

Data management approach to multidisciplinary agricultural research and syntheses

Daryl E. Herzmann, Lori J. Abendroth, and Landon D Bunderson

Agricultural scientists are increasingly asked to address challenges related to natural resource stewardship, agricultural productivity, and environmental protection while simultaneously being mindful of the impact and risks associated with climate change (ASA CSSA SSSA 2011; Hatfield et al. 2011; OSTP CAST 2012). Measuring and predicting the effects of climate on agricultural systems adds a layer of complexity that is challenging using traditional data management methods; these methods also limit the potential for full data discovery and innovation (Overpeck et al. 2011; Wolkovich et al. 2012). To properly address challenges, access to multidisciplinary data spanning environments, timescales, treatments, and management is necessary (White and van Evert 2008; Reichman et al. 2011; Eigenbrode et al. 2014). Disciplinary scientists, data scientists, and data managers need to increasingly work in a collaborative manner in this data-rich era. While scientists generally desire to share data, time constraints, limited funding, a lackluster reward system, and reuse concerns are cited as barriers (Michener et al. 2011; Tenopir et al. 2011; Marx 2012; Wolkovich et al. 2012). A concerted and well-executed approach is necessary to overcome these barriers and move toward transformative science.

The Climate and Corn-based Cropping Systems Coordinated Agricultural Project (CSCAP), referred to as “Sustainable Corn,” was funded by USDA National Institute of Food and Agriculture (NIFA) and is the largest corn (*Zea mays* L.)-based research project funded by the agency to date (USDA NIFA 2011; Eigenbrode et al. 2014). The team is investigating complex carbon (C), nitrogen (N), and water cycles to increase the efficiency and resiliency of

corn-based cropping systems while working to decrease the environmental footprint under climate change. The Sustainable Corn team is a partnership across 11 institutions and 9 states, and includes 140 active members from 19 disciplines working within biophysical and social research, extension, and education (see table 1 in Eigenbrode et al. 2014 for more details). A major component of the proposed work was to collect data from a field research network comprising existing and newly established experimental sites across the US Corn Belt while implementing standardized methodologies that would allow for synthesis across differing soil types, environments, and management operations.

Increasingly, proposals are encouraged (or required) to be multi-institutional and multidisciplinary with a combination of scientists that may or may not have previously worked together. This approach to funding can potentially create fragmented, disparate sets of data that result in a series of publications, but do not encourage the data to live on beyond the funded team due to insufficient funds or lack of professional archivists with the necessary capacity or infrastructure (King 2011; Wolkovich et al. 2012). The type and length of funding should be considered by teams when developing their data management plan; we believe different approaches should be considered given team goals, allocated dollars, existing infrastructure, and expected use of the data during and beyond the funding period. The amount of agricultural data that are publicly available is only a fraction of the research conducted, which leaves much undiscovered, not utilized by other scientists (current and future), and with a high propensity to be lost over time (White and van Evert 2008; Vines et al. 2014).

“Big science” is a loosely used term to describe new scientific approaches available thanks to modern technology, including massively scalable computing and high speed networks connecting nearly every corner of the world. This technology permits researchers to instantly

share and analyze increasingly large and diverse datasets. There is no one technological solution that facilitates big science, and large collaboration projects create a new set of problems traditional scientists have not previously experienced. A key problem teams encounter is the deluge of diverse, unfamiliar datasets with insufficient project preparedness to adequately handle them (Haendel et al. 2012; Michener and Jones 2012). The Sustainable Corn team can be categorized as a big science or big interdisciplinarity team due to the number of research sites, complexity of data, and users dependent on the database (Eigenbrode et al. 2014). Our approach to data management and project collaboration serves as a case study that may be a viable option for consideration by other multi-institution and multidisciplinary teams.

