II Sense of Place and Sonic Topologies: Towards a Telemimetic Sublime in the Data Landscape

Christina McPhee
www.christinamcphee.net

illustrations: digital print series on fujiflex, each 19 x 66 inches, ©Christina McPhee 2003-2004

Background

Global warming implicates the increasing atmospheric level of carbon as a primary agent. Nonetheless, the total worldwide carbon budget, which takes into account all known petrochemical usage on an annual basis, shows that terrestrial systems must be absorbing more carbon than we realize. According to the carbon budget mathematical models, carbon concentrations should be increasing faster than they actually are. The hypothesis is that the carbon flux patterns of selected microsystems worldwide may reveal conditions under which more carbon is been absorbed than is being released. On and near Konza Prairie, since 1997, diurnal and annual data are collected as "eddy correlation" or "eddy covariant" flux measurements. From two of the sites, a located on the Rannels Ranch next to the Konza field station, wireless net carries the live data online for collection and analysis. Jay Ham, PhD, agronomist and climatologist, conducts research into carbon flux dynamics relative to models of climate change, at Kansas State University. He is the scientific partner for the present project, Slipstreamkonza. Slipstreamkonza addresses aesthetics of digital data expression of land as a breathing ecosystem. The time based data stream of carbon flux is interpreted as rhythmic, virtual expression of sound and image in net based and spatial installation.
I net and Gaia

Imagine interpolated virtual and actual spaces thrive and decay, die and live in a riparian zone, watered by pervasive computing as a neural territory or intelligent topology, the net acts as if alive. As a place of continuous ruin and simultaneous regeneration, the networked space of electronic communications is re-presenting, itself. A semiotic model may offer us the net as a subjective topology, a synaptic process-space. This space is not silent. Semiotically, it ‘voices’ itself. A model of the net as a live voice finds some echo in analogy to the Gaia hypothesis on the nature of the physical landscape. As life, Gaia persistently self-represents, or emits information about herself [1]. This is an old idea in new dress. “Day by day pours forth speech,” declares the Psalmist. In semiotic terms, a landscape of voice or self-expressive phenomena, as actual, real information—is both a data landscape and sonic topology. Where and what is this place? What is the sense of place in the data landscape?

ii topology and telemimesis

You might picture the structure, or topology of data streams, whether in the electronic or in the natural ecosystem, as an invisible domain that persists over, and through discontinuities. The leap across the breaks, or breakdowns, can be expressed musically by means of formal structures of recursion and feedback loops, as in classic cybernetic theory, but also as in Baroque fugue structures. I imagine recursion and flow, between natural data and human/machine, an interpolated, mutual consciousness. A topology is a word (logos) of a place (topos). A hypothesis about what constitutes this ‘word’ or voice of a place on the level of artistic process is aesthetic in nature and intent. Aesthetically, such a place may be explored as a process of telemimesis. “Telemimesis” joins tele -- vectors across distance in space, as if space is actually layered time—with mimesis, in the Platonic sense of figuration of a prescient or hidden motif.
As a visual artist, one may turn a gaze to what cannot be ‘seen’. Here we move into a zone of the sublime. Sublimity refers to that which is below, beyond or immanent relative to an ontological or cognitive threshold. I assume that there is a way of expressing this indeterminate zone, or invisible condition, in both the realms of the physical, cultural landscape and in the interior, “behind the screen” landscape of the net.

As an ecosystem, the data landscape may be described as continually subject to entropy, following the second law of thermodynamics. Life itself may be thought of arising, like a phoenix from ashes, as an articulate resistance to entropy. A continuous dialectic between entropy and the architectural self-structuring process of life means that homeostasis is predicated on breakdown, or ruin. Data stream is not always continuous. Scientific instrumentation for measurement and transmission of physical data may fail. Anomalies of landscape data are not always explicable based on known models. Humans struggle with the limitations of their bodies, including, fatigue, inattention, illness and mortality. A *telemimetic* aesthetic of the sense of place in the data landscape accommodates breakdown of the ‘language’ of information streams.

