

# What is the Thing that Changes? Space and Time through the Atlas of Historical County Boundaries

Douglas Knox  
Data Modeling Symposium  
March 16, 2012

Five simple words provide all anyone needs to take the first steps on the path of grappling with complex and rewarding data modeling issues, unwittingly: *How hard can it be?*

Counties in the U.S. ought to be easy to model. There are more than 3,000 of them, their shapes are known, and there are official census records and many other kinds of statistics collected about each of them. It is not difficult to find visualizations based on U.S. county data, and that fact is evidence that counties really are not all that difficult to model. If we had data from another year, we could make another map, and start making comparisons. But history complicates things.

Let's say we are looking at county-level census figures from the 19<sup>th</sup> century, perhaps from the Historical Census Browser that the University of Virginia has hosted for many years. We can look at the figures and infer that the population of Portage County, Wisconsin decline by about 400 in the 1840s—in fact it lost about 700 males and gained about 300 females, so its highly unbalanced sex ratio became slightly less unbalanced as it lost a quarter of its population. We could start speculating about what might have been going on, but our speculations will take on a different character when we realize Portage County in 1850 covered almost entirely different ground from Portage in 1840. And in fact during most of the intervening years it looked dramatically different from both of these shapes, and included a strip running all the way to the northern border. When we tell stories of change over time, we want to know what it is that changes, and necessarily make commitments about we think it is that stays the same. Because we model counties with names, we construct stories about them, and say their population changed, and that they moved.

Before talking any more about particular data modeling issues, I would like to share some brief

project stories, to provide a context. I've been invited here in part to represent libraries and archives. I'm not a librarian or archivist, but I worked for fifteen years on various digital publication projects at the Newberry Library in Chicago. In 2010 the Newberry completed a twenty-year project called the *Atlas of Historical County Boundaries*, edited by John Long. It was a carefully researched project that attracted a dedicated team of remarkable people—John Long, Peggy Sinko, Emily Kelley, Pete Siczewicz, Gordon Den Boer, and Laura Rico, especially, along with others at various times. I assisted part of the project in its later digital phases, which meant that I was in a position to be in awe of what this core team accomplished.

Back in the 1970s historian and archivist Lester Cappon headed a project at the Newberry to create an *Atlas of Early American History*, a beautiful large-format publication of newly prepared historical maps, made possible with funding from the National Endowment for the Humanities. The scholars engaged in creating these maps called themselves editors rather than data modelers. They were making creative, responsible interpretive decisions about how to represent things, what to include and what to leave out.

John Long was an energetic young early Americanist scholar who was assistant editor of the project, in the course of which the team discovered to their surprise that there was no good resource that they could rely on that documented the historical boundaries of U.S. counties. So with a successful NEH grant proposal in the late 1970s John started an innovative project to create a carefully researched reference that would document every change in county boundaries. This work would be grounded in primary sources, and would be expressed using the best technology of the time, in collaboration with the Cartography Lab at the University of Wisconsin. The project produced digital files and printed books for 14 states, but eventually ran aground on technological and methodological issues. I will say a bit more about the difficulties shortly.

At the end of the 1980s John was able to start anew, thanks again to NEH funding, with an improved research and mapping methodology that learned a lot from the earlier effort. The project gave

up its digital ambitions for a while and focused on correctly modeling, on paper, the editors' inferences from primary sources. They published 19 volumes covering 24 states and the District of Columbia by the year 2000, when the print publisher dropped the series. During the 1990s the project eventually moved toward digital methods, as desktop computing and GIS software came within reach inexpensively. Of necessity the project became its own digital publisher, which was not easy. By 2010 the project had completed its research and retrospectively digitized the research from the books to create a consistent set of digital files for the full U.S.

So why did it take so long? How hard can it be?

Counties in the U.S. are creatures of state governments. States (and in earlier times, territories and colonial governments) pass laws creating counties. The states establish the names and boundaries of counties. They set up the framework for county governmental structures and prescribe the terms and conditions under which local residents can organize county governments. County governments often serve as judicial districts, enforce state laws, keep records of births, deaths, marriages, and real estate transactions, and provide other functions of local government.