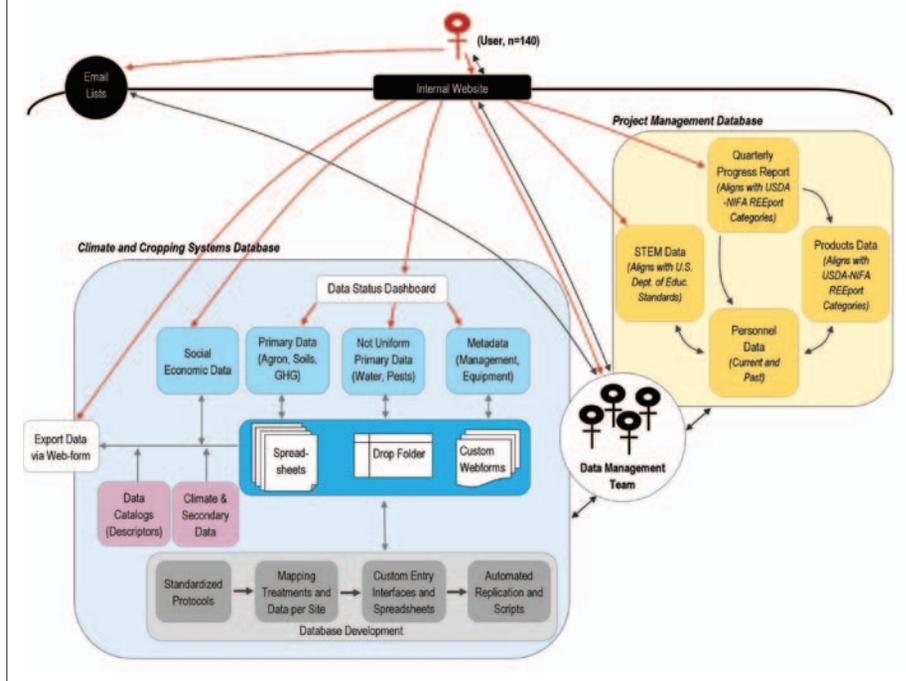
CONCEPTUAL APPROACH

No multi-institution infrastructure was in place at the time of receiving the award to support project and data collaboration needs. Time was a precious commodity as field data collection was to start immediately and the database would need to support the next five years of effort. Also, the expectation of USDA NIFA funded Coordinated Agricultural Projects (CAPs) was that findings related to field research, synthesis, and modeling would be produced during the funded period and not simply toward the end, which necessitated data upload, review, and management occurring concurrently and in a timely fashion to the collection phase to ensure use by other team members. An initial set of needs were identified including (1) an authenticated virtual space that allowed sharing of files, editing of content, and revision control; (2) a system for data entry, review, analysis, and long term storage; and (3) a Wiki website for interactive editing of internal content. Although this paper provides an overview of the team's research data approach, similar strategies and tools were utilized for overall project data to have a streamlined approach and

Daryl E. Herzmann is systems programmer, Lori J. Abendroth is project manager, and Landon D Bunderson is data manager for the Climate and Corn-based Cropping Systems Coordinated Agricultural Project at Iowa State University in Ames, Iowa.

Figure 1

Workflow describing the team's research (blue shading) and project (yellow shading) databases highlighting the one-directional workflow (red lines) and expectation of team members (users) contrasted with the multidirectional workflow (grey lines and arrows) and expectation of data personnel. The entry portal for all databases is through the internal team website. The workflow has been simplified in some aspects graphically. Research and project management databases utilize similar approaches and technology although only the research component is discussed in this paper.



provide consistency and synergy for team members (figure 1).

With members located across nine states, the Sustainable Corn team had to employ methods that were easy to learn, transparent, and powerful for successful virtual interactions. The establishment of a virtual environment to facilitate centralized data collection presents a number of challenges, which include identity management, data provenance (the origin and curation), disparate software platforms, and data formatting. While the Internet has greatly bridged the physical distance gap between scientists, the recent proliferation of web browser based technologies has mitigated many of the aforementioned issues with virtual environments.

The building out of a custom suite of tools on a local server includes the installation and security hardening of server software, designing of web data entry forms, programming of these forms to save data to a database, quality control routines of this information, and download inter-

faces of this data. The aforementioned work does not include identity management of collaborators, the curation process of reviewing data, and training aspects of data entry. Many of these steps are typically accomplished by a dedicated information technology staff member(s) and perhaps a few tech-savvy power users on the project. While certainly possible, these steps would have taken a number of years to complete. The desire to achieve the ideal technical implementation was balanced with the immediate need of project members to start collecting data and collaborating; every effort was made to accomplish the best technical approach while meeting needs of real users (Bach et al. 2011). This was the motivation behind using cloud computing, which in 2011, was rapidly gaining adoption as companies such as Google built their web accessible applications. The Sustainable Corn team's approach uses traditional technologies, like a relational database and conventional Microsoft Excel (Microsoft

Corp., Redmond, Washington) spreadsheets, while leveraging cloud-based tools and services provided by Google with automated syncing processes written, when appropriate.