Imagine recursion and flow, between natural data and human/machine, an interpolated, mutual consciousness. [2] The place of flow is sonically expressive, or so the artist hopes. A possibility is that human synaptic pathway performs as a layer of dynamic connotation. Like a trace, or vector, over and through the data landscape, the synaptic layer is human.
Maybe we primates collaborate in a system of connotation that is never fully seen, heard or actualized: existing in time, the system is grasped in spatial fragments. Or, maybe --- between ‘natural’ morphologies in the brain/mind, and the remediation of landscape as ‘big data’, is a space, or place of telemimesis. It is mimetic in that it represents itself relative to a precessive content (landscape data) and does so at a distance (tele) from itself.

Recursion – could it be a metaphor of breathing, in and out, between the inside of the landscape-body and the outside? Within Gaia, imagine an emission of data, and a gathering-in of data. Global climatology attempts to study climate change by collecting millions of data samples of diurnal carbon absorption and release on the tall grass prairie, an ecologically critical environment in North America and elsewhere. The prairie is implicated within the phenomenon global warming in ways that are not clearly seen at present. Slipstreamkonza is an art/science collaborative project that addresses aesthetics of digital data expression of land as a breathing ecosystem and as manifestation of climate change. Slipstreamkonza uses the time based data stream of carbon flux as a basis for a sonic and telemimetic installation. Like the ancient Greek hero, Orpheus, this project attempts to make music from “Hades”, in other words, the invisible domain of data, like a voice from a hidden subject. At the sequence of delays, or layers in time, the ‘sense of place’ seems to be in a feedback loop. Things move in and out of a condition of being nameable, or heard.
vi konza and telemimesis

Konza is the Osage term for “south wind.” Like breath on a mirror, konza is an evanescent imprint of an invisible dynamic. Prairies worldwide capture and release carbon in a waveform breath. At the threshold of the exchange between atmosphere and surface is the life of the planet; the Konza prairie is a site that can be interpreted aesthetically in terms of a telemimetic topology using sonic forms.

Slipstreamkonza can’t be about conventions of scientific visualization, since the researchers already have many ways of doing this in order to better understand the data. Yet, the encounter between the human response and the landscape's self expression as data, is intriguing as a paradox of technology and/in nature. I am interested in the relationship, or dynamic, between the data and the human imagination. Data goes live as a dialectic or interface between paratopic, polyphonic, and polychromatic volumes. Think about interpolation and superimposition, like montage, as virtual and physical spaces, using layers of content that are expressions of hidden data through a semi-permeable membrane, or data transport mode. Maybe time becomes metabolic: it gives rise to a productive structure, composed of intelligent units, or affective artifacts in continuous movement and states of disclosure.
conclusion, or another beginning

A moving sense of place gathers its momentum and definition on the fly, like a continuous improvisation that is not entirely responsive to human use and reflection. A poetics of that place, both virtual and physical, in the mixed volumes of fluid media, might give rise to a polyphonic design strategy, where arching shifts between recursion and sonority, darkness and density, transparency and light, processional and volume are as responsive interactive structures in multimedia installation. Like a fold or complex cut in the fabric of the data landscape, sonic topologies become a conceptual practice realm in contemporary art.

Notes

[1] Geri Wittig has looked at the Gaia hypothesis relative to the discourse on landscape data, holism and science, and includes a brief, helpful bibliography on this topic, at <http://www.c5corp.com/research/complexsystem.shtml>.

[2] Brett Stalbaum asserts that “data's role in the instantiation of the actual may be a matter of virtual informatic interrelations (or external relations between data sets), forming their own consensual domains that heretofore have not yet been observed as such, but which potentially inflect the operation of actual systems via informational transfer between neighboring systems of interrelations.” (<http://www.noemalab.com/sections/ideas/ideas_articles/stalbaum_landscape_art.html>)

Christina McPhee <www.christinamcphee.net>
Ill Atmospherics/Weather Works: Artistic Sonification of Meteorological Data

Andrea Polli
www.andreapolli.com

PROUD music of the storm!
Blast that careers so free, whistling across the prairies!
Strong hum of forest tree-tops! Wind of the mountains!
Personified dim shapes! you hidden orchestras!
You serenades of phantoms, with instruments alert,
Blending, with Nature's rhythmus, all the tongues of nations

Excerpt from Walt Whitman’s “Proud Music of the Storm” [1]

Introduction

For over ten years, I have been creating art works that translate numerical data to sound, from algorithmic compositions modeling chaos to live improvisation using video analysis systems. Areas of particular interest to my research have been modeling human methods of improvisation in interactive computer systems and using data sonification to illustrate complex information. Visualization is the interpretation of scientific data through the visual image, and likewise sonfication interprets data through sound. Sonifications can help scientific researchers understand data in a different way.

