Algorithmic Object As Natural Specimen: Meta Shape Grammar Objects from ATLAS in silico

Ruth West[1], JP Lewis[2], Todd Margolis[3], Jurgen P. Schulze[3], Joachim Gossmann[3], Daniel Tenedorio[3], Rajvikram Singh[3], Correspondence: [1]UCLA Center for Embedded Networked Sensing (CENS), BOX 951596 3551G Boelter Hall, Los Angeles CA 90095-1596, U.S.A, E-mail: ruth@viewingspace.com. [2]Weta Digital and Massey University, [3]University of California, San Diego

Abstract

High-throughput science and digital social infrastructures transform nature and culture in to vast and abstract data, yet to reap the benefits of this data requires transcending representational challenges. Atlas *in silico* is a large scale art-science collaboration that attempts to bridge this gap by placing abstract data in a form and context that can be experienced. In this aesthetically impelled work, we use n-dimensional glyphs to visually personify metagenomics data from the Global Ocean Sampling Expedition. Viewers experience intuitive, open-ended browsing and exploration of this massive abstract dataset. Purposefully positioned between art/new media and science, Atlas *in silico* resists definition as either to reflect upon the digitization of nature and culture..

Modern scientific practice is often regarded as being remote from experiential and aesthetic concerns. In an increasing number of scientific disciplines, hypotheses are formulated and evaluated as algorithmic queries applied over millions of data records representing the digitized data. The scientist does not see or hear the data and has little human experience of it.

This dissociation brings some risks. Statistics texts give illustrations of false conclusions that can easily occur with such blind approaches: models applied to data that do not fit the models' assumptions silently give incorrect conclusions [1]. More generally, this "data blindness" may encourage orthodoxy. How do you look for the unexpected when the data is only accessed through algorithmic queries? Discovery of unexpected patterns may occur through the programming of new queries, but this is hardly fluid.

Scientific practice was not always so separated from visual and aesthetic experience. In the time of Darwin, visual atlases documented variation amongst biological specimens at a material scale, recording and displaying phenotype at multiple levels of granularity from subtle nuances to radical discontinuities in structure and function. In his E Notebook on transmutations of species, Darwin wrote, "…a grain of sand turns the balance," in regard to the subtleties of variation and its relationship to natural selection [2]. Just as Darwin faced the challenge of representing natural selection in a 19th Century visual culture invested in the concepts of species fixity [3], we in the 21st Century face representational challenges in engaging with massive multidimensional digital data collections that create new views of nature and culture. Can experiential, artistic or aesthetic approaches transcend these challenges and enhance our encounters with such vast data collections in which the subject of study is high-dimensional, invisible, or otherwise abstract?

We created ATLAS *in silico* to explore this challenge of *experiencing massive abstract data*. ATLAS *in silico* is an interactive and immersive virtual environment/ installation that utilizes, and provides an aesthetic encounter with, metagenomics data of marine microorganisms from the Global Ocean Sampling Expedition. This research is a large-scale art-science collaboration (ongoing almost three years, involving up to twelve research collaborators, and over time an additional dozen students) encompassing the disciplines of virtual reality, auditory display and sonification, visualization, electronic music, new media arts, metagenomics, computer graphics, and high-performance computing all of which are engaged in a working process combining artistic practice and inquiry. In this aesthetically impelled work, we explore the use of n-dimensional glyphs to visually personify millions of abstract individual records from the GOS (each record having variation in sequence, molecular structure and function as well as contextual metadata), and place the biological data within each record in a human context.

