# Mapping Oklahoma Mesonet Sensor Datastreams

**Poster** Submitted by

**Betsy Van der Veer Martens**
School of Library & Information Studies
University of Okahoma
bvmartens@ou.edu

**Christopher A. Fiebrich and Bradley G. Illston**
Oklahoma Mesonet
University of Oklahoma
fiebrich@ou.edu; illston@ou.edu

**POSTER ABSTRACT**

This poster describes collaborative work in progress between researchers at the Oklahoma Mesonet of the Oklahoma Climatological Survey and the School of Library & Information Studies, both located at the University of Oklahoma, to identify the patterns of environmental sensor data distribution and dissemination for use in agricultural management, climate analysis and weather forecasting by both commercial and governmental entities, emergency response and management, meteorological education at all levels, and scientific research in a variety of disciplines, ranging from archaeology to zoology.

This work differs from similar projects in its inclusion of ways in which the Mesonet's sensor datastreams are referenced by various "upstream" stakeholders in technological and other communities at different points in their temporal flow as well as by their more traditional "downstream" stakeholders from the scientific community. It includes both bibliometric and datastream analytic components in its methodology.

**Keywords**

Bibliometrics, data, data citation, data curation, data management, datastreams, Oklahoma Mesonet, sensors

## INTRODUCTION

Data, however defined, have always played a central role in information science (Buckland, 1991), and their management has become increasingly critical to information scientists as well as to a growing cohort of data scientists (Buckland, 2011). Behind much of this so-called "data deluge" is the rapid increase in the number of environmentally embedded wireless sensors providing real-time, heterogeneous data for a wide variety of applications.

*ASIST 2015,November 6-10*, 2015, St. Louis, MO, USA.

Sensors of all types are expected to reach some 25 billion by 2020 and, accordingly, sensor datastreams will then exceed current social media "big data" proportions (Gartner Group, 2014).

Comparative studies focusing on the various factors involved in the creation and curation of sensor data, particularly sensor datastreams in the sciences, include work by Borgman (2015), Cragin, Chao, & Palmer (2010), Ganguly, Omitaomu, Fang, Khan & Bhaduri (2007), Mayernik, Wallis & Borgman (2013), Wallis *et al*. (2007), and Wallis (2012). While there is controversy as to whether classic bibliometric techniques can be usefully applied to datasets and datastreams in such diverse environments (*e.g.*, Borgman, 2015) and to such "wicked" problems (*e.g.*, Awre *et al*., in press), their use in conjunction with other "practice field" study techniques (*e.g.*, Martens, 2011) may prove particularly relevant to those responsible for managing these innovative "knowledge infrastructures" (Edwards *et al*., 2012).

The specific Mesonet installation described here predates the popularity of "big data" research, in that it can be said to have begun in the late 1980s, as the result of an ongoing collaboration between agricultural researchers at Oklahoma State University and meteorological researchers at the University of Oklahoma to develop a nearly real-time, extremely reliable source of data about local weather conditions across the state (Brock, 2013; Crawford, 2013). This need was felt to be particularly critical for Oklahoma, given the state's history of drought and drainage issues. Losses due to crop failure (Ding, Hayes & Widhalm, 2011) and urban flooding (Waite, 2011) ran into the billions. The initiative garnered support throughout the state, largely due to the efforts of those involved in this unprecedented partnership among scientists at the state's two major research universities.

As a result, the Oklahoma Mesonet, a statewide network of 120 automated environmental monitoring stations, was officially launched in 1994. These 10-meter-tall towers, one or more of which is located within each of Oklahoma's 77 counties, provide regular measurements of air and soil temperature, barometric pressure, rainfall, relative humidity, solar radiation, soil moisture, and wind speed and direction, both direct and calculated, includng

instrumentation data, all of which are then packaged into "observations" that are regularly transmitted every 5 minutes to the Oklahoma Climatological Survey, where the data quality is immediately verified and then made available to Mesonet users. This direct distribution of data, which is in addition to the Oklahoma Mesonet's several other data distribution channels, provides an interesting contrast to other recent studies in data citation practices (Robinson-García, Jiménez-Contreras & Torres-Salinas, 2015).

