E-Social Science from a Systems Perspective: Applying the SACS Toolkit

Brian Castellani

Department of Sociology
Kent State University, Ashtabula
Ashtabula OH 44004
Phone: (430) 964-4331
Fax: (440) 964-4269
E-mail: bcastel3@kent.edu

Frederic Hafferty

Department of Behavioral Sciences
University of Minnesota, School of Medicine, Duluth
Minnesota (MN), USA

Michael Ball

Services, Kent State University at Ashtabula
Ohio USA.

ABSTRACT

Many digital databases housed on the web today are organized in ways that are problematic for systems researchers, primarily because they are prearranged for conventional, reductionistic, linear, statistically-aggregated research. To make use of such data, systems researchers need an intermediary, e-scientific framework that can translate their digital data into a "systems-oriented" format, so that this data can be modeled and analyzed from a complex systems perspective. We have designed just such an intermediary framework, called the SACS Toolkit. The SACS Toolkit helps systems researchers translate and use digital data trapped in non-useful formats through its unique systems-based ontology and methodology. In the current article, we demonstrate the utility of the SACS Toolkit by applying it to a digital case study: a web-based, community health science database we are currently researching. We begin our article with a bit of background, including a review of e-social science and, more specifically, the SACS Toolkit. Next, we provide a brief description of our digital case study and the challenges it presented us; followed by an explanation of how we used the SACS Toolkit to solve our challenges. We end with a summary of how other systems researchers working with digital data may find the SACS Toolkit useful.

INTRODUCTION

This paper demonstrates how scholars can use web-based, digital data to conduct research from a complex systems perspective by employing a new e-scientific method for modeling social systems, called the SACS Toolkit. Our paper is organized as follows. We begin with a bit of background, exploring the new field of e-social science and its more important terms: digital data, ontology, cyberinfrastructure. Next we turn to a brief overview of the SACS Toolkit. Third, we discuss our digital case study—a web-based, community health science database we are researching—and the challenges it presented; followed by an explanation of how we used the SACS Toolkit to solve these challenges. We end with a summary of how other systems researchers working with web-based, digital data may find the SACS Toolkit useful.

Statement of the Problem

Digital data—housed on the internet, web or other forms of cyberinfrastructure—are everywhere. In fact, our worlds are awash in electronic data (Hine 2006). And yet, this data are not so easily collected or analyzed by systems researchers. Why? The problem is that most digital data are not organized or available in a format that is readily useable for complex systems modeling or analysis (Hine 2006). Instead, most data is arranged to conduct conventional, reductionistic, linear, statistically-aggregated research (Abbott 2001).

The gap between web-based, digital data and systems research constitutes the type of challenge e-social science was created to address. The challenge is to create intermediary tools that systems researchers can employ to collect, analyze and model digital data (Castellani and Hafferty 2009).

One such intermediary tool is the Sociology and Complexity Science (SACS) Toolkit (Castellani and Hafferty 2009). The SACS Toolkit provides researchers a new systems-based ontology and methodology for collecting, organizing, analyzing and modeling digital data, in particular the large, multi-dimensional databases regularly encountered on the web today. The SACS Toolkit does this by functioning as an intermediary between the web and researcher. Its intermediary function provides researchers two major advantages.

In terms of ontology, the SACS Toolkit provides a systems-based filing system (social complexity theory) that helps researchers convert and organize digital databases in a theoretically meaningful manner. The filing system is designed also to form a complex system—to match the complexity of most web-based data.

In terms of method, the SACS Toolkit provides a novel algorithm (assemblage) researchers can use to model complex systems with web-based, digital data. The assemblage algorithm works with any type of digital data; and can be used with most methodological techniques (e.g., field research, statistics, etc), including the latest advances in agent-based modeling, network analysis, e-science and web science. Before we explore further the SACS Toolkit, we need to tour briefly the field of e-science.

E-Science: A Brief Overview

E-science is a new area of study (emerging in the late 1990s) that seeks to develop and employ the latest advances in cyberinfrastructure to help scholars make the most of doing research in a digital world (Hine 2006). John Taylor, who coined the term, specifically defines e-science as: "global collaboration in key areas of science and the next generation of infrastructure that will enable it" (www.e-science.clrc.ac.uk/). Examples of e-science abound, from providing researchers and industry access to distributed computer systems to techniques for visualizing scholarly citation networks to virtual communities where researchers can share and work on databases together. One of the leading programs in e-science is the *UK e-Science Programme* (www.rcuk.ac.uk/escience/default.htm).

A subfield of e-science is called e-social science. Its purpose is to use cyber-infrastructure to develop social scientific inquiry in the digital age (Borgman 2007). One of the leading centers is the *National Centre for e-Social Science* (www.esrcsocietytoday.ac.uk/ESRCInfoCentre/index. aspx). As Paul Tennent explains, e-social science performs an intermediary function (www.ncess.ac.uk/events/conference/programme/fri/3dtennent.pdf). Its goal is to act as a go-between, interpreter, integrator, liaison, conciliator and link between the fields of social and computer/information science. Such a role is not easy. It requires more than social scientists using computers, the grid, cyberinfrastructure or computational thinking. And, it requires more than computer scientists making new tools for social scientist to use. Instead, it requires a role that is more ontological and translational, involving itself in the iterative process of connecting data, computers and people in efficient and effective ways that promote (rather than hinder) scientific knowledge and innovation.

The intermediary work of the e-social scientist revolves around four main interconnected areas: digital data, cyberinfrastructure, ontology and method.

Digital Data: Work on digital data has to do with issues of size and complexity. When scholars use the term digital data (or any of its synonyms, such as web-based data, digital databases, or grid-data), they are referring to the databases typically encountered on the web, internet or grid. The defining feature of digital databases are their complexity: they are most often comprised of a large number of cases, factors, relationships, levels of analysis, types of data, and are often collected across time; and, in some cases, real time, as with economic data. It is also often the case that this data are located on different servers, in different formats, and tend to require different methods of retrieval. Finally, and very important to our paper, when digital databases are created, they are typically assembled according to an ontological system of classification or organization that is not always user-friendly for social scientists conducting research (Borgman 2007). By user-friendly we mean that the data is not in a format that promotes or facilitates data collection, management, analysis or modeling.

Cyberinfrastructure: Work on cyberinfrastructure has to do with what, how and where digital data is housed (Hine 2006). Cyberinfrastructure (and its related terms, such as the grid) refers to any and all research environments designed to support advanced data acquisition, storage, management, integration, mining, visualization and other computing and information processing services over the Internet or web (Borgman 2007). For an excellent overview, visit the National Science Foundation's *Cyberinfrastructure Vision for 21st Century* Discovery (www.nsf.gov/pubs/2007/nsf0728/index.jsp). Concerns related to cyberinfrastructure include such questions as: Are data accessible to the right people, institutions, etc? Are data in the right format? Are data sufficiently compatible for multiple methods of analysis, etc? And, can servers, markup and script languages and search engines interface with one another, etc?

