Smarter Agriculture: A pathway to evidence-based policy and management recommendations for agriculture

Sylvie M. Brouder
Agronomy Dept.
Purdue Univ.
12/4/2013
Dr. Plaut: What is the research case for Big Data?

• Turn information into insights
• Connects and empowers people
• Evidence based decision making
• Can develop integrated solutions to complex problems
• Federal mandate from the OSTP (Office of Science, Technology and Policy) that all federal agencies that award grants must make data available
• Cost Efficiencies
“Smarter Agriculture” ~ What do we need (to have, to be doing, etc.) to use data better in agricultural policy and management (research too)?

Better data skill sets (Data Information Literacy curriculum)

Better framework for organizing & systematically reviewing data *as it accumulates*

Structured analytical framework

Data skill sets / DIL

Data repositories

Better place to store data

Purdue sees the opportunities and is growing programs in all 3 areas...
Why are data not reused?

• **Not useful?** Question has changed... Hmmm: Yes & No

• **Not accessible?** Poor data hygiene...
  
  – Diekmann interviews (J. Ag. & Food Info., 2012):

    “The researcher wanted to reanalyze data from another figure and I couldn’t find it. And I couldn’t; I lost it. It was done on an old computer system and the technician who did [it, had] moved on and I wasn’t able to find it.”

    “We have had a lot of problems in the past of losing data, or just misplacing it. And then we have to backtrack it and that’s taken literally days or weeks to find where this data was stored. So it has been a real problem for us.”
Why are data not reused?

• **Too much work?** Lack of data workflow tools...
  – Diekmann interviews (J. Ag. & Food Info., 2012):  
    “[Another group of scientists and I] were talking about, can we get our data and pull it together? They wanted that data, [but] *it’s the annotation that’s really the hard part* [for] them [to be] able to make sense of it. I would be happy to give [out the data], but [then] *I have to explain whatever I did.*”

• **Too expensive?** > 80% of scientists surveyed in 2010 indicated that they did not have resources to make their data open access (Science. Feb. 2011)

*Question of Money, Motivation, and Mechanics...*
Taking a peek at data caretaking in AGRY... K
Team Fellow (PhD student supported by Mosaic and PCS)
Today, I can tell you what this spreadsheet means but you can’t understand all of it on your own...

What is this???

Tomorrow, we may both be in the dark...
Precarious Nature of Data Lifecycle: Scientifically proven that my ability to understand and find these data will erode extremely rapidly!

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

Death...

Change jobs / professional focus / retire

Time / distance
Pressing technological challenges to informatics for all agronomic efforts concern data workflow...

- **Data dispersion**
  - Take advantage of small datasets collected by many researchers (not everything is “BIG”)

- **Data heterogeneity**
  - Varied protocols reflecting local culture & variation in 1° purpose

- **Data provenance**
  - Need to track data through multi-step process of aggregation, modeling, analysis
How hard/expensive can it be???? (Why is posting data on a website NOT enough...)

• Lack of longevity
  – I depart and my webpage may or may not be in maintained (the original source is responsible for preservation and curation).
  – Web pages are updated and links are not
  – Lack of forward formatting (who remembers Lotus 1,2,3?)
  – Data provenance disappears (where DID this number come from)

• Lack of rigor, standardization and process
  – You annotate your data in a way YOU think describes it adequately but no one else reviews it... not standardized because you are not following a protocol (we love to make up our own terms ~ part of showing novelty of endeavor)
  – Not “tagged” for searching
  – Does not have a Digital Object Identifier (DOI: unique & persistent identifier)
Purdue University Research Repository: What libraries are to books, PURR is to data (plus so much more!)

Start Your Research Project

Create a Data Management Plan
Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. Get Started

Upload Research Data to Your Project
Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. Create a Project

Publish your Dataset
Package, describe, and publish your dataset with a Datacite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long-term. See Published Datasets
Very first PURR data publication: Volenec
Alfalfa data looks like Navarrete K data (mostly)...