The Global Ocean Sampling Expedition

The initial Global Ocean Sampling Expedition (GOS) was conducted by the J. Craig Venter Institute in 2003-2006 to study the genetics of communities of marine microorganisms throughout the world's oceans. This global circum-navigation was inspired by the HMS Challenger and HMS Beagle expeditions. The microorganisms under study sequester carbon from the atmosphere with potentially significant impacts on global climate, yet the mechanisms by which they do so are poorly understood. Whole genome environmental shotgun sequencing was used to overcome the inability to culture these organisms in the laboratory. This resulted in a vast metagenomics data set that is the largest catalogue of genes to date from thousands of new species, with no apparent saturation of the data in regards to gene discovery [4]. The data contains DNA sequences and their associated predicted amino acid (protein) sequences. These predicted sequences, called "ORFs" (Open Reading Frames), candidates for putative proteins, are subsequently validated by a variety of bioinformatics analyses. The data also includes a series of metadata descriptors, such as temperature, salinity, depth of the ocean, depth of the sample, latitude and longitude of the sample location that describe the entire GOS data collection. Values for each metadata category were recorded at geo-coded sampling locations and are stored along with the raw data. As an indication of the massive scale of the GOS, the initial set of data released in 2007 nearly doubled the number of proteins in publicly accessible genetic databases that had taken the world-wide scientific community almost 30 years to amass.
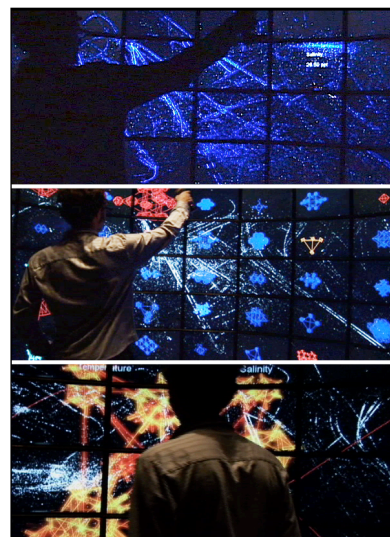

Figure 1: Interaction with overview (top); subset (middle); single GOS record n-glyph (bottom).

ATLAS *in silico*

Within ATLAS *in silico's* virtual environment users explore GOS data in combination with contextual metadata at various levels of scale and resolution through interaction with multiple data-driven visual and auditory patterns at

different levels of detail. The notion of "context" frames the entire experience and takes various forms ranging from structuring the virtual environment according to metadata describing the entire GOS data collection, to the use of socio-economic and environmental contextual metadata (pertaining to geographical regions nearest GOS geo-coded sampling sites) to drive visual and auditory pattern generation, to playing a role in both data sonification and audio design and spatialization that is responsive to user interaction. Participants experience an environment constructed as an abstract visual and auditory pattern that is at once dynamic and coherently structured, yet which only reveals its underlying characteristics as the participant disturbs the pattern through their exploration.

The installation was exhibited at SIGGRAPH 2007, Ingenuity Festival Cleveland 2008, the Los Angeles Municipal Art Gallery 2008 - 2009 [5, 6, 7] and is resident at the UCSD Center for Research in Computing and the Arts (CRCA) and Calit2 (California Institute for Telecommunications and Information Technologies) where it continues in active development. The installation has been presented at venues with a combined audience of over 100,000 visitors. A selection of printed visual altas plates and rapid prototyping sculptures of algorithmic objects as natural specimens was presented as part of the UCDarnet Scalable Relations exhibition in 2009 [8].

In addition to public-facing exhibitions, this ongoing art-science research has generated a strategy for multi-scale auditory display, the Scalable Auditory Data Signature [9], that encapsulates different depths of data representation through a process of temporal and spatial scaling. Our research has also produced a novel strategy for hybrid audio spatialization and localization in virtual reality [10] that combines both amplitude and delay based strategies simultaneously for different layers of sound. This allows each layer to support and highlight its intended role within the evolving sonic environment and achieve increased differentiation and perceptual depth between auditory elements. Research on the multi-scale visualization schemas and interactive framework is ongoing with additional publications forthcoming. Over its almost three-year duration the research collaboration and exhibitions have received support from a variety of academic and industry groups. (Video is at: http://atlasinsilico.net)

VR Systems

The installation has been implemented on two types of virtual reality systems, first on the Varrier™ autosterographic display [11] followed by implementation on a C-Wall passive stereo rear-projection display system for touring. The Varrier™ is a sixty-LCD, semi-circular, 100-million pixel autostereographic tiled display.

Computation is driven by a 16 node dual-Opteron Linux cluster. The display is coupled with 10.1 multichannel spatialized interactive audio and interaction is via wireless head and hand tracking using computer vision and IR-reflective markers [12]. For touring, the installation was modified to use a passive stereo display system. The touring installation is rear-projected in stereo on a polarization-preserving screen. The multi-channel spatialized interactive audio makes use of eight channels. User input/interaction is tracked through an electromagnetic system using both head and hand sensors [13]. Developed in the COVISE VR framework [14], the installation makes use of the OpenSceneGraph API, OpenGL and OpenCover graphics subsystems [15, 16]. Pure Data is used for real-time audio [17]. GOS data and contextual metadata is accessed via a series of SQL queries from a MySQL database [18].
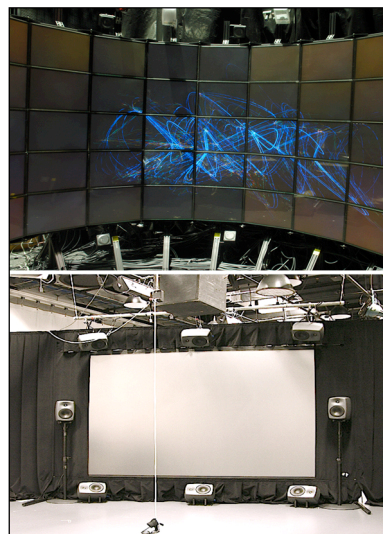

Figure 2: Varrier™ (top); C-wall (bottom).