Ontology: The third area of e-social scientific work, which is strongly tied to issues of data and cyberinfrastructure, is ontology (www.shirky.com/writings/ontology_overrated.html). Ontology concerns itself with the underlying conceptual framework upon which digital databases and their supporting cyberinfrastructure are organized. Information science employs ontology in a very distinct way; one that does not match with the traditional philosophical usage of this term. In philosophy, ontology refers to first principles and the nature of being and Being. One thinks, for example, of Heidegger, Husserl, Sartre and phenomenology in this first and most traditional sense of the term. Information science uses ontology to mean something altogether different (http://en.wikipedia.org/wiki/Ontology_(computer_science)).

For information science, ontology refers to the underlying conceptual framework upon which an information system is grounded, including: (1) what things belong within the domain of an information system (i.e., parts, groups, components, catalogues, classification schemes, servers, databases, storage retrieval mechanisms, computers, software, etc) and (2) what relationships exist amongst these things. The guiding question of ontology is: When considering the development or usage of some database and supporting cyberinfrastructure, what kind of framework or classification system will ensure that scholars, computers and data are connected in the most efficient and effective manner?"

Digital Ontology: The specific type of ontology we address in this paper is digital ontology. Digital ontology (a term we have coined) refers to any electronic classification scheme used to determine what set of things belong to a particular information system and the relationships that exist amongst those things. Examples of digital ontology abound. They include: (1) online catalogues, such as *WorldCat;* (2) search engines, such as *Google* and *Yahoo*; (3) cross-platform markup languages, such as HTML; (4) scripting languages, such as JavaScript; and (5) the numerous techniques from the burgeoning field of e-social science—which includes *Access Grid*, *Map Tube*, etc. (For more examples, see the National e-Science Centre www.nesc.ac.uk/index.html.)

Methodology: The final area of e-social science is method. In terms of methodological innovation, the focus of e-social scientists is the same as scholars in the fields of complexity science and data mining: the focus is to develop the computationally-based tools social scientists need to study the massive, multi-dimensional, multi-platform, complex databases regularly housed and analyzed on the web today.

And why are such tools needed? They are needed because the conventional methods of social science, both qualitative and statistical, were not designed for digital inquiry (Abbott 2000, p. 98). Statistics, for example, unnecessarily reduces the complexity of digital data through its employment of a linear, nomothetic approach to research—identify one or three important social factors and examine their linear impact on some set of dependent variables. As Abbott explains, such an approach is "useless for large-scale pattern-recognition" (2000, p. 298). Therefore, argues Abbott, social scientists "are going to have to jettison the idea of causality that has led us to denigrate precisely the analytic tools necessary to address the problems of huge data sets" (2000, pp. 298-299). And it is not just a matter of "ramping up" statistics to fix the problem. New tools are needed. The same is true of qualitative method, which has done almost nothing to develop innovative ways of analyzing digital data or large databases, numerical or otherwise (Castellani, Castellani and Spray 2002).

E-social scientists have their work cut out for them, particularly when it comes to training social scientists to use these tools. At present, little has changed in the way undergraduates and graduates in the social sciences are taught method. Courses focus on statistics, augmented (maybe) with qualitative inquiry. Little is really offered in advanced modeling, let alone computational modeling—and forget about offering courses that seek to move past the horribly boring yet oddly entrenched divide between qualitative method and statistics. Of the four areas of work in which e-social scientists are involved, this last one presents the greatest challenge.

THE SACS TOOLKIT

While e-social science has made many important advances in the few short years during which it has existed, much remains to be done. One particular area is the development of intermediary toolkits that systems researchers—be they in sociocybernetics, social systems theory, or complexity science—can use to model complex social systems with digital data. Serving such a function is one of the reasons we created the SACS Toolkit (Castellani and Hafferty 2009).

The SACS Toolkit is a new framework for modeling complex social systems. SACS stands for sociology and complexity science. The SACS Toolkit is part of the burgeoning literature in complexity science and e-social science (See Castellani and Hafferty 2009).

The SACS Toolkit was created, in part, to address the increasing theoretical and methodological struggles associated with digital data; in particular, the massive, multi-dimensional, complex data and databases typically found on the web. The SACS Toolkit can handle digital data because of its unique, systems-based, ontological and methodological organization. In what follows, we explore the ontological and methodological strengths of the SACS Toolkit for using, organizing and analyzing digital data. A caveat, however, is in order. Given the focus of our paper, we cannot provide a comprehensive or thoroughgoing review of the SACS Toolkit. A complete review is found in *Sociology and Complexity Science: A New Field of Inquiry* (Castellani and Hafferty 2009).

The Working Parts of the SACS Toolkit

The SACS Toolkit is comprised of three basic parts:

1. A systems-based, ontological and theoretical framework (including related vocabulary) researchers can use to organize their analysis of digital data. This framework is called social complexity theory.

2. A theoretically and ontologically grounded algorithm, called assemblage, which researchers can use to analyze and assemble, from the "ground up," a working model of a social system using web-based data. Assemblage is highly visual, relying upon a rather extensive repertoire of techniques taken from social network analysis, the new science of networks, social simulation, fractal geometry, cluster analysis, grounded theory, and the self-organizing map literature. Integrating these techniques, the SACS Toolkit provides a novel approach to visualizing social systems.

3. A recommended toolset of techniques and methods for modeling with digital data. While the SACS Toolkit can be used with just about any sociological method or technique, our work finds the following techniques indispensible when it comes to analyzing digital data: cluster analysis, neural networking (specifically, the self-organizing map), social network analysis, grounded theory method, Foucault's genealogical method, fractal geometry, chaos theory, computational modeling, and data mining.

Social Complexity Theory

As shown in Figure 1 (see below), social complexity theory is an ontological and conceptual framework for modeling complex systems using various types of data—in particular, web-based, digital data. As a framework, social complexity theory is less interested in explaining things and more interested in providing researchers an effective way to organize, coordinate, categorize, sort, connect, link and manage their data. It does this by providing researchers a theoretical filing system and an associated vocabulary that they can use to create their own model of a social system. Social complexity theory's user-driven filing system is comprised of five organizational folders. In terms of ontology, the most important is the first, the field of relations.