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Rep</th>
<th>Plot No.</th>
<th>Treatment K, kg/ha</th>
<th>P, kg/ha</th>
<th>yield (kg/ha)</th>
<th>P (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>411</td>
<td>19</td>
<td>400</td>
<td>50</td>
<td>24.5</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>412</td>
<td>3</td>
<td>0</td>
<td>50</td>
<td>23.1</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>413</td>
<td>2</td>
<td>0</td>
<td>50</td>
<td>30.2</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>414</td>
<td>2</td>
<td>0</td>
<td>25</td>
<td>18.3</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>415</td>
<td>5</td>
<td>100</td>
<td>25</td>
<td>19.1</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>416</td>
<td>5</td>
<td>200</td>
<td>0</td>
<td>5.2</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>417</td>
<td>5</td>
<td>100</td>
<td>0</td>
<td>10.0</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>418</td>
<td>10</td>
<td>200</td>
<td>25</td>
<td>19.5</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>419</td>
<td>4</td>
<td>0</td>
<td>75</td>
<td>35.1</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>420</td>
<td>11</td>
<td>200</td>
<td>50</td>
<td>26.3</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>101</td>
<td>8</td>
<td>100</td>
<td>75</td>
<td>5566.0457</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>102</td>
<td>9</td>
<td>400</td>
<td>50</td>
<td>5903.9126</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>103</td>
<td>3</td>
<td>0</td>
<td>50</td>
<td>5933.5056</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>104</td>
<td>13</td>
<td>300</td>
<td>0</td>
<td>4455.2046</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>105</td>
<td>17</td>
<td>400</td>
<td>0</td>
<td>5280.7465</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>106</td>
<td>20</td>
<td>400</td>
<td>75</td>
<td>5990.6994</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>107</td>
<td>12</td>
<td>200</td>
<td>75</td>
<td>5673.7468</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>108</td>
<td>9</td>
<td>200</td>
<td>0</td>
<td>5696.6962</td>
</tr>
<tr>
<td>2000</td>
<td>5</td>
<td>1</td>
<td>109</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6102.5182</td>
</tr>
</tbody>
</table>

K (mg/kg)
You can “google” data published in PURR...

https://purr.purdue.edu/publications/1004/about

Abstract
This is a master file of merged excel files with data from several studies on the influence of phosphorus and potassium nutrition on alfalfa. It includes numeric data such as yield, plant mass, plant counts, and tissue concentration of various nutrients. Conventional wisdom states that plants that are poorly fertilized and do not have adequate phosphorus and especially low potassium will become winterkilled. This study showed that plants did not necessarily die in winter with low potassium stress.


Location/Latitude and Longitude: Throckmorton Purdue Agricultural Center located 15 km south of West Lafayette, IN (40°N and 87°W)

Cite this work
Researchers should cite this work as follows:

Tags
- Alfalfa
- Forage composition
- Forage legumes
- Forage quality
- Medicago sativa
- Nutritive value
- Phosphorus fertilizer
- Potassium fertilizer
- Soil fertility
- Winter hardness
- Yield components

NAL terms; important unique terms (Grant #)

Version 1.0 Notes
Citations:
The workflow is predetermined when publishing ~ you are prompted to be comprehensive in the info you provide ~ PU Lib. Information Specialists review it prior to publication...

Phosphorus and Potassium Influence on Alfalfa Nutrition

By Jeffrey J Volenec

Data from several studies on the influence of phosphorus and potassium nutrition on alfalfa. It includes numeric data such as yield, plant mass, plant counts, and tissue concentration of various nutrients

Listed in Datasets

Supporting Docs

DS2.pdf (PDF | 19.98 Kb)
Supporting documents accessible with datasets.... Alfalfa P/K study

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Calendar year when samples were obtained (1997 to 2004)</td>
</tr>
<tr>
<td>B</td>
<td>Month within calendar year when samples were obtained (1 to 12)</td>
</tr>
<tr>
<td>C</td>
<td>Field replicate (1 to 4)</td>
</tr>
<tr>
<td>D</td>
<td>Plot number (101 to 420)</td>
</tr>
<tr>
<td>E</td>
<td>Treatment number (1 to 20)</td>
</tr>
</tbody>
</table>

Summary:
This is a master file of merged excel files with data from several studies on the influence of phosphorus and potassium nutrition on alfalfa. It includes numeric data such as yield, plant mass, plant counts, and tissue concentration of various nutrients. Conventional wisdom states that plants that are poorly fertilized and do not have adequate phosphorus and especially low potassium will become winterkill. This study showed that plants did not necessarily die in winter with low potassium stress.


Location/Latitude and Longitude: Throckmorton Purdue-Agricultural Center located 15 km south of West Lafayette, IN (40°N and 87°W)

Background:
In April 1997, a 1.4-ha site was seeded to Pioneer Brand ‘5454’ alfalfa. This site was selected for study because soil tests indicated low concentrations of extractable P (9 to 15 mg kg⁻¹ Bray P1) and low to
“Smarter Agriculture” ~ **Evidence & Eminence** in recommendation and policy development

Better data skill sets (Data Information Literacy curriculum)

Better framework for organizing & systematically reviewing data as it accumulates

Better place to store data

**Structured analytical framework**

**Data skill sets / DIL**

**Data repositories**

*Purdue sees fantastic opportunities for Extension...*
Ag. Extension Model for Knowledge Creation/Translation Pathway

Basic Research Results

Applied Research Results / Synthesis to Recommendation

Management Practice

Policy (Guidelines) Assistance or Regulatory Instruments
“Eminence” Model: The Extension Specialist’s Purview to…

- Transform data / research into practical knowledge
- Adapt research results to the farm / farmer environment / context
Current “conventional” way of synthesizing applied & basic research pieces into “big picture” results (e.g. fertilizer recommendations, policy) ~ selection bias favors “peers”

Applied Res. Philosophy:
- Data owned by PI
- Don’t share raw data
- Peer review pubs not essential

Reliance on “best professional judgment” (qualitative review)
What’s the problem with a strictly or primarily “eminence” –based framework?