Since 2001, I have been working on the sonification of meteorological data in collaboration with Dr. Glenn Van Knowe at MESO, Mesoscale Environmental Simulations and Operations <http://www.meso.com> a leading firm in the development and application of atmospheric and other geophysical models for research and real-time applications. MESO works with the Mesoscale Atmospheric Simulation System (MASS) to create a highly detailed simulation of the weather based on terrain, initial conditions, and other factors. The atmospheric data sets produced by MESO are extremely detailed, and although they have a variety of visualization tools to interpret the data, much of the data represented is not visual in nature (temperature and atmospheric pressure for example). Through the project we wanted to learn what would happen if the data was interpreted sonically. In April 2003, we completed a series of multi-channel sonifications
of two historical storms, a tropical hurricane and a winter snowstorm at five elevations as part of a storm sonification project called Atmospherics/Weather Works.

The Atmospherics/Weather Works project has three primary goals: the development of a software system for the creation of sonifications based on meteorological and other data to be used in performances and installations, live and recorded musical performances, and a web site for the presentation and distribution of the recordings and software. [2]

The first public installation of the project was in April, 2003 at Engine 27 <http://www.engine27.org>, a non-profit organization devoted to the research, creation and dissemination of multi-channel sound works in New York City. A 16-channel sound installation spatially re-creates two historic storms that devastated the New York/Long Island area first through data, then through sound. The resulting turbulent and evocative compositions allowed listeners to experience geographically scaled events on a human scale and gain a deeper understanding of some of the more unpredictable complex rhythms and melodies of nature.

Why is scientific data so often presented as visual information and much less often presented as sound? One reason might have to do with time. A still visual image can be scanned over time, allowing a viewer to study various aspects of an image. A soundscape or piece of music, although it is also temporal, cannot be examined in detail without the destructive process of stopping, selecting, and replaying various parts. Aspects of the visual image are also easily defined by viewers. Specific colors and shapes can be described and understood more often than specific notes or musical phrases. Specific sounds also can have a level of ambiguity. Although some sounds are easily identified (like a barking dog or a cat's meow for example) the source of other sounds are not quite as clear. If noise or an echo interacts with a sound, it is like looking at a visual image wearing glasses that are heavily fogged, making recognition more difficult.

However, unlike a still visual image, music and soundscapes are inherently narrative. For example, as I listen to footsteps and voices outside my apartment door, I can determine that two people are walking up the stairs of my apartment building. I can determine approximately what floor they are on and even gather a little information about their relationship (are they a couple? a mother and child? have they been recently arguing or laughing?) In a visual image, a photograph of a family for example, unless the emotional states of the subjects are highly exaggerated, an observer is likely to encounter a certain amount of ambiguity in determining the relationships between the subjects.

Can an enhanced narrative and emotional content enhance the understanding of meteorological data? Some meteorologists call themselves 'storm hunters'. They travel far and wide at considerable physical risk in order to experience a hurricane or tornado. Is it because the physical and emotional exhilaration enhances the scientist's understanding of the storm? The storms hunters would most certainly answer in the affirmative. They experience the sound, scale, and physical properties of the storm as well as its direct
effect on the environment. A storm experienced only through visualization, whether animated or static, does not convey this visceral information. Scientists must use their imagination to create a mental image of a storm's potential devastation. A sonic experience of a storm can benefit communities beyond the meteorologist's lab. If a scientist is alerted by a visceral experience that a storm is likely to cause destruction, communities may be more quickly notified to prepare a proper response to the storm.