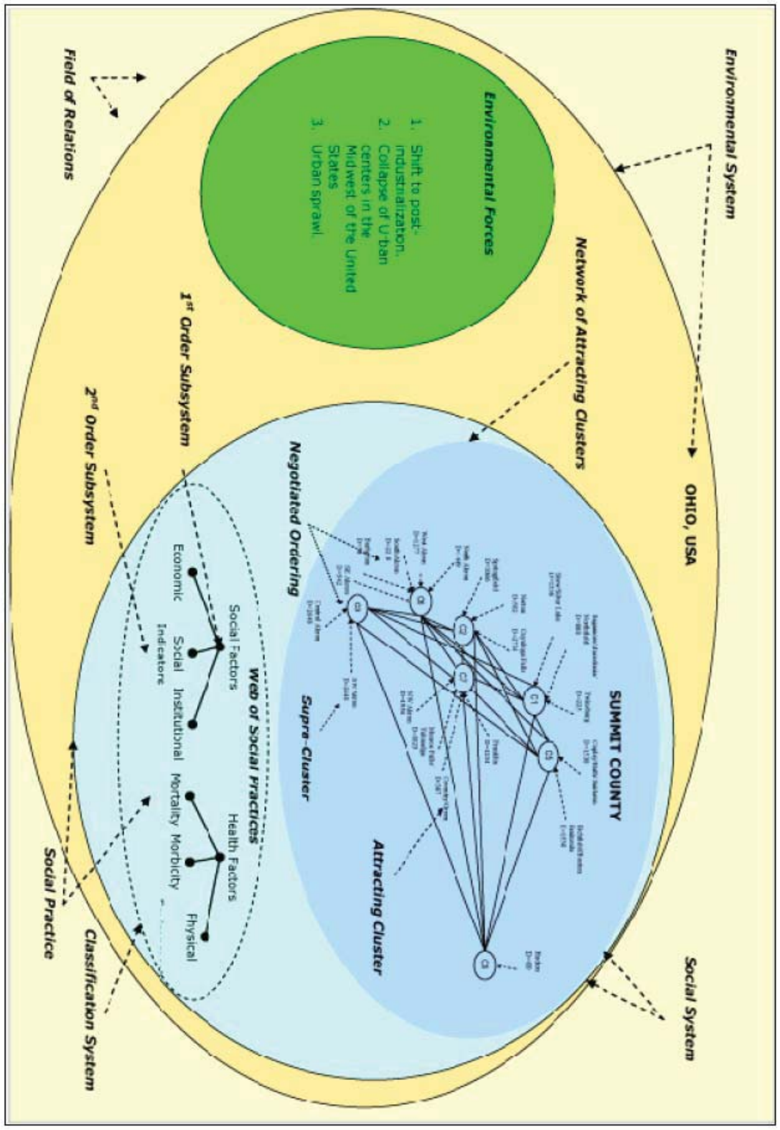


FIGURE 1: Map of Summit County using the SACS Toolkit

Field of Relations:

As shown in Figure 1 above, the field of relations is the intellectual arrangement and bracketing of all information necessary to construct a model of a complex social system. We borrow the term from Michel Foucault (Dreyfus & Rabinow 1983). For us, this term has three ontological functions: conceptual, organizational and methodological.

1. Conceptually, the field of relations functions as the grid of analysis (Dreyfus & Rabinow, 1983, pp 118-125). Its purpose is to articulate the domain in which all the elements of a social system of study, and their relationships, can be located and coaxed into coming together. What makes the field of relations unique in terms of e-science is that it is highly flexible and user-driven. The field of relations is flexible because it changes according to the topic of study; and it is user-driven because the researcher defines what the domain of relations will be, what to include within it and what relationships are the most important.

The user-driven nature of the field of relations is very important. The SACS Toolkit's utility comes from its ability to act as a supra-ontological framework, which can be placed upon any existing framework found on the web. As such, regardless of the databases, servers, data formats being used, the researcher has a guiding ontological framework that helps to organize, analyze and model some topic of study in complex systems terms.

2. In terms of digital data, the second ontological purpose of the field of relations is organizational. Social complexity theory is a rigorous framework of classification. Social complexity theory provides a way for researchers to make sense of the chaos of digital data, which it does by giving the researcher a set of conceptual folders, sub-folders, a filing system, and so forth for organizing everything in a set of predetermined format—see Figure 1.

3. The third ontological purpose of the field of relations is methodological. The strength of using the field of relations is that it can be directly applied to the management of one's database, as well as the analysis of empirical data. This is of particular importance when working with digital data because there is no loss of information as the researcher moves from theory to data collection to analysis.

The Other Four Folders:

As shown in Figure 1, what makes social complexity theory so rigorous and yet flexible when it comes to organizing digital data is that its filing system is designed to form a complex social system. Said more specifically, the four major folders within the field of relations—(1) the web of subsystems, (2) the network of attracting clusters, (3) environment, and (4) system dynamics—represent each of the major domains of a complex social system. In turn, each of these folders is comprised of their own sub-folders and files. Furthermore, this system of folders comes with a corresponding vocabulary, including such concepts as attractor points, negotiated ordering, system trajectory, social practice, emergence, and self-organization (Castellani and Hafferty 2009). By using this filing system and vocabulary, researchers can empirically investigate the structure and dynamics of a complex social system, confident that they have an effective way to manage their data and their study, as it develops over time.

Assemblage:

Assemblage is a case-based, system-clustering algorithm for modeling social systems. It is built on the organizational framework of social complexity theory and represents the procedural component of the SACS Toolkit. As shown in Figure 2 below, the goal of assemblage is to move researchers through a six-step algorithm for constructing a model of some social system of study. This algorithm roughly proceeds as follows:

STEP 1: Help the researcher define a set of research questions in systems terms.

STEPS 2-4: Establish the social system's field of relations and begin to "file and fill-in" the information for all of the major folders (web of social practices, network of attracting clusters, etc). Examine the internal structure and dynamics of the model for a particular moment in time-space—a snapshot of the model, if you will—including its interactions with key environmental forces and its trajectory within key environmental systems. Assemble these discrete, cross-sectional snapshots of the system into a moving model, providing some overall sense of the system as a whole.

STEPS 5-6: Once done, researchers can "data mine" this model to answer the initial study questions or to generate new questions or models.

The Uniqueness of Assemblage:

As a set of procedures, assemblage has seven key features which, when combined, make it unique amongst e-social science and complexity science methods. This is not to say that some of the features of assemblage (such as its case-based approach to analysis) are not found in other methods and techniques. However, it is to say that no other method has all seven features. We will briefly review these features here. See Castellani and Hafferty (2009) for more information.

1. Assemblage was designed to address the unique challenges associated with modeling complex social systems.

2. Assemblage is ontologically grounded in social complexity theory. Few methods in e-social science or complexity science come with their own systems-based ontology. Assemblage does.

3. Assemblage has no data preference. Unlike the majority of e-social science or complexity science methods, which tend to focus on numerical data, assemblage works equally well with any and all data types—from numerical to visual to historical.

4. Assemblage works with just about any statistical, qualitative, historical or computational technique. As such, it works with (rather than against) the existing repertoire of a researcher, rather than unnecessarily pushing the researcher into new techniques. This is an important point because, to date, the methods of e-social science and complexity science are, for the most part, computationally based. The reason assemblage can be used with such a wide variety of tools and toolsets is because these tools do not drive the model building process. Instead, the six-step algorithm of assemblage, along with the theoretical framework upon which it is grounded, drives model building. Any tool can be used as long as the researcher uses it in service of modeling a social system.

5. Assemblage employs a case-based, constant comparative approach to modeling complex social systems. Following Ragin (2008) and colleagues (e.g., Ragin and Byrne 2009), we find that the best way to preserve the complexity of any system of study and to make sense of this complexity at the same time is to adopt a case-based approach to analysis.

A case-based, constant comparative approach to digital data treats a social system as a set of cases, each of which represents one of the multiple ways that a complex social system is practiced by the agents of which it is comprised. An easy example: in a political system, the various values of its agents will couple together to form different political parties (e.g., conservative, moderate, liberal, etc). These political parties will, in turn, have their own coupled divisions: liberal conservatives, moderate conservatives, etc.

From the perspective of social complexity theory, a case represents each and every way a complex system of study can be practiced. In other words, a case represents one example, expression, instance or illustration of a social system of study. A "case-based" approach is useful because it allow us to build a social system from the ground-up, by exploring and comparing cases, one or several at a time, to profile and catalogue the various ways that a web of social practices is expressed. Once this process is complete, the researcher is ready to move to the next major step in the assemblage process.