• We are just not using all the data we could or should be using & the amount of data that we could use is skyrocketing (The “Moneyball” Extension Allegory)
  – Extension Specialists are human beings and prone to biases and lapses in judgment at a rate similar to the general population
  – We tend to rely on data that we are familiar to a greater extent than data more distant from us.
  – Training in scientific theory and practice may help overcome this.

• “Best professional judgment” ought to be considered in this light to ensure equal parts eminence and evidence (Note: as experts we are not alone in this problem)

If you’ve seen it, you expect to see it, you are prone to diagnose it... the power of the “change up” pitch!
What’s the problem cont.

• Can’t find the original data/publications to reanalyze/augment → unnecessarily have to redo portions of research ($$$")
• Guidelines and recommendations not specific enough to satisfy most farmers
• Policy instruments unsatisfying / contentious:
  – Assistance programs don’t seem to be achieving desired outcome (e.g. CEAP); costly
  – Regulations perceived as burdensome, inappropriately targeted, inefffectual, suppressing of profitability and economic growth
We need to use data... differently, more comprehensively, more quantitatively...
Medical Data and the Grassroots Effort to Make it Available for Evidence-based, Clinical Use

Kay Dickersin, MA, PhD

Smarter Agriculture: A Dialogue on Critical Data for Agriculture
Potomac, Maryland
October 10, 2013
Dickersin: Reviews essential in health

• Research evidence generally available in short published papers. Researchers try to publish a lot of them

• Literature is large and growing (eg, 20,000 biomedical journals)

• Literature is not organized

• Reviews of primary research necessary for coping with information overload
Dickersin: Knowledge translation: From clinical research to practice decisions

Evidence generation

Clinical trials, observational studies

Evidence Synthesis (systematic reviews)

Cochrane Collaboration, others

Clinical policy (guidelines)

Professional Societies, others

Application of policy: Evidence, Clinician expertise, Patient values

Evidence-based healthcare

Knowledge translation
Dickersin: What is a systematic review?

• A review of existing knowledge that uses explicit, scientific methods:
  – Structured and transparent process
  – Comprehensive search for relevant articles
  – Explicit methods of appraisal and synthesis
• Summarizes methods and results of similar but separate studies
• May or may not combine results quantitatively (meta-analysis)
Explicit framework for rigorous SR Vs. reviews conducted as low cost, ad hoc, un-structured, grad student level, endeavor...
Dickersin: A “cumulative meta-analysis” showed us how important it is that we synthesize what we know in an ongoing fashion.

Antman et al JAMA 1994

Allows updating!!!
Dickersin: Who is doing systematic reviews?

- Selected topic areas
  - Clinical interventions
  - Animal studies
  - Assessment of risk
  - Toxicities
  - Methodologists
  - Education
  - Social welfare
  - Crime and justice
  - International development

- Cochrane Collaboration

- Groups interested in policy (professional societies, governments, payers)
  - Dept. Education, EPA, health insurance

- Businesses: Hayes, ECRI (contracting to pharma)
Closer to home: Use of science in conservation policy...

Roberts, Stewart, Pullen (2006)~

Asked: "Are review articles a reliable source of evidence to support conservation and environmental management?"

Approach: A comparison to medicine

Answer: Ecological reviews more prone to bias, lacking in details; less likely to address study relevance & quality & to quantitatively synthesize evidence
Collaboration for Environmental Evidence now using this framework to improve use of science in conservation policy
More than Meta-Analysis: Systematic Review (SR) methodology widespread / standard for accessing, appraising and synthesizing scientific information...

SR process helps ID types of bias in independent studies and give pragmatic quality ranking...