Viewers position themselves at a location in front of the Varrier™ display and inside the tracking volume. Within this central viewing area immersion and autostereographic viewing are optimal. ATLAS *in silico* makes use of this optimal viewing and tracking volume artistically and conceptually to enable the blending of quantitative and qualitative representations. This is accomplished by positioning the entire data set in its contextual overview mode predominantly behind the implied image plane created by the surface of the LCD tiles, and progressively moving the data-patterns forwards towards the user in the coordinate space (Fig.1, above) while retaining this contextual overview in the background. Implementation on the C-wall display required modification of all system components yet this sequence of position and presentation was retained.

Shape Grammar Objects: N-dimensional glyphs

Our approach for visualization and sonification of GOS data and contextual metadata is data-driven yet non-hypothesis driven. It is a hybrid multi-scale strategy that merges quantitative and qualitative representation with the aim of supporting open-ended, discovery-oriented browsing and exploration of massive multidimensional data collections in ways that do not require *a priori* knowledge of the relationship between the underlying data and its mapping.

Data is presented visually within the virtual environment as dynamic and abstract patterns in different positions relative to the overall virtual world coordinate system and the user (real-world) coordinates. Data is also presented not only as audio objects within the virtual environment, but using spatialization strategies that position and move audio objects relative to the user according to both their interaction with the data patterns, and the relation between and within data objects themselves. As users select a small subset of records from the larger data set, their representation transforms from luminous colored particles (one per GOS record) to a series of geometric shapes of different size, color and number of vertices (Fig.1, above). These shapes are n-dimensional glyphs generated by a custom meta shape grammar.

We aim to generate glyphs that reflect differences in the underlying data or metadata so that differing data can potentially be visually sorted in an exploratory fashion, by an observer who is not familiar with the details of the mapping from data to glyph attributes.

This requires an approach that generates as wide a range of distinctive patterns as possible. We select shape grammars for glyph generation because they generate a relatively large range of distinctive patterns relative to the coding complexity. While shape grammars were one of the earlier approaches to computer pattern synthesis [19], they have remained relatively underexplored compared with popular approaches such as fractals and L-systems. Through preliminary work, we observed that picking rules without constraints generates chaotic shapes that are not memorable. Thus, it is desirable to have some structure in the rule selection process, yet still allow it to reflect the data.

We used an approach [20] in which the rule generation and interpretation is mapped to higher-order functions, a feature of functional languages such as Lisp and ML. An example of a meta-rule is a procedure that recursively builds and places other rules at the vertices of a polyhedron guided by some data attribute mapped to this rule.
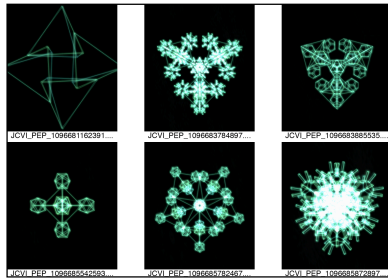

Figure 3: Shape grammar objects for several GOS records (2D projections).

Data values driving glyph generation for each GOS record include the amino-acid (ORF) sequence, bio-physicochemical properties of this sequence such as amino acid composition, molecular weight, hydrophobicity and secondary structure characteristics, as well as the metadata values associated with its physical sampling location, and those associated with the sampling locations in which it was algorithmically identified. (ORFs) that were physically sampled in one location were found to algorithmically assemble in collections of sequences from up to 24 other geo-graphical locations from which they were not physically sampled [21].

Reaping the full benefits of the GOS and metagenomics research overall requires the development of next generation computational and

networking technologies to enable the scientific community to work with the scope and massive scale of this type of data [22]. To reflect this broader socio-technological and ultimately human context, a set of contextual socio-economic, technological and environmental metadata descriptors for the regions nearest the geographical locations where the microbial samples were collected was selected and compiled to add this secondary level of context to the GOS data. Contextual metadata representing CO_2 emissions per capita, infant mortality rates and internet users per capita was collected from publicly available data repositories and combined with biological data for each ORF. A custom meta shape grammar combines these heterogeneous data types to seed each level of iteration. This visualization strategy combines both quantitative and qualitative data in one high-dimensional representation. Research on our multi-scale visualization schemas is ongoing and our algorithms will be presented in future publications.