6. Assemblage is a data-compressing, system-clustering method. The ultimate goal of assemblage is to help the researcher cluster the social system into its key attractor points. In this way—and here we draw directly from Kohonen (2001) and his self-organizing map technique—assemblage is a data reduction technique. Assemblage tries to reduce and compress the complexity of a social system into a simpler and more understandable form. The product

of this simplifying process is the network of attracting clusters. As shown in Figure 1, the network of attracting clusters looks like a typical network, except it is organized around the dominant ways a system in practiced—otherwise known as the system's attracting clusters. Once this network of attracting clusters has been created, it is then reconstructed over a series of discrete moments in time-space and put together to create a moving picture of the system's dynamics, along with its trajectories within various environmental systems. If greater detail is needed, this can be done post hoc. Or, if one wants a more complete picture, one can "drill down" (to use a data mining term) into a particular cluster to construct a more refined and focused map of a particular section of some social system of study.

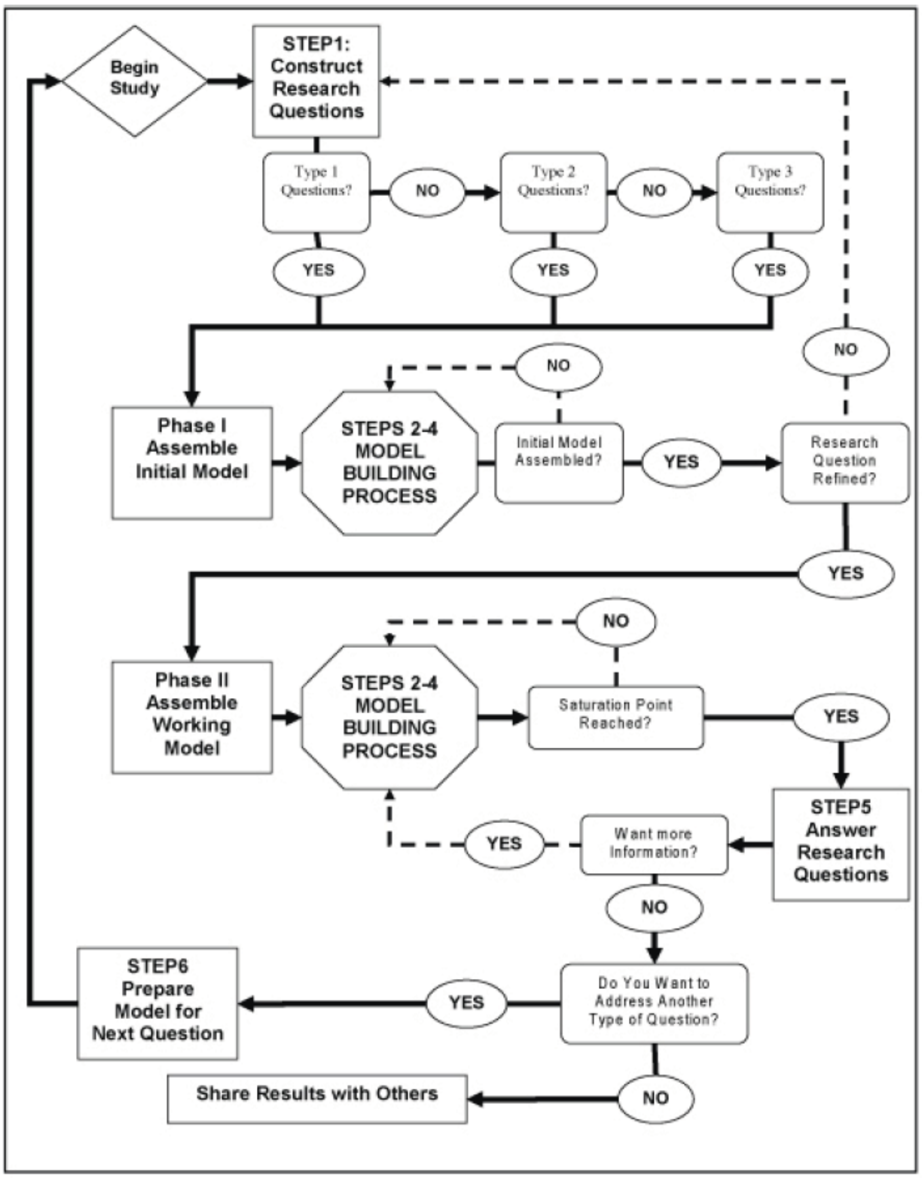


FIGURE 2: The Assemblage Algorithm

7. Finally, assemblage provides a novel approach to visualizing social systems. As a data compression technique, the goal of assemblage is to help the researcher create a low-

dimensional picture of high-dimensional data (Castellani and Hafferty 2009). For examples, see Figure 1 and Figure 7.

Summit County 2010: A Case Study

At this point we have accomplished two of our three goals. We have provided a basic overview of e-social science and we have outlined the SACS Toolkit. Our third and final goal is application. Below, we provide an example where we used the unique, systems-based ontology and methodology of the SACS Toolkit to solve a particular challenge we had with a specific set of web-based, digital data. The digital data in question is the *Healthy Summit 2010 Quality of Life Project*. But, before we get to the website, we need a bit of background on our case study.

Communities as Complex Systems

Our case study involves a county in Northeastern Ohio, USA we have been studying for the past two years, called *Summit County*. Our goal has been to understand how the 20 communities of Summit County function as a complex system and the impact this county-level system has had on the health of its various communities. Our case study is grounded in the community health science literature (Cummins, Curtis, Diez-Roux and Macintyre 2007; Curtis and Riva 2009; Robert 1999).

Over the last several years, a major shift has taken place in the community health science literature. The conventional, simplistic, reductionistic, statistically-driven view of communities as "little more than context" is being replaced by a more spatially and conceptually complex view (Cummins, Curtis, Diez-Roux and Macintyre 2007; Curtis and Riva 2009). In this new view, communities are thought of in holistic or systems terms and are seen as complex, emergent entities. This view also holds that communities function at multiple levels of scale; they operate with open-ended boundaries; they are fluid, mobile and evolving; they are not constrained by traditional notions of space and time; they are comprised of nonlinear feedback loops and causal pathways; they have histories and multiple social meanings; they emerge out of the intersection of the micro and macro, the local and global, and agency and structure; and they are nodes in a larger network of places and environmental forces. In short, communities are complex systems (Blackman 2007; Curtis and Rivera 2009a, 2009b; Gatrell 2005).

The Data Challenge:

While some of the leading scholars in the community health science literature are making the complexity turn (see Cummins, Curtis, Diez-Roux and Macintyre 2007), most public and community health science databases, particularly those housed on the web, have not. The e-scientific gap between these complexity-thinking scholars and most of the current digital data is due to two major reasons. First, the majority of scholars in community health science (particularly those working in conventional public health facilities throughout the United States, where a significant amount of public health data is housed) do not yet endorse a complexity science view of communities (Curtis and Rivera 2009a, 2009b; Gatrell 2005). Second, as with any new ontological framework for organizing our understanding of data, there is a lag between the new idea and supporting cyberinfrastructure. And so, when it comes to the community health science literature, an e-scientific gap exists that breaks down, rather than facilitates systems-oriented research. It is this same e-scientific gap that exists in the web-based, digital database we are using for our study of Summit County. Let us explain.