Types of bias:
- Selection
- Performance
- Attrition
- Measurement

### Field
<table>
<thead>
<tr>
<th>Field</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline appropriate / representative</td>
<td>Category (Yes, No, Uncertain) &amp; comment on rationale</td>
</tr>
<tr>
<td>Management Intervention relevance</td>
<td>Category (High, Low, Uncertain) &amp; comment on rationale (e.g. if farmer survey, potential confounding factors are distributed across baseline &amp; intervention treatments (improved seed, access to inputs, labor, etc.))</td>
</tr>
<tr>
<td>Experiment execution concerns</td>
<td>Descriptive: problems with design relative to question (bias), methods concerns (publication bias if a review), externalities, data loss, attrition</td>
</tr>
<tr>
<td>Data Extraction concerns</td>
<td><em>Not all pieces of data appropriately presented; data in the wrong format; no variance measure</em></td>
</tr>
<tr>
<td>Level of record/data review</td>
<td>Categorical: Peer-reviewed journal (impact factor), other peer-reviewed format, agency report, unpublished data, etc.</td>
</tr>
<tr>
<td>Overall subjective quality assessment</td>
<td>Category (High, Medium, Low, Uncertain) with respect only to the a priori question</td>
</tr>
</tbody>
</table>
**Extension’s Future**: A decade of reflection & two “shoulds” for ensuring we are the/an unbiased source of knowledge for mgmt practice and policy in a data-rich world...

Repositories: Make data (Vs synthesized results) available for additional/future research

- Requires:
  - Infrastructure
  - Change in culture
  - Better data hygiene

Meta-analysis: Make systematic reviews w/ quantitative methods a “tool of the trade” for tomorrow’s Extension Specialist

- Requires:
  - Infrastructure
  - New knowledge
  - Better quantitative skills

Goal: Using disparate data – in aggregate – to strengthen “inference space”
Dickersin: Knowledge translation: From clinical research to practice decisions

Evidence generation → Evidence Synthesis (systematic reviews) → Clinical policy (guidelines) → Application of policy: Evidence-based healthcare

Clinical trials, observational studies
Cochrane Collaboration, others
Professional Societies, others
Clinician expertise
Patient values

Knowledge translation
Evidence-Based Healthcare

“The integration of best research evidence with clinical expertise and patient values”

Evidence-Based Agriculture

“The integration of best research evidence with management expertise and stakeholder priorities?”


T. Scott Murrell, IPNI
Who has data w/ potential for evidenced-based ag. in the clinical setting... e.g. for a fertilizer rec. or for policy on conservation structure installation/cost share, etc.?

New Concept: “Living Recommendations”

Intensive research (fewer)
- Long-term exp. w/ regionally relevant benchmark trts for “Mitigation” mgmts
- Evolved agro-ecosystems w/ data records
- Intensive C/N/H₂O measures

Replic. / Demo. exp (more)
- Fully-replicated, experiments
- Newly est. trts. of imp. mgmt. variants.
- High risk &/or emerging tech.
- Less intensive routine measures + targeted sampling campaigns

Commercial Sites (many)
- Farmer Practice vs. improved mgts. (EI)
- “Adaptive Mgmt.” approach
- Limited # of trts. & measures
- Focus on trts. w/ near-term, measurable impacts
“Smarter Agriculture” ~ New Curriculum for DIL

Better data skill sets (Data Information Literacy curriculum)

Better framework for organizing & systematically reviewing data as it accumulates

Structured analytical framework

Data repositories

Data skill sets / DIL

Better place to store data

Purdue sees fantastic opportunities for education mission...
Farmer participation in “open access” data & the data value chain requires…

- Data Modification: Scholarly data not directly useful to farmer – data needs to be subsetted/distilled in the context of important questions.
- Access to more environmental data (weather & climate specified)
- Data “translation” tools for multiple users
- Multiple forms of delivery to span synthesis for quick digestion (“does it work?”) to more nuanced analysis for iterative practice improvement (“adaptive mgmt.”)
- Protocols & minimum datasets for on-farm research; farmer input must be sought & protocols should not be arbitrarily inflexible
- (Privacy / security policy – assumed to be taken care of?)
Blending different ag data streams at different ed. levels requires new skills & DIL curricula (“Library Sciences should be solicited to educate all…”)

Future farmer or ag. industry employee (BS level)

- Everyone needs environmental info. mgmt that teaches how data are produced/used (“data in my life”)
- Array of educational trajectories are needed from most basic level to specific endpts.
- Future farm managers need data skills in context of business mgmt & systems analyses
- Be able to understand data from outside their degree & be able to ascertain data quality

Future consultant, CCA, policy maker, Agent, Ext. Specialist (MS, PhD level)

- Understand exp. design, statistics & probability (risk)
- Understand geospatial data
- Curricula should use open-source software & “workforce-available” statistical tools
- Be able to translate science into lay language w/ context
- CCA: Certificate in Ext. Prgm should cover 12 data competencies
- Capstone data experience
- Ext. Spec. competent in Systematic Reviews; data mgmt plans / repositories part of degree