>

#CASTreports2019
Unique Formats for Unique Needs

**Issue Papers.** 8-20 pages; more narrowly focused.

**Special Publications.** 25-150 pages; often follow-on to workshop.

**Task Force Reports.** 40-200 pages; comprehensive overview.

**Commentaries.** 4-8 pages; quick turnaround.

**Ag quickCASTs.** 1-page excerpts from full CAST publications.

The Science Source for Food, Agricultural, and Environmental Issues
Other Forthcoming Publications

• Reducing the Impacts of Agricultural Nutrients on Water Quality across a Changing Landscape [April 15, 2019, rollout]

• GMO Free—The Impact on Consumers, Retailers, Farmers, and the Environment [Summer 2019 rollout]

• Interpreting Agricultural Chemical Residues Measured in Food or Milk [Summer 2019 rollout]
Friday Notes

- Published 48 times annually
- Lead articles on timely ag topics
- Dozens of ag news briefs from 100+ sources with live links to original articles
- An international news section
- Legislative updates from D.C.

#CASTreports2019
Submit an idea!

Do you have an idea for a CAST publication, video, or project you would like to share?

CAST®
Borlaug CAST Communication Award

2019 Winner Announcement
April 16, 2019, at 3:00: USDA Whitten Patio
Keynote Speaker: Dr. Marty Matlock

*Sustaining Prosperity from The Land in the 21st Century*


Marty Matlock (2018)
Sun...Earth...Water...Mankind
In synergy with science and technology
to create a sustainable world
supported by plants and animals

The CAST website has had visitors from every U.S. state, Canada provinces, and 181 countries.

#CASTreports2019
Enabling Open-source Data Networks in Public Agricultural Research

CAST Commentary QTA2019-1

Presented by
Sylvie Brouder, Ph.D.
Purdue University, Department of Agronomy
March 2019
Task Force Members

Authors:

Sylvie Brouder (Chair)
Department of Agronomy, Purdue University
West Lafayette, Indiana

Alison Eagle
Sustainable Agriculture, Ecosystems Program,
Environmental Defense Fund, Raleigh, North Carolina

Naomi K. Fukagawa
Beltsville Human Nutrition Research Center, USDA–ARS
Beltsville, Maryland

John McNamara
Teaching Academy, Washington State University
Pullman (retired)

Seth Murray
Department of Soil and Crop Sciences
Texas A&M University, College Station

Cynthia Parr
National Agricultural Library, USDA–ARS
Beltsville, Maryland

Nicolas Tremblay
Agriculture and Agri-Food Canada
St-Jean-sur-Richelieu, Canada

Reviewers:

Marianne Stowell Bracke
Harriet Cheney Cowles Library, Whitworth University
Spokane, Washington

Paul Fixen
International Plant Nutrition Institute
Brookings, South Dakota (retired)

Jeffrey Volenec
Department of Agronomy, Purdue University,
West Lafayette, Indiana

CAST Liaison:

Drew Lyon
Extension and Research, Washington State University
Pullman

#CASTreports2019
What motivates this commentary?

Grand Challenge Questions

Next generation problem solving in agriculture:

- “Big” science
- Transdisciplinary research linking disciplines and DATA

Major barrier to making better agricultural decisions:

- Lack of data sharing and data accessibility

Publically funded science belongs to the public and should be available for their use
“Enabling Open-source Data Networks in Public Agricultural Research”

Goal
Advance the conversation among agricultural science partners to create a system conducive to data sharing and the team science to address grand-challenge questions in food systems.

Purpose
Document the need for and anticipated benefits of developing data standards, incentivizing data sharing and building data-sharing infrastructure to meet the varied needs of agricultural researchers.
Topics for today...

• Historical perspectives on data and the process of conducting agricultural research
  • Small science, the small-science environment, outcomes, and limitations
  • Why we currently don’t share data
• Vision for the new data ecosystem
  • FAIR data, repositories, “knowledgebases”
• Four immediate imperatives
• Four critical strategies to facilitating data sharing
• Partnering for success
• The business model
  • Open data is not free ~ suggestions for who pays and how

What is “small” vs “big” science?

Hypothesis-driven research, led by single researcher or small team, generating and analyzing their own results
History Lesson: Why did we do “small” agricultural research?