Our work represents a part of a growing movement in data sonification research. In 1997, The Sonification Report was prepared for the National Science Foundation by members of the International Community for Auditory Display (ICAD).[3] This report provides an overview of the current status of sonification research and proposes a research agenda. Most significantly to us as interdisciplinary collaborators, the report stressed the need for interdisciplinary research and interaction. Our project is well-suited to sonification according to the findings of ICAD. The data sets produced by MASS are extremely large and complex, and although there are a variety of visualization tools in use to interpret the data, much of the data represented is not visual in nature (temperature and atmospheric pressure for example). The data represented often portrays complex changes over time, an aspect of data particularly suited for sonification.

My personal interest in data sonification is in the artistic creation of new languages of data interpretation. As individuals and groups are faced with the interpretation of more and more large data sets, a language or series of languages for communicating this mass of data needs to evolve. Data interpreted as sound can communicate emotional content, and I am particularly interested in the sonification of data related to the atmosphere and the weather because of the long history of the weather used as a metaphor for emotion in the arts.

2 Project Planning

The project began when I met Dr. Van Knowe in the summer of 2001 at the first meeting of Bridges, an International Consortium on Collaboration in Art and Technology, a joint project of The USC Annenberg Center for Communication & The Banff Centre for the Arts New Media Institute [4]. Dr Van Knowe had joined MESO as a Senior Research Scientist after 24 years as a meteorologist for the Air Force. He was Chief of Meteorology at Rome Lab in New York where he directed the meteorological aspects of all research and was chief of the modeling and simulation development branch for the Air Force's Combat Climatology Center (AFCCC) at Scott AFB, IL.

Dr. Van Knowe and I brainstormed at that meeting and then continued to communicate via email and telephone to develop a project plan. After developing a proposal and being invited to participate in one of the first spatialized sound production residencies at Engine 27 to create a storm sonification, we met at MESO to plan the project. We wanted to create a spatial sonification of one or more storms that occurred in the New York area in the recent past in the hopes that some members of the audience would remember the
specific storms.

Dr. Van Knowe and Dr. John Zack of MESO suggested we try to create a sonification of a major winter snowstorm that in 1979 was not foreseen by the existing meteorological models and inspired years of research and development into improving the models. The "President's Day Snowstorm" initially formed as a weak wave of surface low pressure on a front in the Gulf of Mexico on 18 February 1979. Since this storm was not predicted by the existing meteorological models of the time, a large amount of data on this storm was available.

Later, Dr Van Knowe found a strong tropical Hurricane, Hurricane Bob, that passed though the same coastal region. We decided to attempt to sonify two storms that have a very different physical structure to see if the sonifications would yield insight into the nature of these two different types of storms.

3 Modeling the Storms for Spatialized Sound

Since the Engine 27 space has a very specific and unusual 16-channel speaker arrangement, we decided to map each speaker to a specific point in space proportional to the area spanning from Northern Florida to Northern New York State and from the Eastern tip of Massachusetts to Western New Jersey with New York City situated near the center. Simulated point data was to be modeled for an area of approximately 1000km. This area was mapped to the size and shape of the Engine 27 space. (see figure 1)

The kind of model output needed for sonification was very different that the output formats already in use by MESO for visualization. Dr. Van Knowe and his colleagues use the Mesoscale Atmospheric Simulation System (MASS) to create a highly detailed simulation of the weather based on terrain, initial conditions, and other factors. MASS takes real data inputs from satellite or surface readings and couples the information with global and regional models. There are several MASS output file formats: 3D array files, 2D horizontal (x-y plane) files, 2D vertical cross sections (x-z plane), 1D x,y simulated point observations, and 1D vertical profile (x,z) simulated point atmospheric soundings.

Our project required files of individual variables output for each geographical point at regular temporal intervals. Dr. Van Knowe and Dr. Kenneth Waight of MESO created a custom piece of software to output the data in this format. Kenneth T. Waight joined MESO in October 1987 after completing his Ph.D. in atmospheric science at the University of Wyoming. His first three years at MESO were spent on a project funded by the NASA Marshall Space Flight Center. Dr. Waight relocated to MESO's Troy, New York office in 1990 to assist in the development of MESO's real-time operational
4 Creating the Sonifications

After the storms were modeled and the data output, we were left with 720 data files of 481 values each and the daunting task of translating these numbers into sound. Engine 27 master programmer Matthew Ostrowski joined us at this stage and he and I worked at the Engine 27 space for a period of about four weeks creating a system for reading and translating the files to spatialized sound using Max/MSP.