The Case of Summit County 2010

The database for our study is entirely web-based. It was put together by the Summit County Combined Health District. The website is the *Healthy Summit 2010 Quality of Life Project*, which we will abbreviate as Summit 2010 for the rest of this paper (www.healthysummit.org/). It was designed to bring together the activities, concerns, data, and research agendas of all the health providers in Summit County, including its various public health centers. In terms of our research, we chose the Summit 2010 website because of the wealth of data it provides. The database is organized into two major types of data: reports and maps (See Figures 3 and 4 below respectively).

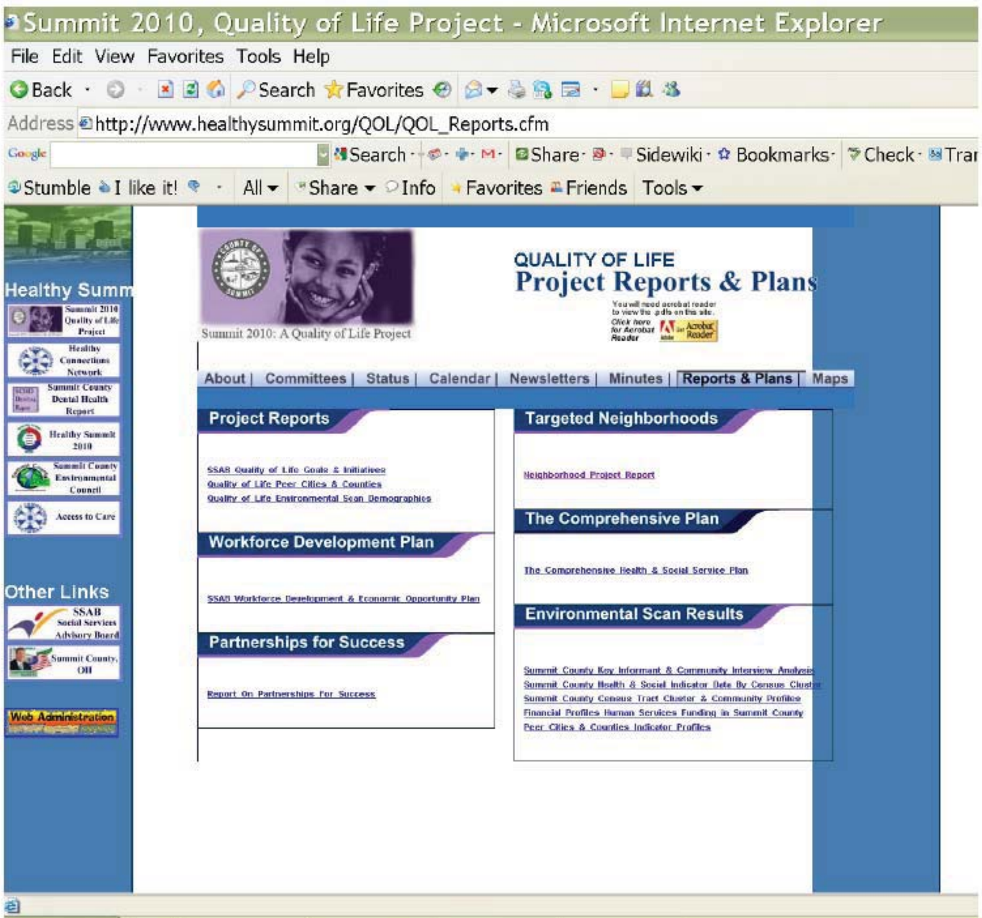


FIGURE 3: Summit 2010 Website

Reports: As seen in Figure 3, the reports found on the Summit 2010 website provide numerous types of data, including: (1) listings of all the health agencies in Summit County; (2) historical narratives; (3) in-depth neighborhood studies of three of the poorest communities in Summit County; and (4) statistical summaries of the county as a whole, including a long list of economic (e.g., household income, job growth, etc), institutional (e.g., immunizations, education levels, etc), and health outcome indicators (e.g., mortality rates, morbidity rates, etc).

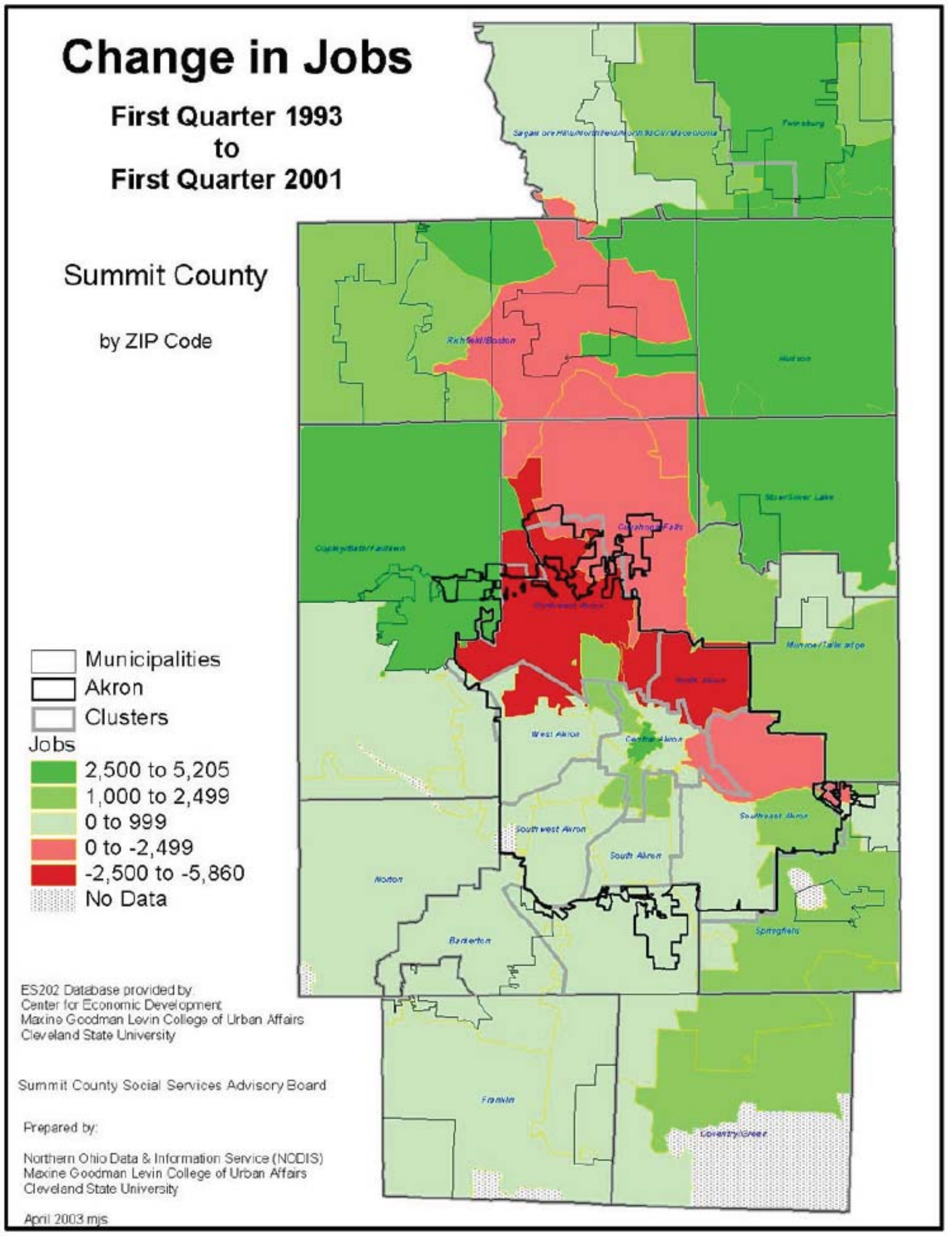


FIGURE 4: Job Growth Map of Summit County

Maps: As seen in Figure 4—which is one of 25 maps—the maps on the Summit 2010 website show how various social and health factors are spatially distributed across the 20 major census clusters in Summit County. Together, these maps provide a detailed overview of the economic and health inequalities that exist within Summit County.