Authoritative evidence for what counties are and have been exists primarily in the legislative records of the states, supplemented by other sources. A primary source for this research looks something like this:

*Be it enacted by the General Assembly...*

This is an imperative sentence. The legislators of a state define a portion of territory by its bounds, starting at a specified point, proceeding to another point, and another, etc., and eventually returning to the initial point. This portion of territory they then declare to be a county, and then they assign a name to that county.

This is a performative speech act. Counties are things that legislators in the past made up. They have already been modeled, and they in fact come into being only by being modeled. This is perhaps an exception to the principle that the map is not the territory, or the model is not the thing. There is no

thing to begin with, in this case, except through its establishment by an intentional act of modeling. When we add a historical dimension, historical modeling is necessarily metamodeling. If the passage of time has done anything interesting with the original model, we can neither simply reproduce it nor ignore it. If that past model has changed, or is inconsistent, or incoherent, or unintelligible, we may be able to correct it, or we may need to incorporate it and model its difficulties and historicity as well, depending on what we are trying to accomplish.

The late 1970s county boundary project looked at documents like this and the national map they wanted to draw, and saw a series of line segments. Each segment in the interior of the country is of course a boundary between two counties, one to the left and another to the right. Rather than duplicate segments, we can treat them as atomic, the fundamental unit of what we want to model, and then just record which counties are on each side at various times. This was a reasonable way to think of the problem given the technology of the time. It would minimize the duplication of coordinates. Working with digital coordinate data was cumbersome in the early 1980s. Quality control consisted of editors pairing up to read aloud printed lists of coordinates, digit by digit. The feedback loop took weeks as the Newberry and the Wisconsin Cartography Lab traded printouts of coordinates and digital plotter results back and forth. This kind of model and the available methods for working with the model were ultimately too unwieldy to be sustainable with an acceptable level of quality. The project stopped after producing five printed volumes covering 14 states. The project's digital files have been on deposit with ICPSR since 1984, but the information is incomplete and has mistakes.

Having learned from this experience, when the project started up again it dropped digital methods altogether at first and focused on getting the research right. The key to the method was a step called "historical compilation." The compiler read through the laws, consulted references, and started drawing boundaries on tracing paper placed over a USGS base map, putting as many counties as possible together in a composite map in order to ensure that they had a complete, consistent set of boundaries and dates with the desired coverage. The new project succeeded with this method. The

editors thought through the information design of the books, and prepared boundary maps for each county on USGS base maps of constant scale (generally 1:500,000). The map scale and the width of the drafting pen set a limit to the precision of these maps, and that limit in turn influenced editing choices. Some changes were meticulously documented in the county chronologies but had to be identified as too small to map.

What they called “historical compilation” involved thinking about modeling several kinds of things at once. Most obviously, they were modeling counties as a succession of events and boundaries. Every county has its own story, told from its own point of view, with its own bounds. This was a primary way of organizing the books. Redundant line segments were no longer much of an issue.

Organizing information around each county individually is important, but historical compilation also had to consider how these individual counties are related to each other in space and time. Counties generally shouldn't overlap or leave awkward gaps. The compiler put as much information as possible on a single sheet of tracing paper in order to be able to work out the relations between counties. Sometimes state legislatures defined and named counties without yet organizing county governments, and they might “attach” such a proto-county to another county, so that the organized county would temporarily provide judicial and other services to residents of the area that was attached. Sometimes, counties did in fact overlap historically, and an overlap might have its own story to be worked out in a series of legal and legislative actions. Because a state legislature can make mistakes, and can discover and correct those mistakes, sometimes over a period of many years, the Atlas model had to be able not simply to represent the ideal model of counties intended by the legislature, but to have its own broader meta-model that can accommodate mistakes, ambiguities, and corrections.