**Research tools and technologies**
- **limited data size and accessibility**
  - Rudimentary data collection and data-management tools
  - More limited capabilities of field and laboratory analytical tools
  - No or limited access to computational power
  - Data storage an individual decision and enterprise

**Lots to be gained and was gained from “incremental” science**
- Gregor Mendel (1800s)
- Ronald Fisher (1900s)
- Gebisa Ejeta (now) ...
How have we pursued agricultural research questions in the past?

**Environmental Outcomes:**

- Article data not “open” (subscription paywalls)
- No explicit linkages among knowledge fragments
- Research not transdisciplinary
- Lots of inaccessible and, often, lost data
- Data accessed via journals partial and potentially biased

The “Small” Science Environment
Policy and recommendations need to reflect all results; in journals, the **null result is an endangered species!**

- Review of 4,600 papers – all disciplines – 242 papers in agriculture
- Evaluated them for positive support of a “tested hypothesis”
  - Found 22% increase from 1990 to 2007
  - Asia > U.S. > Europe
- Hypothesize that:
  - Research becoming less pioneering and/or
  - Objectivity in research/publication is decreasing...

Negative results are disappearing from most disciplines and countries (Fanelli et al. 2012)
Achieving desired outcomes to complex problems with science: Single panacea vs. articulating strategies toward solutions...

Small science can’t address system tensions and trade-offs to weight a *plurality of possible outcomes for multiple objectives*

“Abandon reductive approaches that imply one solution, and instead acknowledge the necessity of a non-reductionist approach that is integrative over different levels of detail.”

A complex systems approach to address world challenges in food and agriculture (van Mil et al. 2014)
Other reasons to share: Increasing public trust in and use of science is founded on open data

Sharing of research materials and raw data

Transparent disclosure of methods and results

Collaboration to increase “power” and replicate findings

Enhanced reliability of research results

Increased public trust in and use of science

Make “big” data out of “small”
Why is so much data not shared and/or lost?

The “scoop” and other past perceptions of ownership...

Multiple “vocabularies” among disciplines and domains

---

The response of the Contrarian Curmudgeon:

“You can have my data when you pry it out of my cold, dead hard drive.”
Why is so much data not shared and/or lost?

Lack of proper, prior planning and low-barrier desktop tools ~ the effect of time on the “knowledge value” of data...

Knowledge Value of Data

Phase 1:
Data collection during exp. / Analysis for thesis or manuscript prep.

Phase 2:
On to the next project...

Phase 3:
Time or circumstance create distance from the topic

None
Proximate
Remote

Time / distance / both

Reality ~ data sharing is hard...

• It takes time and money
• Data literacy has not been part of undergraduate and/or graduate curricula
• Excel is low-barrier but it is inadequate to the task
• Once data are prepared for sharing, what do you do with it…?

Modified from W. Michener, DataONE
FAIR data to facilitate data sharing and extend data lifecycles

Typical agricultural data lifecycle...

Agricultural data should be FAIR.
FAIR data are/can be...

**Findable**: described with a digital object identifier and rich metadata indexed in a searchable resource

**Accessible**: retrieved using a standardized communication protocol (free, open, and universally implementable)

**Interoperable**: represented with a formal, shared, and broadly applicable language with FAIR vocabularies

**Re-usable**: richly described by a plurality of attributes (clear provenance, and usage license, and meets domain standards)

(Adapted from Wilkinson et al. 2016)
Key attributes:

- Data collection *a priori* anticipating reuse
- Article data is published; all project data are FAIR
- “Knowledgebases” with expert services enhance data collections and reuse
- Dedicated experts collaborate with teams and stakeholders for high-value data products (“fusions”)
- Stakeholders access knowledge and data...
Facilitating data sharing to fully realize open access to public agricultural research

**Imperatives**

- Development and implementation of “best practices” for data management in all federally funded projects

**Data and Metadata Standards**

Necessary for getting from here to there...
Facilitating data sharing to fully realize open access to public agricultural research

**Imperatives**

- Development and implementation of “best practices” in all federally funded projects
- Incentives and mechanisms for making “grey and dark” data available

Fertilizer Industry’s 4R Research Fund and Repository

**4. Funding Schedule**

The 4R Fund requires 10% of the funding be withheld until this is submitted.