We decided to create a composition of each day’s storm activity in full at each of the five elevations. We started by simply and directly mapping each variable to the pitch of a sound sample of a distinct timbre. We somewhat arbitrarily used long tones for temperature and pressure related variables and percussive tones for water related variables. The bank of sound samples used included vocal sounds, sounds created by wind instruments, and environmental sounds including the sounds created by various insects. The resulting sound compositions were interesting, but listeners found it difficult to hear the changes in each individual variable.

We then decided to map the total wind speed to the amplitude of the sound. Directly mapping loudness to wind speed for every speaker (every geographic point) created a dramatic spatialization effect. The fastest wind speeds, representing the greatest storm activity, created the most sonic activity and excitement.

However, the combination of timbres was still overwhelming to the listener, limiting the listener’s ability to make sense of the data. At this point, we decided not to limit the number of variables presented through the sonification for the sake of the public presentation. Had we been creating the sonifications for research only, at this stage we might have brought Dr. Van Knowe and his colleagues into the space to listen to and compare and contrast sound compositions created by single variables. However, there was a deadline for a public presentation of the work to a general audience and aesthetically we felt that the single variable compositions lacked the fullness necessary to engage a general audience expecting to hear a musical composition.

Dr. Van Knowe then created a complete model of each storm at 5 points of elevation: sea level, approximately 8500 feet, approximately 18,000 feet, approximately 35,000 feet, and approximately 60,000 feet (or, the top of the atmosphere). Each variable was output every three minutes for a 24 hour period of the greatest storm activity. The model grid resolution was 10km. Nine variables were modeled at this stage, but only six variables were used in the final sound compositions: atmospheric pressure, water vapor, relative humidity, dew point, temperature, and total wind speed.
The first aesthetic choice was to translate the atmospheric pressure data to a very low frequency sound. In doing so, listeners lost the ability to hear a detailed melody line describing the pressure changes, but gained a visceral sense of the storm.

Then, we began experimenting with using some of the variables as filter variables for sound samples representing other variables. Some of the variables in the model were highly coupled or inversely related to other variables. We created a band-pass filter that filtered a sound representing temperature with dew point values and filtered water vapor with relative humidity values. We found at this point that we needed to choose sounds with a wide spectrum in order to hear the filtering most effectively. White noise has the widest spectrum, and selecting ‘noisy’ sound samples proved the most effective in communicating the data and also was the most effective aesthetically due to the variation in the resulting sounds.

The scaling of the data for sonification presented particular challenges. Although the overall wind speeds varied with elevation levels, we decided to use global scaling for wind speed. This created the effect of the compositions building and receding in intensity. However, using global scaling for variables such as temperature mapped to pitch or water vapor mapped to a band pass filter proved to be much less dramatic that creating a scaling system for each elevation level of each storm since the variables differed widely between levels.

Finally, since the sonifications were to be performed in the format of a spatialized sound installation, we developed a daily schedule in which various compositions present the data sets at the five elevations, moving from ground level to the top of the atmosphere. In the installation, each storm was performed for approximately 1/2 hour six times each day. A storm consisted of six approximately five minute compositions presenting all variables at a single elevation and one combination of elevations based on the heights of the speakers. These compositions were marked by a number of ringing bell sounds, marking time and elevation like the ringing of church bells.

**Conclusion**

The final compositions were well received by both the general and the scientific audiences. Visitors to the installation particularly enjoyed remembering where they were during Hurricane Bob and the President’s Day snowstorm while listening to the sonifications. Some audience members found a metaphorical meaning in the series of rising elevations, finding the compositions nearer to the ground to be more visceral while those compositions representing activity closer to the top of the atmosphere were felt to be more ethereal and spiritual.
Dr. Van Knowe was particularly intrigued by the spatialization of the sound, and was interested in how the wave patterns of the storms were moving in space. The sonifications reinforced some known aspects of the particular storms. The winter storm was more intense near the top of the atmosphere while the hurricane’s fastest wind speeds occurred at lower elevations. This change in intensity was communicated very clearly through the varying degrees of loudness of the compositions. The patterns of movement of the tropical hurricane were known to be more chaotic than the winter storm, and the resulting compositions also reinforced this concept. Most listeners found that they could understand more the more they listened to the compositions, and there was an overall consensus that the work opens up doors for more research both in science and the arts.