Often, too, in the early years some portions of territory were not yet assigned to any county, and these areas had their own shapes and histories that needed to be modeled to provide continuity in the larger network of stories. The editors created of their own accord what they called “non-county areas,” often several contemporaneous ones distinguished by number, and they constructed county-like

chronologies for them. This approach ensured more or less full coverage over space and time within a state. The nature of these non-county areas is a consequence of the editorial commitments of the project.

Digital methods eventually came back into the project. GIS software from ESRI enabled boundaries to be recorded digitally as polygons (technically multipolygons) in shapefiles. This was now a pretty good match for what the project was already doing. There was a difference in precision, of course. Changes that had been too small to draw with a pen at 1:500,000 scale could now be recorded. In fact every coordinate of every line could be recorded with more decimal points of precision than any historical model really deserved.

In addition to shapes and boundaries, there are other aspects of organizing information about counties that are just as important. For each state, the project created a comprehensive data table with one row for every version or change to every county-like thing, including larger significant jurisdictional events in the history of the colony, territory, or state, whether or not those changes were mapped. The project invented its own identifiers for counties. Counties sometimes change names while remaining otherwise the same, and identifiers enable these versions to be brought together. In recent decades federal data has assigned counties a FIPS code (Federal Information Processing Standard), but there are no federal codes for counties that no longer exist, and FIPS codes have not been maintained consistently from decade to decade, but seem to have been reassigned at various times in past decades to maintain an ordering consistent with alphabetical order.

In the Atlas data, records that correspond to mapped boundaries are assigned sequential version numbers. The rows of the comprehensive table each describe a change from the perspective of a particular county, and so there is some redundancy as two or more counties often participate in a change. The column in the data describing each change was carefully edited, and a considerable amount of the effort of the project went into editorial work such as this, in addition to the boundary drawing and mapping. Each recorded change also had a citation to original laws and other primary and

secondary sources.

All of these changes were described in the context of a single state, and occasionally states have different models of what a county is at all. Louisiana calls its subdivisions “parishes” rather than counties, though in the past it had some counties as well. South Carolina at various times had parishes and districts as well as counties, and in some periods all three coexisted. Taking this into account adds a layer of complication to the model—another field in the data table, another optional layer in the interface, more complex relations to describe in prose.

When we talk about historical models, I suggest we should not think first and primarily about models that have time as a dimension—adding start and end dates to a table—but about past modeling choices that time has changed. This is often where important historical and ontological questions assert themselves. We may want to ask how the population of a particular county changes, and then go look up some census figures, assuming that counties themselves don't change. But they do.

Despite all the difficulties with a naïvely presentist model of what counties are, despite the many years of human labor that they have structured and recorded, not just in my colleagues' project but in county courthouses across the country, counties are still somewhat easy. They are built on a scaffolding of power, governance, instrumentation, and bureaucracy that has many layers of history, with complex changes in implicit and explicit models. The early colonies were surveyed by familiar landmarks, and the states west of Ohio had a virtual grid laid over them through a rectangular survey. The county boundary project staff grappled with some of these layers as they needed to, and edited and abstracted away some of them to create something useful. With counties now formally historicized in space and time, one might take them in all their changes as stable model for reference, and historicize again.

Counties provide scaffolding for building up other kinds of models, implicit and explicit. With the knowledge that county boundaries change, we can return the question of how population figures at the county level have changed over time. What do we mean by population figures? The categories used

by the census to record population characteristics also have changed. It might not be difficult to assign a URI to each of these characteristics in any single census year, but putting this information in RDF could be a disservice to historical inquiry if we implicitly commit ourselves to inferences that insist that the same term had to mean the same thing over time. In the history of the U.S. Census, the system of categories changed over time, and even if the same term is continuous, one might argue that its meaning changes because of changes in the system of which it is a part.

Our models are built of many components. Some of them we may treat as stable for the sake of having points of reference for measuring change, but for any of them one might at some point want to ask, or need to ask, whether that thing, too, has a history.