It is important to utilize cloud technologies that are similar in appearance and functionality to existing software packages used for data entry. Agricultural researchers are accustomed to the Microsoft Excel style of data entry and manipulation (Hunt et al. 2001). For example, Google's spreadsheet application is similar to Excel, yet runs within a web browser and interfaces with data within Google cloud, allowing synchronization to a server at Iowa State University (ISU) and immediate viewing by team members. This varies substantially from the more traditional workflow in which data are shared with other collaborators via email or stored on some shared network file storage space.

Technology alone is not the panacea of collaboration, but requires individuals (data personnel) that can build and manage the environment. A thorough comprehension of the project, workflow, and its various components allows technological solutions to be developed that directly meet the needs of the collectors and users of the data for the project. Data team members must have extensive knowledge of the project science, data types, protocols, and methods. They should also be familiar with the current data provenance methods, tools for entry, analysis, and modelling. Beyond having a strong technical knowledge, it is crucial that these individuals are able to translate concepts and requirements in ways that users can easily understand along with having an innate ability to work within the full spectrum of the data cycle. They must also be able to work within the "grey" areas that often impede team science because of an inability of others to know what is missing or needed by others. Multidisciplinary science includes data types that can be messy or unfamiliar to many and must be distilled and explained into simpler information types. Although technological approaches often receive the majority of time and discussion in papers, we believe the right personnel and skillsets can be equally important, are a key indicator of success, and should be examined

just as thoroughly. The Sustainable Corn data team members (e.g., authors of this paper) have backgrounds in climate science, agronomy, biology, ecology, field plot research, statistical packages, programming languages, technology, and management. These qualifications are important when determining the appropriate schema to represent collected data, the infrastructure to support the data collection, the ability to conduct quality control and checks on data entered, and the needs of others on the team for data.

DEVELOPMENT

An initial estimate of data collection volumes determined there would not be any technical limitations to storing these data on the cloud with a redundant, noncollaborative copy stored on local preexisting infrastructure at ISU. While the automated sensor deployments common within environmental data sets will collect up to a few million observations, the total size of Sustainable Corn data will only reach a few gigabytes. While size was not a constraint, the complexity came in the diversity and methods by which the data were collected. The implementation of accounting for this complexity was not appealing to accomplish via the generation of data entry software on the local ISU data server. Instead, we determined customized entry cloud spreadsheets and interfaces would best handle the complexity with the option of writing software against the data stored which provides a level of abstraction to the data within the project itself. Others are free to write code to interface with the web services to manipulate the data just as the official Sustainable Corn project interfaces do.

Today's web browser allows for centralized identity management (session cookies), data provenance through transactional web services tracking access and changes, the standardization of web applications using HTML/Javascript technologies, and the development of community data formats and transports. Web browser users are able to seamlessly pass between federated websites using a single identity, modify shared data stored in the cloud through no-software-installed applications, and export this data in many

different formats on demand. Google's identity management allows for accounts to be used by team members to access the project's internal team site and the database all within the same entry portal.

An expansive field research network was established in all project states (Iowa, Illinois, Indiana, Michigan, Missouri, Minnesota, Ohio, Wisconsin), except South Dakota, and includes 35 experimental sites including long-term and newly established research plots. As described in Kladvik et al. (2014), protocols were developed by team scientists and encompass treatments within five categories (tillage, crop rotation, drainage water management, N fertilizer management, and landscape position) and measured data spanning agronomic, soil, water, greenhouse gas, integrated pest management, and weather. The establishment and use of standard sampling methodologies across research sites and personnel helps to minimize variations, ensure scalability, and allows for transparent management of data. The worth and value of standard methods in agricultural research has been recognized, and there is great potential for more widespread utilization (Hunt et al. 2001; Tenopir et al. 2011; Olson 2013; White et al. 2013). Team protocols have helped to ensure consistency in data collected among sites; accountability of data originating from each site; and a consistent set of data for systems analysis and predictive modeling across a suite of local, regional, and national scale models.