The Ontological, E-Scientific Gap of Summit 2010

The e-scientific gap created between our systems approach to community health and the Summit 2010 website is ontological in nature. The Summit 2010 website is organized according to a conventional, simplistic, reductionistic and largely statistically-driven view of communities. Here are the specific issues with the website we had to address:

1. While the Summit 2010 website contains some qualitative reports and interviews with local residents, the majority of its reports are quantitative in nature.

2. Second, as shown in Figure 5 (see below), all the statistical reports are descriptive and summative in nature. Furthermore, these summaries are all in PDF form.

3. Third, none of the reports examine the relationships amongst communities, or how changes in one community or the county affect other communities. Instead, they examine communities individually, or they examine the county as a whole.

4. Fourth, while most of the statistical data was collected at two major points in time (early and late 1990s), changes are examined in broad strokes, as *trends* across the county or specific communities. None of the reports examine how the county or its communities changed over time together, or what influence their mutual, interdependent change might have had on one another.

5. Finally, Summit 2010 does not allow access to the databases used to generate its reports.

Priority Indicator #1: **Overall poverty**

Defined as: The number of persons living within the census tract cluster borders where those borders are defined by United States Census Bureau census tracts who live "below the poverty level" as defined by the U.S. Census. Following the Office of Management and Budget's (OMB's) Directive 14, the Census Bureau uses a set of money income thresholds that vary by family size and composition to detect who is poor. If the total income for a family or unrelated individual falls below the relevant poverty threshold, then the family or unrelated individual is classified as being "below the poverty level."

Source of Data: United States Census Bureau 1990 and 2000 Decennial Censes.

Cluster	1990		2000		Change	
	#	%	#	%	#	%
Sagamore / Macedonia / Northfield	567	2.7	634	2.3	67	11.8
Twinsburg	774	5.7	702	3.3	(-) 72	(-) 9.3
Richfield / Boston / Peninsula	381	5.1	298	3.6	(-) 83	(-) 21.8
Hudson	168	1.0	380	1.7	212	126.2
Copley / Bath / Fairlawn	679	2.6	951	3.0	272	40.1
Cuyahoga Falls	3,140	6.5	3,052	6.1	(-) 88	(-) 2.8
Stow / Silver Lake	926	3.0	1,371	3.9	445	48.1
Northwest Akron	2,770	9.6	3,234	10.7	462	16.7
Munroe Falls / Tallmadge	1,007	5.0	860	4.0	(-) 147	(-) 14.6
North Akron	2,919	15.4	2,107	11.8	(-) 812	(-) 27.8
West Akron	6,729	18.7	5,347	15.7	(-) 1,382	(-) 20.5
Central Akron	7,621	44.3	6,804	34.8	(-) 817	(-) 10.7
Southwest Akron	9,501	26.8	7,617	23.0	(-) 1,884	(-) 19.8
South Akron	5,526	20.8	4,307	16.8	(-) 1,219	(-) 22.1
Southeast Akron	9,555	17.5	8,921	16.2	(-) 634	(-) 6.6
Norton	606	5.4	665	5.8	59	9.7
Barberton	4,649	16.6	3,721	13.1	(-) 928	(-) 20.0
Springfield	1,575	8.2	1,292	6.5	(-) 283	(-) 18.0
Franklin	726	4.5	653	4.1	(-) 73	(-) 10.1
Coventry / Green	1,672	5.6	1,760	5.2	88	5.2
Summit County	61,491	12.1	52,991	9.9	(-) 8,500	-13.8
United States	31,742,864	13.1	33,899,812	12.4	2,156,948	6.8

Table created by: Thomas Quade, MA, MPH, Epidemiologist, Akron City Health Department, Akron, Ohio 44308 quadeto@ci.akron.oh.us

FIGURE 5: Typical Style of Presentation in Summit 2010 Statistical Reports

Applying the SACS Toolkit:

Let us briefly summarize what we so far know. As systems researchers, our work for the last two years has sought to study community health from a complexity science perspective (Castellani and Hafferty 2009). The case study upon which our research is based is Summit County, Ohio, USA. We treat this county as a complex system comprised of 20 communities. Our database for this study is the Summit 2010 website. While this database is rich in detail, it is organized according to a conventional, non-systems view of communities—as detailed in our above five points. Our challenge, therefore, was to translate this website's data into a format for doing systems research. To solve our challenge, we employed the SACS Toolkit, a new framework for modeling complex systems. The strength of the SACS Toolkit is its user-oriented, systems-based ontology and methodology.

In what follows, we outline how we used the SACS Toolkit to solve the ontological challenge the Summit 2010 website provided us. It is important to note, however, that our goal here is not to provide a detailed, step-by-step account of our research procedure. That is forthcoming (Castellani, Buckwalter, Hafferty & Ball forthcoming). Instead, our goal is to highlight our research process, sufficient for readers to see the potential of the SACS Toolkit.

Steps to Solve Our Ontological Challenge:

1. Employing the first step in the assemblage algorithm (See Figure 2), we began by constructing a map-based understanding of Summit County and its 20 communities. Remember that, according to assemblage, the first step in the research process is to formulate a series of systems-based questions and to construct a preliminary model of one's topic as a complex system.

We therefore began with the maps because (while not intended by the creators of the website) they provide a *systems* view of Summit County. We chose Figure 4, a map of job growth in Summit County. Following assemblage (and its case-base, bottom-up approach to model building), we made this map our first case. What was great about this first case is that, distinct from the community-by-community report on job growth provided on the Summit 2010 website (which is formatted similar to Figure 5), Figure 4 gave us an immediate, systems-based, spatially arranged understanding of job growth and the lack thereof in Summit County.

Looking at the middle of the map in Figure 4, for example, one sees Akron, the major city in Summit County (population 217,000). This city, which has been hit hard by post-industrialization, has had trouble generating new jobs. This lack of job growth is seen primarily in the communities just north of the City's downtown. Conversely, one sees that in some of the suburban communities surrounding north Akron (as well as those in the northeast corner of Summit County), major job growth has taken place. In other words, the pattern is not random; instead things seem to cluster together into system-wide patterns. The immediate question is why?

We cannot stray too far in discussing this question, but it has a lot to do with the out-migration of the middle-class and affluent residents of Akron to the suburbs. What we can, however, spend time discussing is the way our examination of these maps helped us build a systems understanding of Summit County.

