


 Search JDH

 Search JDH

- [Subscribe to the RSS](#)
- [About](#)
- [Volumes](#)
- [Submissions](#)

## Table of Contents for Vol. 1, No. 1 Winter 2011

- [Introduction](#)
- [Articles](#)
  - [Academic History Writing and its Disconnects](#)  
Tim Hitchcock
  - [Defining Data for Humanists: Text, Artifact, Information or Evidence?](#)  
Trevor Owens
  - [Demystifying Networks, Parts I & II](#)  
Scott B. Weingart
  - [Clustering with Compression for the Historian](#)  
Chad Black
  - [Spatializing Photographic Archives](#)  
Marc Downie and Paul Kaiser
  - ["Humanities in a Digital Age" Symposium Podcasts](#)  
Jeremy Boggs, Alison Booth, Daniel J. Cohen, Mitchell S. Green, Anne Houston and Stephen Ramsay
  - [Philosophical Leadership Needed for the Future: Digital Humanities Scholars in Museums](#)  
Nik Honeysett and Michael Edson
  - [Critical Discourse in Digital Humanities](#)  
Fred Gibbs
- [Conversations](#)
- [Reviews](#)
- [Authors](#)

## Defining Data for Humanists: Text, Artifact, Information or Evidence?

### [Trevor Owens](#)

Data seems to be the word of the moment for scholarship. The National Endowment for the Humanities and a range of other funders are inviting scholars to “dig data” in their [“Digging into Data” grant program](#). Data itself is now often discussed as representing a [fourth paradigm for scientific discovery and scholarship](#) (PDF). What is a humanist to do in such a situation? Does data, in particular big data, require humanists to adopt a new methodological paradigm? Or, are the kinds of questions humanities scholars traditionally have explored through close reading and hermeneutic interpretation relevant to big data? In this brief essay I suggest some of the ways that humanists already think about and analyze their sources can be employed to understand, explore, and question data.

### What is Data to a Humanist?

We can choose to treat data as different kinds of things. First, as constructed things, data are a species of artifact. Second, as authored objects created for particular audiences, data can be interpreted as texts. Third, as computer-processable information, data can be computed in a whole host of ways to generate novel artifacts and texts which are then open to subsequent interpretation and analysis. Which brings us to evidence. Each of these approaches—data as text, artifact, and processable information—allow one to produce or uncover evidence that can support particular claims and arguments. Data is not in and of itself a kind of evidence but a multifaced object which can be mobilized as evidence in support of an argument.

## Data as Constructed Artifacts

Data is always manufactured. It is created. More specifically, data sets are always, at least indirectly, created by people. In this sense, the idea of “raw data” is a bit misleading. The production of a data set requires choices about what and how to collect and how to encode the information. Each of those decisions offers a new potential point of analysis.

Now, when data is transformed into evidence, when we isolate or distill the features of a data set, or when we generate a visualization or present the results of a statistical procedure, we are not presenting the artifact. These are abstractions. The data itself has an [artifactual](#) quality to it. What one researcher considers noise, or something to be discounted in a dataset, may provide essential evidence for another.

In the sciences, there are some tacit and explicit agreements on acceptable assumptions and a set of statistical tests exist to help ensure the validity of interpretations. These kinds of statistical instruments are also great tools for humanists to use. They are not, however, the only way to look at data. For example, the most common use of statistics is to study a small sample in order to make generalizations about a larger population. But statistical tests intended to identify whether trends in small samples scale into larger populations are not useful if you want to explore the gritty details and peculiarities of a data set.

## Data as Interpretable Texts

As a species of human-made artifact, we can think of data sets as having the same characteristics as texts. Data is created for an audience. Humanists can, and should interpret data as an authored work and the intentions of the author are worth consideration and exploration. At the same time, the audience of data also is relevant. Employing a reader-response theory approach to data would require attention to how a given set of data is actually used, understood, and interpreted by various audiences. That could well include audiences of other scientists, the general public, government officials, etc. When we consider what a data set means to individuals within a certain context, we open up a range of fruitful interpretive questions which the humanities are particularly well situated to explicate.

## Data as Processable Information

Data can be processed by computers. We can visualize it. We can manipulate it. We can pivot and change our perspective on it. Doing so can help us see things differently. You can process data in a stats package [like R](#) and run a range of statistical tests to uncover statistically significant differences or surface patterns and relationships. Alternatively, you can deform a data set with a process like [Spoonbill's N+7 machine](#), which replaces every noun in a text with the seventh word in the dictionary that follows the original, thus prompting you to see the original data from a different perspective, as Mark Sample's [Hacking the Accident](#) did for [Hacking the Academy](#). In both cases, you can process information—numerical or textual—to change your frame of understanding for a particular set of data.

Importantly, the results of processed information are not necessarily declarative answers for humanists. If we take seriously Stephen Ramsay's suggestions for [algorithmic criticism](#), then data offers humanists the opportunity to manipulate or algorithmically derive or generate new artifacts, objects, and texts that we also can

read and explore.<sup>[1]</sup> For humanists, the results of information processing are open to the same kinds of hermeneutic exploration and interpretation as the original data.

## Data Can Hold Evidentiary Value

As a species of human artifact, as a cultural object, as a kind of text, and as processable information, data is open to a range of hermeneutic tactics for interpretation. In much the same way that encoding a text is an interpretive act, so are creating, manipulating, transferring, exploring, and otherwise making use of data sets. Therefore, data is an artifact or a text that can hold the same potential evidentiary value as any other kind of artifact. That is, scholars can uncover information, facts, figures, perspectives, meanings, and traces of thoughts and ideas through the analysis, interpretation, exploration, and engagement with data, which in turn can be deployed as evidence to support all manner of claims and arguments. I contend that data is not a kind of evidence; it is a potential source of information that can hold evidentiary value.

## Conclusion

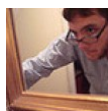
Approaching data in this way should feel liberating to humanists. For us, data and the capabilities of processing data are not so much new methodological paradigms, rather an opportunity for us to bring the skills we have honed in the close reading of texts and artifacts into service for this new species of text and artifact. Literary scholar Franco Moretti already has asked us to pivot, to begin to engage in [distant reading](#). What should reassure us all is that at the end of the day, any attempt at distant reading results in a new artifact that we can also read closely.

In the end, the kinds of questions humanists ask about texts and artifacts are just as relevant to ask of data. While the new and exciting prospects of processing data offer humanists a range of exciting possibilities for research, humanistic approaches to the textual and artifactual qualities of data also have a considerable amount to offer to the interpretation of data.

Originally published by Trevor Owens on [December 15, 2011](#). Revised March 2012.

[1]Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Champaign: University of Illinois Press, 2011).[↩](#)

## About Trevor Owens



Trevor Owens is a Digital Archivist with the National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress. He is finishing a PhD in social science research methods and instructional technology in the Graduate School of Education at George Mason University. Trevor wrote a book for Arcadia Press about the history of Fairfax County told through postcards, and published articles in the journals *Simulation & Gaming*, *Science Communication*, *Cultural Studies of Science Education*, and *On the Horizon*. In 2009 he was chosen by Read Write Web as one of the 50+ Semantic Web Pros to Follow on Twitter, and won the CW Bright Pixel Prize for building *PlayingHistory*, an open collaborative directory of digital historical games and interactives.



ROY ROSENZWEIG  
Center FOR  
History AND  
New Media



This work is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/).