The management of project research data consists of a hybrid use of Google collaboration tools along with traditional relational database software. For a discussion of relational databases for agroecological research, see van Evert et al. (1999). Through the use of Google web services, these data repositories are automatically synced via custom software to the local database at ISU. This syncing limits issues with data versioning and allows data providers to manipulate and data users to download data through customized interfaces. The manipulation and download interfaces are organized within an internal website (Google Sites) and accomplished through two mechanisms: Google Drive (collaboratively edited spreadsheets and

work documents) and Google Gadgets (customized interfaces). Using Google Gadgets, which reside on the Google Sites Wiki pages and interface with data stored within Google Spreadsheets, we created a custom interface for all management metadata (figure 2). A feature to the Google Drive is data versioning, which allows oversight of what data are changing when and by whom allowing data changes to be rolled back if necessary. This is a very important feature that some data management systems do not support.

In general, team members utilize the following workflow when interfacing with the team's "Climate and Cropping Systems" database (see figure 1):

1. Team members enter data into a customized Google Gadget (webform) or site-specific generated Google Drive spreadsheet. Unstructured data sets are uploaded to Google Drive for manual processing by the data manager.
2. Data are gathered, processed, and quality controlled within the ISU housed database. Data are synchronized between the ISU database and Google Drive.
3. Team members who wish to use the data may request it via Google Gadgets (webforms) that materialize data from the ISU database and/or directly from the Google Spreadsheets.

To allow simple, straightforward, and transparent data entry portals that also force uniformity and structure, we began by dictating the exact combination of treatments and data variables that would originate from each site over the course of the five years. Two relational spreadsheets were developed to encompass the complexity of the field research network using unique site identifiers and coding systems. One is a master treatment matrix encompassing all treatments (55), and the other is a master data matrix including all data variables (95). Although consistent, standardized protocols were developed, variations existed in the specific treatments or data collected due to differing capacities across the 35 research sites. Codifying these protocols into database schema and metadata was completed in the first year. Data catalogs documenting standards, units, and explanations for the metadata, treatments, and measured data were devel-

Figure 2

Screenshot of the web interface by which collaborators enter, edit, and delete field management information (metadata). This is an example of the field operations tab entry webform for a research site in the 2013 season. This application is a Google Gadget, which resides within a Google Site, and saves the data to a Google Spreadsheet. The code for the interface is written in Javascript and HTML.

Select Research Site:

Select Year. Enter information under all 5 tabs. Please click the "SUBMIT" button to save your edits for each tab you enter data for.

2011 Crop 2012 Crop **2013 Crop** 2014 Crop 2015 Crop

Field Operations Management Pesticides Site Map & GPS Notes

Add to or edit previous Field Operations entries for 2013:

- 10/8/2012 plant_rye-soy-res [Edit](#) [Delete](#) [View](#)
- 11/13/2012 tillage_chisel [Edit](#) [Delete](#) [View](#)
- 4/25/2013 termination_rye_corn [Edit](#) [Delete](#) [View](#)
- 5/14/2013 plant_corn [Edit](#) [Delete](#) [View](#)
- 5/22/2013 sample_soilnitrate [Edit](#) [Delete](#) [View](#)
- 5/23/2013 termination_rye_soy [Edit](#) [Delete](#) [View](#)
- 6/6/2013 plant_soy [Edit](#) [Delete](#) [View](#)
- 6/13/2013 fertilizer_synthetic [Edit](#) [Delete](#) [View](#)
- 10/21/2013 harvest_soy [Edit](#) [Delete](#) [View](#)
- 10/29/2013 harvest_corn [Edit](#) [Delete](#) [View](#)

Select Operation:

Fertilizer Date:

Form of Fertilizer:

Which crop was this applied to:

Application Type:

What is the chemical formulation of the fertilizer applied? (example: urea, superphosphate, etc.)

Was a stabilizer (e.g. inhibitor or slow release) used?

oped to ensure all are clearly interpreted by team members.

To create customized entry spreadsheets that aligned with users' Excel files, each treatment occurring at a particular site was assigned a unique plot identifier by the user; this plot identifier was typically a number sequence identical to that located on the plot stake. These plot identifiers are then combined with the master treatment and master data matrices to generate custom entry spreadsheets that only show the viable treatment and data combinations for a particular site. The primary sorting criteria for field personnel when entering

into these spreadsheets are the plot identifiers as these can be numerically sorted and align directly with their Excel spreadsheets allowing them to copy and paste.