Andrea Polli

References

In 1998, C5 had a problem; two problems actually.

We had organized that year as a business without a model to do a data collection and analysis project at SIGGRAPH 98, called the Remote Control Surveillance Probe project. The impetus for the founding of C5 was to see what kinds of business opportunities were available to a collaborative group of artists and theorists already working for many years with information as our primary medium. The expertise of C5 members was brought under one umbrella to tackle problems in domains relative to our collective experience, which includes autopoietic theory, artificial intelligence, information systems design and programming, public relations, emergent behavioral systems, semiotics, literary criticism, military studies, library science, and fine art.
Shortly after organizing, we were invited by Steve Dietz of the Walker Art Center in Minneapolis to do a net.art project related to a work by C5's president Joel Slayton, “Not to See a Thing”. The project had been exhibited as part of the 1997-98 "Alternating Currents: American Art in the Age of Technology" exhibition at the San Jose Museum of Art, in collaboration with the Whitney Museum of American Art. The “Not to See a Thing” project collected about 10 gigabytes of information about audience participation with the work during the time it was installed in the SJMA. What Steve Deitz was interested in was how we might hybridize the “Not to See a Thing” data with the infrastructure of the Internet itself to create a net.art project. This in essence created our two problems.

On the one hand we had a fairly large, but still manageable set of biometric data from Slayton's installation, which we had to mingle with the tremendous infrastructure of the Internet itself. And of course we had to find a way to make the manifestation of that data mingling visible/navigable to the user. Thus the first problem was related to the size of the datasets, and the need to develop a strategy for exploring them and exposing something about them. The second problem was that we were faced with two large sets of data that were superficially unrelated to one another. Our efforts culminated in the “16 Sessions” project, and the realization of the C5 IP database that Lisa Jevbratt developed to facilitate the mingling between the “Not to See a Thing” data and IP space. This paper focuses on the strategies that emerged from these projects and how they inform the matter of how artists can and should contribute solutions to these kinds of problems.

I'll begin with the scale problem first, because it is the less interesting of the two, and the solution is more obvious. The question is "How do you create a context in which information artists with different experiences and different sets of IT skills can participate in the exploration of and experimentation with large data sets?" We believe it is important to create a context that is amiable to both collaboration and independent endeavor at a variety of interface levels. Technically this requires the development of multiple interfaces to the data which are congruent with the experience of the various groups of people who will be working with it. To ensure this, whenever possible, artists should be involved with or completely responsible for the development of the various interfaces. Given that artists today are also computer programmers, database administrators, information architects, engineers, and theorists, it’s important that the data to be worked with be arranged for maximum access. Access which ranges from the raw
layers of software development, and importantly experience working with each other, so the process was relatively smooth. Of course, this is not the situation with larger sets of institutionally collected data, where the standards, data formats, and API's can often be quite obtuse.\textsuperscript{viii}

Different challenges exist with the emergence of large collections of public data such as is available from the United States Geological Survey, NASA, NOAA, and the Human Genome Project. Not only challenges presented by the technical sophistication of the data and the tremendous size of the data, but in strategizing appropriate interfaces to the data that allow users of very diverse backgrounds to participate in the process of consuming the data and generating new knowledge from the data. C5 has been active in this area. For example, the C5 Landscape database is a relational database, Perl API and set of sample interfaces designed specifically to help users in creating their own programs that can easily access, analyze and display information about the shape of the earth.\textsuperscript{x} The database is designed to eliminate much of the complexity in acquisition, database interface, processing and imaging common in the manipulation of geo-data, so that artists have a manageable platform in which to write their own software and perform mapping experiments. Artists using the software can work with the database from various levels of technical sophistication. These levels range from a web-based GUI
directly through SQL, Perl DBI and Java JDBC programming techniques. An API also provides a variety of features and capabilities through easy to use Perl modules.