<table>
<thead>
<tr>
<th>Date</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>year 1</td>
<td>$113,724</td>
</tr>
<tr>
<td>year 2</td>
<td>$106,903</td>
</tr>
<tr>
<td>year 3</td>
<td>$76,252</td>
</tr>
<tr>
<td>upon completion</td>
<td>$32,986</td>
</tr>
</tbody>
</table>

(remaining 10%)
Facilitating data sharing to fully realize open access to public agricultural research

**Imperatives**

• Development and implementation of “best practices” in all federally funded projects

• Incentives and mechanisms for making “grey and dark” data available

• **Coordination among existing and emerging data initiatives, networks, and repositories**

**Examples of initiatives, networks, repositories ~ illustrative not exhaustive!**

<table>
<thead>
<tr>
<th>Alliances, Coalitions, Networks</th>
<th>Repositories and Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Research Data Alliance</td>
<td>DataONE</td>
</tr>
<tr>
<td></td>
<td>Ag Data Commons (ARS NAL)</td>
</tr>
<tr>
<td>The Ag Data Coalition</td>
<td>Cy Verse</td>
</tr>
<tr>
<td></td>
<td>Purdue Univ. Research Rep (PURR)</td>
</tr>
<tr>
<td>AgBioData</td>
<td>GODAN</td>
</tr>
<tr>
<td></td>
<td>Dryad</td>
</tr>
<tr>
<td></td>
<td>Maize DB</td>
</tr>
</tbody>
</table>

Sheer number poses a challenge to a coordinated landscape; each faces **sustainability** issues...
Facilitating data sharing to fully realize open access to public agricultural research

**Imperatives**

- Development and implementation of “best practices” in all federally funded projects
- Incentives and mechanisms for making “grey and dark” data available
- Coordination among existing and emerging data initiatives, networks, and repositories
- **Dedicated and sustainable infrastructure** – hardware, software, and human resources – to curate, preserve, add value...

**Critical need: A knowledgebase at the heart of the data ecosystem**
Strategies

Data to support ecosystem service valuation, food supply chain sustainability metrics, and user “Apps”

Curriculum for “digital natives”

Rescuing long-term data records documenting farm management impacts on water quality (N, P)

Moving well beyond an agronomist simply asking an economist to play a cursory role or a data scientist to help with analytics or code
Commitment to collaborative, iterative analysis of successes and failures in design, implementation, and utility...
Open access ≠ cost free: Who pays for the ecosystem?

Open data infrastructure = public good

Core institutional support → 18-year survival / database support via short-duration research grants (current model) wasteful and inefficient

Longevity of biological databases (Attwood et al. 2015)

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alive</td>
<td>53</td>
<td>16.3%</td>
</tr>
<tr>
<td>Alive - rebranded</td>
<td>23</td>
<td>7.0%</td>
</tr>
<tr>
<td>Archived</td>
<td>47</td>
<td>14.4%</td>
</tr>
<tr>
<td>Dead</td>
<td>203</td>
<td>62.3%</td>
</tr>
<tr>
<td>Total</td>
<td>326</td>
<td>100%</td>
</tr>
</tbody>
</table>
Who pays? User?

“who’d pay for YOUR data?!”

Subscription fee = open access?

Sustainable funding for biocuration: The Arabidopsis Information Resource (TAIR) as a case study of a subscription-based funding model (Reiser et al. 2016)

“Freemium” → data access is free but premium payments for additional services
12 funding models for data (knowledgebases) by origin of revenue

Funding knowledgebases: Towards a sustainable funding model for the UniProt use case (Gabella et al. 2017/2018)

Which are open access compliant / equitable? #1 – 4

Which are stable? #1, 2, maybe 3...

“Razors and blades” = We give you the razor, you buy blades
Final thoughts

✓ Infrastructure: Most cost-effective, robust solutions may involve a mix of the proven with the innovative

✓ Leadership and oversight: USDA Research and Economics Office
  ➢ Office of the Chief Scientist in partnership with the Office of the Chief Information Officer
  ➢ Why? Stewardship of public research data is a natural extension of their historic roles and responsibilities.

✓ 2018 Farm Bill opportunities?
  ➢ Creation of the Agriculture Advanced Research and Development Authority (AgARDA)
  ➢ Full appropriation of authorized funds will position AgARDA to address the data infrastructure needed for agricultural research to “make big data out of small.”

✓ Getting started in D.C with AgARDA leading a partnership delivering benefits to all stakeholders in the agriculture data value chain?
  ➢ Convenings led by OCS and OCIO: Stakeholders in public listening sessions, USDA leadership in high-level meetings → create plan, operationalize, deploy
Questions/Discussion

Sylvie Brouder
sbrouder@purdue.edu

This CAST Commentary is available at www.cast-science.org/publications.

#CASTreports2019