Metadata detailing site management operations are entered through a customized interface which prompts members with a menu of options that are populated with drop-down lists and open text boxes where information pertaining to land management, treatments applied, and practices important for data interpretation are entered (figure 3). Existing standards have been incorporated as applicable, such as use of SI units and those stated within the

International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) and the Agricultural Information Management Standards (AGROVOC) (Hunt et al. 2001; FAO 2012; ASA CSSA SSSA 2013). We currently have approximately 100 metadata types that can be entered per research site. This Google Gadget interface populates into and reads out from a Google Spreadsheet that can then be related to field data. Metadata can be entered, edited, or deleted by team members in real-time via the Google Gadget. Python (van Rossum 1990) scripts have been developed to interface with these spreadsheets via Google's GDATA Application Programming Interface (API). These scripts download the raw data from the Google Spreadsheets and mine them into a traditional relational database for aggregation, quality control, and backup purposes. The Google Drive API also provides metadata on spreadsheet changes that allows the download scripts to efficiently process only new data.

Datasets generated via automated methods or continuous (such as weather data, moisture sensor readings, etc.) are uploaded directly to the Google Drive and stored in their native format, be it a structured text file or binary data. Processing scripts are then built to download these files when the Google Drive API advertises them as new or modified. These scripts process the raw data into the local database.

With the data centrally collected within the Google cloud and relational database, the third step was to provide download interfaces that collated and presented data in a user-friendly format. Again, Google Gadgets were chosen to provide metadata-driven and dynamically generated website forms that present users with download options to select from the diverse datasets collected. These forms call web services that package the data into a format that is importable into a desktop spreadsheet program or statistical software package. At any point, users are free to directly interrogate the Google Spreadsheets and manually collect the data they need as it is a completely accessible and transparent system.

IMPLEMENTATION

Our strategy has been a phased approach, releasing entry interfaces and spreadsheets to the team as they are finalized (following testing) to allow data upload and increase user familiarity and ownership. With a phased approach, it was crucial to get the data entry portal live as entry and quality control are the most time consuming activities; this potential bottleneck had to be avoided to ensure project goals were met. The entry portals (metadata and field data) were complete in years 1 and 2 (2011 and 2012) of the project, and the download interfaces, similar to the customized entry interfaces, were completed in year 3 (2013).

Tracking the inflow of data is essential in order for the database team to ensure uniformity and quality of data, compliance of research team members, and proper storage of all data types. The database has two virtual dashboards where team members can view which data were collected, which were marked as “did not collect,” and which data are not yet entered into the system (figure 3).

Dashboards are invaluable when managing the inflow of data. As data upload deadlines approach, team members are directed to view the dashboard and can

easily see where their data input may be lacking. Also, a yearly audit is made on each spreadsheet, and the data are reviewed with researchers who have responsibility for uploading data. This is essential to ensure missing data are not still pending, data are in the correct units, and all required data have been collected. The audit also affords the opportunity for researchers and the database team to review requirements and give feedback to team leadership. In addition to yearly audits, the database team is in constant contact with each researcher as data are coming in. In order for datasets to be used by the modeling team and future researchers, metadata have to be extremely detailed, thus there is a requirement for constant communication.

The combined approach of using a traditional relational database in tandem with cloud technologies has been very positive overall from a data management perspective as well as for users. There are technical nuances when working with a third-party free software (Google) that do require some time and attention by database personnel. The cloud is rapidly being developed, however, and the benefits of leveraging this have far outweighed these minor irritations and allowed us to have

a fairly robust and functional system running very quickly.

The team’s data management approach and research database (Climate and Cropping Systems database) has provided structural and collaborative advantages for team members and working groups. By the end of the project (2016), we believe this approach will be documentable in terms of increased collaboration, synthesis, and greater deliverables (such as publications) to our funding agency. Positive outcomes of the team’s approach encompass the data and research cycles including the following:

- Standardization and decoding of soil, water, and crop datasets for greater application across disciplines.
- Expedited discovery of relevant project data through integrated search provided by the Google cloud platform.
- Minimal loss of data and supporting information due to centralized storage and metadata assigned to data.
- Improved transparency and reproducibility of findings as data are centrally located for all team members.
- Increased speed and mobilization of the team when addressing these emerging issues or grand challenges.

Figure 3

Screenshot of the data status dashboard per research site and measured variable with the number of data points denoted. The example here has been simplified by eliminating research site and variable identifiers. This web interface dynamically generates mini progress bars based on the amount of data uploaded versus the amount of data expected, which is derived from entered metadata. Four status options exist per variable including: data have been entered (green), missing data (blue), did not collect (yellow), and not entered yet (red).