There are of course many projects that incorporate the idea of artists working with data at all levels. Notable are Lisa Jevbratt’s “Mapping the Web Infome”, and Rhizome’s “alt.interface” projects. Rhizome's “alt.interface” project involves exposing (to artists) the database API of the Rhizome website and its large text object collection, such that they can create alternative interfaces. Jevbratt's web crawling project is especially notable because of the way that she worked with the invited artists to create both an interface for the 'alternatively' technical artists involved, as well as working at the database and API levels with many of the artists to collaboratively implement features suggested by artists. It is appropriate for artists to be involved in the development of the public API's and application layer interfaces through which the public at large will have access to large data, because in many cases artists working collaboratively already have experience in working out the inherent interface issues that are involved in making data available to 'technically diverse' or even non-technical users. Artists in both new media academia and fine art practice have been involved in this kind of work for many years.

The second issue is a deeper one involving how artists have and can contribute to dealing with inter-relations between very different datasets, as well as unexplored intra-relations within single large datasets of considerable complexity. The exploration of large datasets is one of the most provocative and interesting issues for artists today because of the explosion of availability of such large data sets being made available to the public.

Why? Artists as cultural workers have always sought to contribute to the state of our knowledge near the edges of human understanding. Among the new cultural problems we face today are the problems of big data. And lest you assume that this is exclusively the domain of computer science, the large datasets of today present new kinds of problems which computers and networks are not traditionally used to solve, and perhaps even that the traditional use of computers and networks can not solve. The familiar notion of the "information processing life cycle" is the basis of contemporary data processing. This is the very colonial idea that data is
notion holds that data must be processed into useful information, and to accomplish this you normally start by considering the output you want, the available input, and then determine the algorithm that will take your raw and untreated data and turn it into a manageable, cognizable, useful thing we call information. The entire field of Data Mining and Knowledge Management as we know it today is predicated on the pre-existence of semantic models that allow data to be algorithmically mined for meaning. This is the basic philosophy and approach to data and information, and is of course profoundly successful, but its application reaches severe limitations in dealing with contemporary data and the new kinds of problems it presents.

For example, traditional problem solving is not at all applicable to the situation C5 faced with “16 Sessions”. We had two very different data sets, and although we had some preconceptions of what they meant, we had no idea how they were related or if they were related, and no clear idea of what kind of question to ask. Neither set of data was collected with a protocol that was designed to facilitate the type of endeavor that we were charged with performing. Again, standard information processing techniques are not useful for all problems, especially when you do not have a question, when you have a poorly formed question, or when the dataset itself is not entirely understood or contains information potentials that were unplanned at the time it was collected. Data may have non-transparent semantics, or may be so complicated that you do not know where to begin to search, or it may take on new roles as new needs emerge after the data is collected. These issues are of course also related to the problem of what questions to ask. When you don't understand your data, you will naturally have poorly formed questions about it.

Why is this an important problem? The answer is that there is ever more data being collected in various endeavors about which we don't know what questions to ask. For example, the Human Genome Project has sequenced and published the entire human genome, but that tremendous data set is largely unexplored, because in part, scientists have not sought the answers to questions not yet raised. While this may seem quite tautologically obvious, it is simultaneously a tremendous and real problem. As put by Lisa Jevbratt, the process of exploring genomic data can be "described as that of a group of people in a dark room fumbling around not knowing what is in the room, how the room looks or what they are looking for." Genomic data is non-unique in this respect. There are, for example, vast datasets available from the United States and other governments regarding all kinds of interesting things that we don't yet fully understand, or that we think we understand but which has behavior and relations that have been overlooked. And artists, who do not always participate in the scientific method, may well make discoveries or observations in their aesthetic and conceptual pursuits with such data that lead to such questions, even if the artists are participating as blind probe heads in data space.

The exploration of such data, I argue, is the most productive and culturally useful positions from which to perform as an artist in the 21st century. It is hard now to make
faced with large sets of data without a map nor a clearly defined problem definition is one of the most interesting and provocative problem types we face in an era where our ability to collect data outpaces our ability to generate knowledge from it. Asking questions and exploring spaces in poorly defined problem domains consisting of huge datasets is the natural, useful, and potentially highly productive cultural role in which artists should play a part.