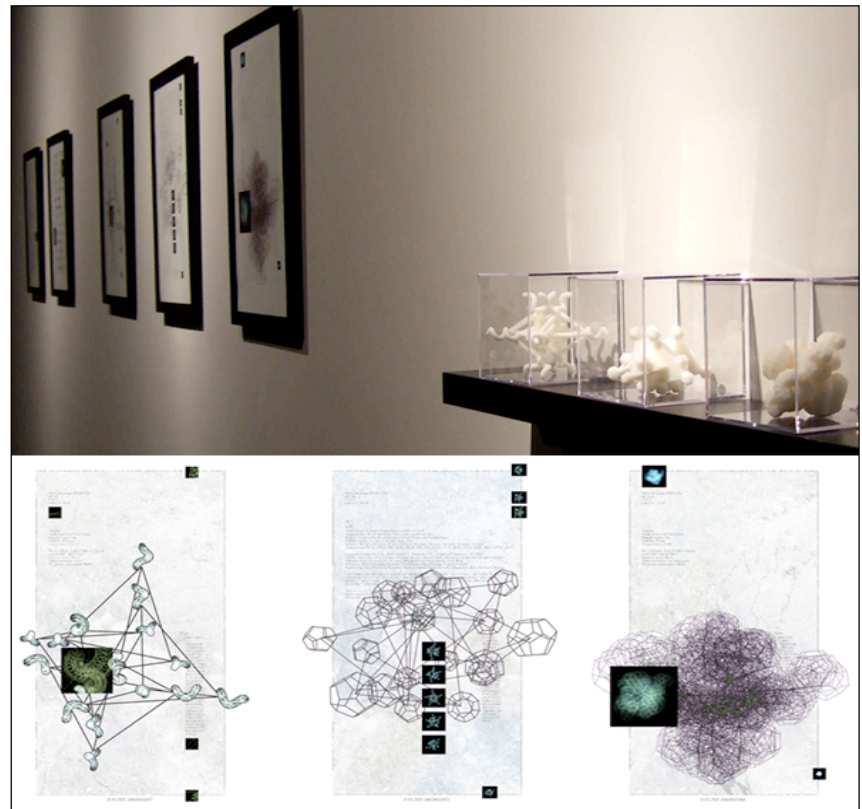## Algorithmic Object as Natural Specimen

Figure 4: Algorithmic objects as natural specimens - atlas plate prints and rapid prototyping sculptures (top); printed atlas plates for three GOS records (bottom).

In addition to their dynamic representation within the virtual environment, 2D and 3D projections for a subset of these n-dimensional glyphs were manifested as archival digital prints and rapid prototyping sculptures. Each glyph is presented as a print and sculpture along with the underlying data fore-grounded in the context of satellite imagery of its geo-coded physical sampling location.

Our images and sculptures of the shape grammar objects return to the formalisms of biological atlases and preserved specimens of the 19th Century, documenting variation at a material scale that allows a visual experience of the data to augment blind algorithmic queries. This historic linkage is also reiterated in the virtual world by combining the aesthetics of fine-lined copper engraving, lithography and grid-like layouts of 19th Century scientific representation with contemporary digital aesthetics including 3D wireframes, particle systems, and spatialized multichannel audio. The printed atlas plates are presented alongside the corresponding rapid prototyping sculptures enclosed in specimen cases.

To highlight the historical and ongoing interplay between scientific discovery and culture, the installation and these algorithmic objects as natural specimens contextualize GOS ORF

sequence data (representing individual protein molecules) within environmental and social metadata from geographic regions nearest the ocean locations at which the microorganisms were sampled, and combine these to generate the n-dimensional glyphs by driving the shape grammar, as well as driving all of the visual and auditory elements of the virtual environment.

Reflections

This work explores the potential interplay between artistic and data-driven strategies, based on visual and auditory pattern, in working with massive multidimensional multi-scale, multi-resolution data. Taking pattern and its application in both science and the arts as a focal point, a point of intersection, and a point of departure, this work focuses on alternative representational and interactive approaches to support open-ended, intuitive, discovery-oriented browsing and exploration of massive multidimensional data in both research and public/artistic contexts. It explores the potential for aesthetic, artistically-impelled experiences to support data exploration and hypothesis-generation in ways that transcend disciplinary boundaries and expertise while simultaneously interrogating the cultural context of the underlying science and technologies to produce a work with multiple entry points. Purposefully positioned on the edge between art/new media and science, the balance in this work at times shifts from one to the other, yet overall in both process and product it resists definition as either.