SITEID	AGR1	AGR2	AGR3	AGR4	AGR5	AGR6	AGR7	AGR8	AGR9	AGR10	AGR11	AGR12	AGR13	AGR14	AGR15	AGR16
ALL SITES	345 / 200	181 / 131	160 / 200	335 / 210	130 / 200	212 / 200	247 / 200	18 / 200	31 / 200	311 / 100	34 / 90	28 / 102	200 / 200	111 / 144	185 / 91	190 / 91
SITE_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SITE_2	20	0	0	0	0	0	0	0	0	20	20	0	0	0	0	0
SITE_3	84	0	0	0	0	0	0	0	0	64	31	0	0	0	0	0
SITE_4	100 / 22	87 / 41	86 / 64	87 / 41	0	0	0	0	0	120	97 / 41	0	0	0	0	0
SITE_5	12	12	0	0	0	0	0	0	0	12	12	0	0	0	0	0
SITE_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SITE_7	24	24	0	0	0	0	0	0	0	18	18	18	18	24	24	24
SITE_8	12	12	0	0	0	0	0	0	0	12	12	0	0	0	0	0
SITE_9	12	0	0	0	0	0	0	0	0	12	12	0	0	0	0	0
SITE_10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SITE_11	12	12	0	0	0	0	0	0	0	12	12	12	12	12	12	12
SITE_12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SITE_13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SITE_14	12	12	0	0	0	0	0	0	0	12	12	12	12	12	12	12
SITE_15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Copyright © 2014 Soil and Water Conservation Society. All rights reserved. Journal of Soil and Water Conservation 69(6): 180A-185A www.swcs.org

ACKNOWLEDGEMENTS

The authors greatly appreciate the feedback and willingness to explore this new approach to data management by Sustainable Corn team members. The team's centralized Climate and Cropping Systems database has been developed for a regional collaborative project supported by the USDA National Institute of Food and Agriculture (NIFA), Award No. 2011-68002-30190, Cropping Systems Coordinated Agricultural Project: Climate Change, Mitigation, and Adaptation in Corn-based Cropping Systems. Project website: sustainablecorn.org. The 11 institutions comprising the project team include the following land grant universities and USDA Agricultural Research Service (ARS): Iowa State University, Lincoln University, Michigan State University, The Ohio State University, Purdue University, South Dakota State University, University of Illinois, University of Minnesota, University of Missouri, University of Wisconsin, and the USDA ARS in Columbus, Ohio.