Although some commercial and media entities are licensed users, the data are also made freely available to other users who wish to access them. These users include construction firms, emergency and public safety personnel, farmers, gardeners, ranchers, local government entities, media outlets, science teachers, transportation firms, water experts, weather forecasters, and a variety of others interested in Oklahoma's drought conditions, heat impact, dry line movements, snow fall, squall lines, thunderstorms, and tornadoes.

Researchers using the data represent a variety of fields, from the team of zoologists studying the effect of prolonged drought on bison in the Tallgrass Prairie Preserve near Pawhuska, Oklahoma (Maichak, Schuler, & Payton, 2004) to the atmospheric science graduate student exploring potential climate change impacts on wind turbine development in Oklahoma (Dryden, 2011).

The Oklahoma Mesonet has been termed the "gold standard" among statewide climate and weather networks because of its well-known attention to quality assurance (National Research Council, 2008, p. ix). Nevertheless, its existence continues to be somewhat contingent upon constant communication of its value to federal and non-federal funders, state legislators and other governmental entities, local community supporters in each of the Oklahoma counties containing a Mesonet site, media outlets, public safety officials, commercial users, and the scientists who cite its data in their own work.

These "social" aspects underlying the more formal knowledge infrastructure (Borgman, Darch, Sands, Wallis & Traweek, 2014) in this project may differ from previously studied data sharing practices among particular groups of scientists (*e.g*, Kowalczyk & Shankar, 2011; Mayernik, 2015). They represent another important aspect of data management today, and the Oklahoma Mesonet provides a unique opportunity to better understand them.

## METHODOLOGY

The current project, begun in the summer of 2015, is using a multi-method approach in the study of the Mesonet sensor datastreams. In addition to performing bibliometric analysis of the scientific journal articles in different disciplines that are citing Mesonet-related authors and papers, the specifics of the individual datasets in these articles are being mapped to the actual temporal and geographical sensor datastreams from which the data was drawn. A similar effort will be piloted on publications in various media used for non-academic purposes, such as weather forecasting.

All initial categorization and coding of data usage are first done by project researchers not affiliated with the Oklahoma Mesonet, then checked for accuracy by researchers affiliated with the Oklahoma Mesonet, then re-checked by the researchers not affiliated with the Oklahoma Mesonet. This is to insure that both scientific and non-scientific aspects of the data usage are captured correctly.

This categorization of technical and temporal aspects of data usage will precede the development of a model of sensor datastream "flow" specific to the Mesonet based on a preliminary schematic based on similar models for other sensor datastream applications, such as that for earth sciences, as reported in Borgman, Wallis & Mayernik (2012), and that for ecology, as reported in Porter, Hanson, & Lin (2012). The effort also intends to add new detail to the data types and data stages identified in those two initiatives. The resulting schematic will then be cross-checked with different Oklahoma Mesonet stakeholders for its potential utility and validity.

## PROJECTED PROJECT OUTCOMES

The goals of this research in progress are (1) to identify specific uses of the Oklahoma Mesonet datastreams in different scientific and technical communities; (2) to produce an easily understandable, widely applicable data usage schematic for the Oklahoma Mesonet community that offers empirically-grounded insights from the different user communities for these data and that also has a practical purpose in promoting this scientific data resource to both existing stakeholders and potential stakeholders that may help to sustain it; and (3) to provide unique data citation information from the Oklahoma Mesonet sensor datasteams to the bibliometric and data curation/data management research communities to encourage additional investigation.

## REFERENCES


Awre, C. *et al.* (in press). Research data management as a "wicked problem." *Library Review 64(4/5)*. Retrieved July 30, 2015 from http://dx.doi.org/10.1108/LR-04-2015-0043

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge MA: MIT Press.

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014). The ups and downs of knowledge infrastructures in science: Implications for data management. In *Proceedings of the EEE/ACM Joint Conference on Digital Libraries. September 8-12, 2014, London, U.K.* (pp. 257-266).

Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and

technology collaborations. *Computer Supported Cooperative Work 21(6)*, 485-523.
Brock, F. (2013). Oral history interview with Fred V. Brock. Retrieved July 30, 2015 from http://dc.library.okstate.edu/cdm/compoundobject/collection/Spot/id/539/rec/3
Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science & Technology 42*(5), 351-360.
Buckland, M. K (2011). Data management as bibliography. *Bulletin of the American Society for Information Science and Technology 37*(6), 34-37.
Cragin, M. H., Palmer, C. L., & Chao, T. C. (2010). Relating data practices, types, and curation functions: An empirically derived framework. *Proceedings of the American Society for Information Science and Technology 47(1*), 1-2.
Crawford, K. (2013). Oral history interview with Ken Crawford. Retrieved July 30, 2015 from http://dc.library.okstate.edu/cdm/compoundobject/collection/Spot/id/621/rec/2
Ding, Y., Hayes, M. J., & Widhalm, M. (2011). Measuring economic impacts of drought: A review and discussion. *Disaster Prevention and Management 20(4*), 434-446.
Dryden, J. M. (2011). Potential climate change impacts on wind resources in Oklahoma: A focus on future energy output. M.S. Thesis. Oklahoma State University.
Edwards, P. *et al*. (2013) Knowledge infrastructures: Intellectual frameworks and research challenges. Ann Arbor MI: Deep Blue. Retrieved July 30, 2015 from http://deepblue.lib.umich.edu/handle/2027.42/97552
Ganguly, A. R., Omitaomu, Fang, Y., Khan, S. & Bhaduri, B. L. (2007). Knowledge discovery from sensor data for scientific applications. In J. Gama *et al.* (Eds.), *Learning from data streams: Processing techniques in sensor networks*. (pp. 205-228). Berlin: Springer.
Gartner Group. (2014). Gartner says 4.9 billion connected "things" will be in use in 2015. Retrieved July 30, 2015 from http://www.gartner.com/newsroom/id/2905717
Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology 45*(1), 247-294.
Maichak, E. J., Schuler, K. L., & Payton, M. E. (2004). Daily and seasonal behavior of bison on an Oklahoma tallgrass prairie. *Prairie Naturalist 36(2*), 103-118.
Martens, B. V. (2011). The production of practice theories. *Journal of the American Society for Information Science & Technology 62*(3), 586-593.
Mayernik, M.S. (2015). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*. Retrieved July 30, 2015 from DOI: 10.1002/asi.23425
Mayernik, M. S., Wallis, J. C., & Borgman, C. L. (2013). Unearthing the infrastructure: Humans and sensors in field-based scientific research. *Computer Supported Cooperative Work 22(1)*, 65-101.
National Research Council. (2008). *Observing weather and climate from the ground up: A nationwide network of networks*. Washington DC: National Academies Press.
Porter, J. H., Hanson, P. C., & Lin, C. C. (2012). Staying afloat in the sensor data deluge. *Trends in Ecology & Evolution 27(2)*, 121-129.
Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2015). Analyzing data citation practices using the Data Citation Index. *Journal of the Association for Information Science and Technology*. Retrieved July 30, 2015 from DOI: 10.1002/asi.23529
Waite, J. L. (2011). Using the urban regime framework to analyze hazard mitigation sustainability at the local scale. Ph.D. Dissertation, University of Maryland, College Park. Retrieved July 30, 2015 from http://drum.lib.umd.edu/handle/1903/12290
Wallis, J. C. (2012). The distribution of data management responsibility within scientific research groups. Ph.D. Dissertation, University of California, Los Angeles. Retrieved July 30, 2015 from http://escholarship.org/uc/item/46d896fm
Wallis, J. C., Borgman, C. L., Mayernik, M. S., Pepe, A., Ramanathan, N., & Hansen, M. (2007). Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries. *11th European Conference on Research and Advanced Technology for Digital Libraries,. Budapest, Hungary, September 16-21, 2007*. (pp. 380-391). Berlin: Springer.