Metagenomics offers us a new view of nature that is sequence-centric rather than organism-centric. In parallel to challenges that Darwin's work on natural selection posed to 19th Century representations of nature based on concepts of species fixity, the new view of nature provided by vast and abstract metagenomics data, such as the Global Ocean Sampling Expedition, poses a fundamental challenge for our ability in the 21st Century to represent and intuitively comprehend nature as well as our evolving relationship to it in a time of global environmental concern. Within ATLAS *in silico*, this challenge becomes a visceral, sensate experience of the abstraction of nature and culture in to vast databases — a practice that reaches back to the history of expeditionary science of the 19th Century and which continues in 21st Century expeditions such as the Global Ocean Sampling Expedition.

References and Notes

1. Cleveland, W, Visualizing Data, Hobart Press (1993).

2. Darwin, C. R. Notebook E: [Transmutation of species (1838-1839)] Online: "http://darwin-online.org.uk/content/frameset?viewtype=side&itemID =CUL-DAR124.-&pageseq=99"

3. Simith, J. Charles Darwin and Victorian Visual Culture. Cambridge University Press, 2006.

4. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, et al. "The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families." PLoS Biol 5(3) (2007).

5. West R, Margolis T, Mostafavi I, Lewis JP, Gossmann J, Hackbarth B, Horn AS, Fernald S, Schulze J, Li W, Henthorn T, Yamada T, Singh R, Chheng T, Cassey T, Lien J. "ATLAS in silico," ACM SIGGRAPH 2007 Art Gallery, (2007) p. 225.

6. West R, Margolis T, Mostafavi I, Lewis JP, Gossmann J, Hackbarth B, Horn AS, Fernald S, Schulze J, Li W, Henthorn T, Yamada T, Singh R. "ATLAS in silico," Ingenuity Festival Cleveland (July, 2008)

7. West R, Margolis T, Mostafavi I, Lewis JP, Gossmann J, Hackbarth B, Horn AS, Fernald S, Schulze J, Li W, Henthorn T, Yamada T, Singh R. ATLAS in silico, Los Angeles Municipal Art Gallery, (November 6, 2008 – January 19, 2009).

8. West R, Lewis JP, Margolis T, Gossman J, Schulze JP. "…a grain of sand turns the balance." UCDarnet Scalable Relations Exhibition, UC Santa Barbara, curated by Christiane Paul, (February 13 – March 13, 2009).

9. J. Gossman, B. Hackbarth, and R. West, with T. Margolis, J.P. Lewis, and I. Mostafavi. "Scalable Auditory Data Signatures for Discovery Oriented Browsing in an Expressive Context." Proceedings of the 14th International Conference on Auditory Display, Paris, France. (June 24 - 28, 2008).

10. West R, Gossman J, Margolis T, Schulze JP, Lewis JP, Tenedorio D. Sensate abstraction: hybrid strategies for multi-dimensional data in expressive virtual reality contexts. Proceedings of 21st Annual IS&T/SPIE Symposium on Electronic Imaging and conference on The Engineering Reality of Virtual Reality, (18-22 January 2009 San Jose, California USA).

11. Sandin D, Margolis T, Ge J, Girado J, Peterka T, and DeFanti T. "The VarrierTM autostereoscopic virtual reality display." Proceedings of ACM SIGGRAPH, 24(3): 894-903, (2005).

12. Online: "http://www.ar-tracking.de."

13. Online: "http://www.ascension-tech.com/realtime/RTflockofBIRDS.php"

14. Rantzau D, et al., "Collaborative and Interactive Visualization in a Distributed High Performance Software Environment." Proc. International Workshop on High Performance Computing for Graphics and Visualization, Swansea, Wales, (1995).

15. Online: "http://www.openscenegraph.org/projects/osg"

16. Online: "http://www.hlrs.de/organization/av/vis/covise/"

17. Online: "http://www.crca.ucsd.edu/~msp/software.html"

18. Online: "http://www.mysql.com/"

19. Stiny, G. and Gips, J. Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts, U. California Press. (1978).

20. Lewis, J.P. , Rosenholtz, R., Fong, N., Neumann, U. VisualIDs: Automatic Distinctive Icons for Desktop Interfaces, ACM Trans. Graphics Volume 23, #3 (August 2004), pp. 416-423.

21. Li. W. personal communication (2007).

22. Online: "http://camera.calit2.net/"