REFERENCES

- ASA CSSA SSSA (American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America). Climate Change Position Statement Working Group. 2011. *In* Position statement on climate change. Working Group Report. Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.
- ASA CSSA SSSA. 2013. Publications handbook and style manual: Chapter 7. *In* Units and measurements. <https://www.agronomy.org/files/publications/style/chapter-07.pdf>.
- Bach, K., D. Schäfer, N. Enke, B. Seeger, B. Gemeinholzer, and J. Bendix. 2011. A comparative evaluation of technical solutions for long-term data repositories in integrative biodiversity research. *Ecological Informatics* 11:16-24, doi:10.1016/j.ecoinf.2011.11.008.
- Eigenbrode, S., L. Wright Morton, and T. Martin. 2014. Big interdisciplinarity to address climate change and agriculture: Lessons from three USDA Coordinated Agricultural Projects. *Journal of Soil Water Conservation* 69(6):170A-175A, doi:10.2489/jswc.69.6.170A.
- FAO (Food and Agriculture Organization of the United Nations). 2012. Agricultural information management standards: AGROVOC. <http://aims.fao.org/standards/agrovoc>.
- Haendel, M.A., N.A. Vasilevsky, and J.A. Wirz. 2012. Dealing with data: A case study on information and data management literacy. *PLoS Biology* 10(5):1-4, doi:10.1371/journal.pbio.1001339.
- Hatfield, J.L., K.J. Boote, B.A. Kimball, L.H. Ziska, R.C. Izaurralde, D. Ort, A.M. Thomson, and D. Wolfe. 2011. Climate impacts on agriculture: Implications for crop production. *Agronomy Journal* 103:351-370, doi:10.2134/agronj2010.0303.
- Hunt, L.A., J.W. White, and G. Hoogenboom. 2001. Agronomic data: advances in documentation and protocols for exchange and use. *Agricultural Systems* 70(2-3):477-492, doi:10.1016/S0308-521X(01)00056-7.
- King, G. 2011. Ensuring the data-rich future of the social sciences. *Science* 331:719-721, doi:10.1126/science.1197872.
- Kladivko, E.J., M.J. Helmers, L.J. Abendroth, D. Herzmann, R. Lal, M. Castellano, D.S. Mueller, J.E. Sawyer, R.P. Anex, R.W. Arritt, B. Basso, J.V. Bonta, L. Bowling, R.M. Cruse, N.R. Fausey, J. Frankenberger, P. Gassman, A.J. Gassmann, C.L. Kling, A. Kravchenko, J.G. Lauer, F.E. Miguez, E.D. Nafziger, N. Nkongolo, M. O'Neal, L.B. Owens, P. Owens, P. Scharf, M.J. Shipitalo, J.S. Strock, and M.B. Villamil. 2014. Standardized research protocols enable transdisciplinary research of climate variation impacts in corn production systems. *Journal of Soil and Water Conservation* 69(6):532-542, doi:10.2489/jswc.69.6.532.
- Marx, V. 2012. My data are your data. *Nature Biotechnology* 30(6):509-511, doi:10.1038/nbt.2243.
- Michener, W.K., S. Allard, A. Budden, R.B. Cook, K. Douglass, M. Frame, S. Kelling, R. Koskela, C. Tenopir, and D.A. Vieglais. 2011. Participatory design of DataONE – Enabling cyber infrastructure for the biological and environmental sciences 11:5-15, doi:10.1016/j.ecoinf.2011.08.007.
- Michener, W.K., and M.B. Jones. 2012. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 27(2):85-93, doi:10.1016/j.tree.2011.11.016.
- Olson, K.R. 2013. Soil organic carbon sequestration, storage, retention and loss in U.S. croplands: Issues paper for protocol development. *Geoderma* 195-196:201-206, doi:10.1016/j.geoderma.2012.12.004.
- OSTP CAST (Office of Science and Technology Policy: President's Council of Advisors on Science and Technology). 2012. Report to the President on agricultural preparedness and the agriculture research enterprise. Washington, DC. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_agriculture_20121207.pdf.
- Overpeck, J.T., G.A. Meehl, S. Bony, and D.R. Easterling. 2011. Climate data challenges in the 21st century. *Science* 331:700-702, doi:10.1126/science.1197869.
- Reichman, O.J., M.B. Jones, and M.P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331:703-705, doi:10.1126/science.1197962.
- Tenopir, C., S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6(6):1-21, doi:10.1371/journal.pone.0021101.
- USDA NIFA (USDA National Institute of Food and Agriculture). 2011. NIFA announces grant to study the effects of climate change on agricultural and forest production. Washington, DC. http://www.csrees.usda.gov/newsroom/news/2011news/02181_climate_change_cap.html.
- van Evert, F.K., E.J.A. Spaans, S.D. Krieger, J.V. Carlis, and J.M. Baker. 1999. A database for agroecological research data: II. A relational implementation. *Agronomy Journal* 91:62-71, doi:10.2134/agronj1999.00021962009100010010x.
- van Rossum, G., et al. 1990. The python language reference. Python Software Foundation. <http://docs.python.org/py3k/reference/index.html>.
- Vines, T.H., A.Y.K. Albert, R.L. Andrew, F. Débarre, D.G. Bock, M.T. Franklin, K.J. Gilbert, J.-S. Moore, S. Renaut, and D.J. Rennison. 2014. The availability of research data declines rapidly with article age. *Current Biology* 24:94-97, doi:10.1016/j.cub.2013.11.014.
- White, J.W., and F.K. van Evert. 2008. Publishing agronomic data. *Agronomy Journal* 100:1396-1400, doi:10.2134/agronj2008.0080F.
- White, J.W., L.A. Hunt, K.J. Boote, J.W. Jones, J. Koo, S. Kim, C.H. Porter, P.W. Wilkens, and G. Hoogenboom. 2013. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture* 96:1-12, doi:10.1016/j.compag.2013.04.003.
- Wolkovich, E.M., J. Regetz, and M.I. O'Connor. 2012. Advances in global change research require open science by individual researchers. *Global Change Biology* 18:2102-2110, doi:10.1111/j.1365-2486.2012.02693.x.