C5's approach to these types of problems is to explore the application of autopoiesis as a conceptual framework for understanding the behavior of data and information. Autopoiesis takes place in systems that differentiate themselves from other systems on a continual basis through operational closure, and that produce and replace their own components in the process of interaction with their environment (structural coupling), that occurs via a membrane containing the organization of the unity in question, thus allowing distinction between it and its environment. A basic question for any analysis of the autopoietic potentials of data involve distinguishing a membrane, or the interface, where operational closure (inside) and structural coupling with an environment (outside) are expressed. It is in patterns of structural coupling that relations between complex data can be analyzed. If you can find a membrane, you have revealed a relation between or within data sets. To find membranes, you need to mingle data. For example, there are contemporary explorations within the social sciences that demonstrate that relations exist containing information the landscape (for example drainage, land cover, or topography), can reveal insights when mingled with historical data. C5 views these types of data processing explorations as very interesting instances of structural coupling between data sets, even those as superficially different as geological and historical data.

Most of C5's approach to autopoietic frameworks for the understanding of large data has been developed by Joel Slayton and Geri Wittig. Perhaps the key idea that emerges from their work is the notion of a composibility of relations, in that composibility indicates the potential for autopoietic membranes existing as data relations via third order structural coupling in a coded environment. This allows for the analysis of data sets where the semantic relationships are uncertain. In a sense, this idea can be described as the search for algorithms in which superficially different data sets might be shown to couple based on their subject-less form through inherent sans-semantic or pre-semantic models, and to seek these relations specifically to flag the potential for the presence of immanent, unplanned, or otherwise unrecognized semantics flowing from mingled relations, thus revealing something about the ontology of the sets that produces new knowledge about them. It is unlikely that there is a universal algorithm for this, (such as a universal visualization system for all data), but if there is, it is likely to be accidentally discovered by researchers searching for inter-relations between data sets. Obviously, artists should be involved in this endeavor.

This is only one approach, undertaken by a small self-funded organization that believes a very particular theoretical framework can be expressed in coded relations that deliver their own answers. To explore this, we of course need a lot of data. It is important
that science organizations create the circumstances that will allow a diversity of independently theorized approaches to emerge based on public interest in and public access to the data. It is in casting large sets of scientific data into the realm of artists, and indeed the public at large, which will allow a multitude of self-organized modes of discovery to develop.

Brett Stalbaum

NOTES
First Published as:
Software Development Platforms for Large Datasets: Artists at the API,
Leonardo Electronic Almanac volume 11, number 5, May 2003 ISSN #1071-4391
This essay began as speaking notes for a talk of the same title delivered at a rhizome.org event sponsored by Qbox in San Francisco CA, April 26th 2002. http://rhizome.org/events/rhizome_sf_apr.php3


2 http://www.c5corp.com/walker/gateway.html

3 http://www.c5corp.com/projects/16sessions/index.shtml

4 The internet protocol is the numerical addressing scheme used to identify devices on the internet.

5 This later became the technical basis for 1:1, http://www.c5corp.com/projects/1to1/index.shtml

6 API is the acronym for application programming interface, which is a group of public functions that exist in a library of code that other programmers can make use of to implement their own code. Artists should design API's as well as use them.

7 http://www.c5corp.com/softsub/index.shtml

8 A good example of this is the Spatial Data Transfer Standard. According to computer scientist Gregg Townsend, "The adoption of SDTS was a giant step backwards. While previous DEM files could be read by relatively simple programs, SDTS file are difficult to read even with the help of a large external library." http://www.cs.arizona.edu/topovista/sdts2dem/

9 http://cadre.sjsu.edu/~gis

10 http://dma.sjsu.edu/jevbratt/lifelike/

11 http://rhizome.org/interface/

11i For example, see http://fisher.lib.virginia.edu/projects/salem/ The GIS of "Salem Village in 1692" is part of an electronic Research Archive of primary source materials related to the Salem witch trials of 1692


11iv http://cse.ssl.berkeley.edu/nvo/nvo.htm