Going back to Figure 4, by studying this map we gained a holistic view of job growth as a dynamic dimension of the system of Summit County; changing over time, situated within and across communities. Furthermore, we were immediately challenged to consider what underlying social forces might account for this system-wide difference in job growth. Is it out-migration? Is it the emergence of poverty traps? As we worked to develop our questions, we moved our focus on the map back and forth between the local (individual communities) and the global (the county), attempting to more accurately formulate our questions. This type of back and forth movement is exactly what the SAC Toolkit is designed to help researchers, at this initial stage, do. Only by purposely engaging the data in such a bottom-up, theoretically grounded manner (research questions and preliminary model being built simultaneously) can one quickly obtain a systems view, regardless of the data's particular ontological format.

2. Our next step was to construct a preliminary model of Summit County. To construct this model we continued analyzing, from the ground-up, the website's maps. To organize our mounting visual data—there are 25 maps on the website—we turned to the SACS Toolkit filing system, via social complexity theory and its major folders: field of relations, environment, etc. Using this filing system and its folders, we generated the model shown in Figure 1 (See above).

Looking at Figure 1, one can see all the major folders of social complexity theory. For example, there is the *environmental folder*, which contains our ideas about the major forces impacting Summit County and its health. There is the *web of social practices folder*, which outlines the major factors out of which our 20 communities emerge. And, there is the *network of attracting clusters*, the structure of which is displayed as a network—see Figure 7 for a magnified version of this network. As we hope Figure 1 helps to demonstrate, by following the SACS Toolkit's algorithm we were able to construct, from the bottom-up, our own systems-based, ontological framework for our study. With this framework developed, we were able to organize the rest of website's data, particularly the statistical data, and construct a grounded-theoretical model of Summit County as a complex system.

3. With our research questions and preliminary model created, we needed to prepare for the second phase of the assemblage algorithm: constructing a working model of Summit County. To construct our working model, however, we needed to move from the maps on the Summit 2010 website to the statistical reports. The result of this move from the maps to the statistical data resulted in the database shown in Figure 6 (See below). To construct this database, we went through all the reports, copying and pasting information about each of the 20 communities into an SPSS database. (Figure 5, shown previously, shows a page from one of the reports we used to construct our database.) The result was a vector matrix. As shown in Figure 6, reading from left to right, the list of variables for each of our 20 community vectors included all of the various economic, social and health outcomes indicators discussed in the various reports, from household income to teenage pregnancies to educational levels to mortality rates to job growth rates, at two major points in time: the early 1990s and 2000.

Challenge Solved:

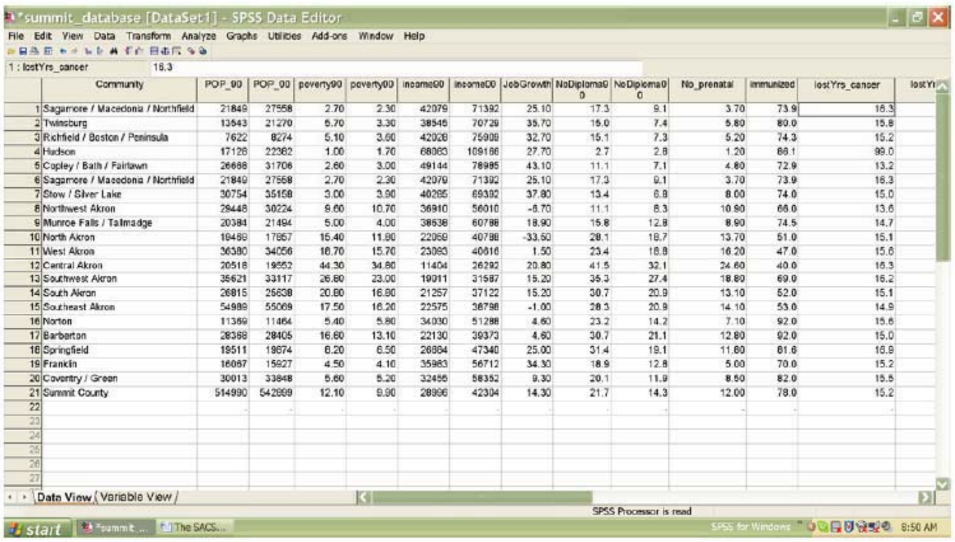
At this point in the research process, we had solved the major ontological challenge the Summit 2010 website had presented us. First, by following the assemblage algorithm—which requires researchers to (a) develop a preliminary model of their topic as a complex system and (b) construct the initial systems-based questions guiding their study—we quickly found a systems perspective of Summit County through its maps. Second, to re-organize our map-based data according to a systems-based ontological framework, we employed the filing system and folders of social complexity theory. With this preliminary model developed, we were able to build our numerical database, with data for two points in time: early 1990s and 2000.

Phase 2 – Method:

With our preliminary model and database complete, we were ready to move to the second phase of our study. In this phase, the goal is to use the various complexity-science methods employed by the SACS Toolkit to construct a working model of Summit County as a complex system.

It is at this phase in the research process, however, that we come to the end of the current article. Nonetheless, while we cannot go into detail about all the various methods we used in our study, we do want to leave the reader with a visual compare and contrast.

Because of the database we constructed, we were able to employ a variety of complexity science techniques that we would not have otherwise used on the Summit 2010 digital data. Such techniques included k-means cluster analysis, the self-organizing map algorithm (a neural net technique for data compression, clustering and visualization), agent-based modeling (specifically the cellular automata), network analysis and the qualitative complexity method, qualitative comparative analysis (QCA). (For more on QCA, see Ragin 2008; Ragin & Byrne 2009.)



	Community	POP_90	POP_00	poverty90	poverty00	income90	income00	JobGrowth	NoDiploma90	NoDiploma00	No_prenatal	immunized	lostYrs_cancer
1	Sagamore / Macedonia / Northfield	21849	27558	2.70	2.30	42079	71392	25.10	17.3	9.1	3.70	73.9	16.3
2	Twinsburg	13543	21270	6.70	3.30	38546	70729	35.70	16.0	7.4	5.80	80.0	15.8
3	Richfield / Boston / Peninsula	7622	8274	5.10	3.60	42028	75909	32.70	15.1	7.3	5.20	74.3	15.2
4	Hudson	17128	22382	1.00	1.70	68063	109166	27.70	2.7	2.8	1.20	88.1	99.0
5	Copley / Bath / Fairlawn	26668	31706	2.60	3.00	49144	78985	43.10	11.1	7.1	4.80	72.9	13.2
6	Sagamore / Macedonia / Northfield	21849	27558	2.70	2.30	42079	71392	25.10	17.3	9.1	3.70	73.9	16.3
7	Stow / Silver Lake	30754	35158	3.00	3.90	40285	69392	37.80	13.4	6.8	8.00	74.0	15.0
8	Northwest Akron	29448	30224	9.60	10.70	36910	56010	-6.70	11.1	8.3	10.90	66.0	13.6
9	Munroe Falls / Tallmadge	20384	21494	5.00	4.00	38638	60786	18.90	15.8	12.8	8.90	74.5	14.7
10	North Akron	19469	17857	15.40	11.80	22069	40788	-33.60	28.1	18.7	13.70	51.0	15.1
11	West Akron	36380	34056	18.70	15.70	23083	40616	1.50	23.4	18.8	16.20	47.0	15.6
12	Central Akron	20518	19652	44.30	34.80	11404	26292	20.80	41.5	32.1	24.60	40.0	16.3
13	Southwest Akron	35621	33117	26.60	23.00	19011	31587	15.20	35.3	27.4	18.80	69.0	16.2
14	South Akron	26815	25638	20.80	16.80	21257	37122	15.20	30.7	20.9	13.10	52.0	15.1
15	Southeast Akron	54989	55069	17.50	16.20	22575	38798	-1.00	28.3	20.9	14.10	53.0	14.9
16	Norton	11369	11464	5.40	5.80	34030	51288	4.60	23.2	14.2	7.10	92.0	15.6
17	Barberton	28368	28405	16.60	13.10	22130	39373	4.60	30.7	21.1	12.80	92.0	15.0
18	Springfield	19511	19674	8.20	6.50	26884	47340	25.00	31.4	19.1	11.80	81.6	16.9
19	Franklin	16067	15927	4.50	4.10	35983	56712	34.30	18.9	12.8	5.00	70.0	15.2
20	Coventry / Green	30013	33848	5.60	5.20	32456	58352	9.30	20.1	11.9	8.50	82.0	15.5
21	Summit County	514990	542899	12.10	9.90	28996	42304	14.30	21.7	14.3	12.00	78.0	15.2

FIGURE 6: SPSS Database Constructed for Working Model of Summit County

The result of all this analysis was a rather detailed systems model of Summit County. Figure 7, for example (see below), is a network analysis of the 20 communities in Summit County, based on their relative wealth and wellbeing. If the reader compares this systems-based map to Figure 5, which is how the statistical data is presented on the Summit 2010 website, one can see that the SACS Toolkit allowed us to go very far beyond the data. We also hope that this comparison shows readers the e-scientific strength and utility of the SACS Toolkit for doing systems research with digital, web-based data.

Conclusion

In this paper, we have explored how the SACS Toolkit functions as an effective e-social science method for modeling complex social systems using digital data. It is effective because of its unique ontological and methodological approach to modeling, which is systems-based, rigorous and yet very flexible. In the case of the community health science, for example, the SACS Toolkit allows researchers to model communities as complex social systems using digital data, which is a major advance in the literature.

As a side note, the Summit 2010 website is not the only example we have of the mediating, e-social scientific utility of the SACS Toolkit. We also used the SACS Toolkit to build a virtual map of the new science of complexity, called the Complexity Science Map (www.artsciencefactory.com/complexity-map_feb09.html). We make this side note to illustrate that the SACS Toolkit is very flexible. It can just as easily be used to create a web-based map of complexity science that connects data, computers and people in efficient and effective ways, as it can be used as a systems-based ontology and methodology for modeling complex social systems with digital data.

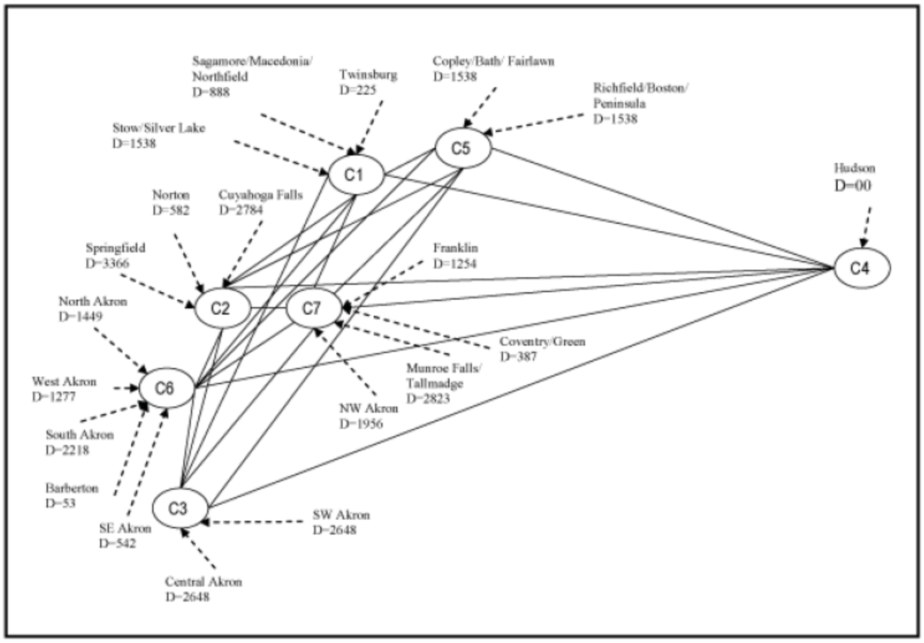


FIGURE 7: Network Model of Summit County and Its 20 Communities

What makes the SACS Toolkit so useful in both these instances is: (1) its explicit, complex systems approach; (2) its systems-based ontology; (3) its rigorous yet flexible filing system, which is designed to function as a complex system; (4) its case-based, data- compression, visual algorithm for modeling complex systems from the bottom-up; and (5) its tremendous flexibility with all types of data and methods. Given this list of attributes, researchers may find the SACS Toolkit similarly effective in other instances where they seek to model a topic as a complex social system using digital data.

REFERENCES

Abbott, A. (2001). Reflections on the future of sociology. Contemporary Sociology, 29: 296-300.

Borgman, C. (2007). Scholarship in the digital age: Information, infrastructure and the internet. Boston, MA: MIT Press.

Bowels, S., Durlauf, S., & Hoff, K. (2006). Poverty Traps. Princeton NJ: Princeton University Press.

Castellani, B., Castellani, J., & Spray, S. (2002). Grounded neural networking: modelling complex quantitative data. Symbolic Interaction 26(4): 577-589.

Castellani, B., & Hafferty, F. (2009). Sociology and Complexity Science: A new area of Inquiry. Germany: Springer.

Castellani, B., Buckwalter, JG., Hafferty, F., & Ball, M. (forthcoming). Community health: A complexity science approach.

Cummins, S., Curtis, S., Diez-Roux, A., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. Social Science and Medicine, 65: 1825-1838.
Curtis, S., & Rivera, M. (2009). Health geographies I: Complexity theory and human Health. Progress in Human Geography, 2009: 1-9.
Dreyfus, H., & Rabinow, P. (1983). Michel Foucault: Beyond Structuralism and Hermeneutics, 2ndEdition. Chicago IL: University of Chicago Press.
Gatrell, A. (2005). Complexity theory and geographies of health: A critical assessment. Social Science and Medicine, 60:2661-2671.
Hine, C. (2006). New Infrastructures for Knowledge: Understanding E-Science. Hershey, PA: Information Science Publishing.
Ragin, C. (2008). Redesigning Social Inquiry: Fuzzy Sets and Beyond. Chicago: University of Chicago Press.
Ragin, C., & Byrne, D. (2009). The SAGE Handbook of Case-based Methods. Los Angeles: SAGE.
Robert, S. (1999). Socioeconomic position and health: The independent contribution of community socioeconomic context. Annual Review of Sociology, 25: 489-516.