DataRemix: Designing the Datamade

Ruth West, Roger Malina, John Lewis, Scot Gresham-Lancaster, Alejandro Borsani, Brian Merlo, Lifan Wang

Leonardo, Volume 48, Number 5, 2015, pp. 466-467 (Article)

Published by The MIT Press

DataRemix: Designing the Datamade

Ruth West, University of North Texas, ruth.west@unt.edu; Roger Malina, UT Dallas, roger.malina@utdallas.edu; John Lewis, Victoria University, john.lewis@vuw.ac.nz; Scot Gresham-Lancaster, UT Dallas, scotgl@utdallas.edu; Alejandro Borsani, University of North Texas, alejandro.borsani@unt.edu; Brian Merlo, UT Dallas, brian.merlo@utdallas.edu; Lifan Wang, Texas A&M, wang@physics.tamu.edu.

See <www.mitpressjournals.org/toc/leon/48/5> for supplemental files associated with this issue.

Abstract

ArtScience is emerging as one approach for creating novel ways of seeing and new ways of knowing. The authors propose a role for ArtScience research and creative work in contributing to the necessary shifts to go beyond the current crisis of representation. DataRemix, a recombination and reappropriation practice intended to trigger novel subjective experiences and associations, is described.

The increasingly rapid pace of digitization of nature and culture will result in a "digital universe" by 2020 estimated at 40 trillion gigabytes [1]. What defines data as "big" is shifting. In some fields, such as the digital humanities, access to large-scale data enables new kinds of analyses but does not necessarily require use of supercomputers. boyd and Crawford [2] observe that big data is no longer solely characterized by its size, but by the "capacity to search, aggregate, and cross-reference large data sets." And as Gantz and Reinzel point out, ironically, "as soon as information is created or captured and enters the digital cosmos, much of it is lost" [1]. An aura of incompleteness is thus associated with the accumulation of data of unprecedented resolution, size and scope.

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 from instruments that detect phenomena to which our own senses are blind. Anticipating this trend, historian of science Daniel Boorstin identified "epistemological inversion," a shift from primarily hypothesis-driven to discovery-driven approaches as one of the profound implications of big data for science; he joked that we were moving from a "meaning rich, data poor: to a data rich, meaning poor" situation [3]. While transforming and accelerating discovery, algorithmic approaches often do not provide a direct or embodied experience of the data. The scientist does not necessarily 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 [4]. Often non-uniqueness problems occur when many different models can be matched to the same data. More generally, this "meaning blindness" may encourage orthodoxy.

Complex relationships between the practices of science, communication about science, aesthetic experience and visual culture are long-standing. In the time of Darwin, visual atlases produced by skilled artists, engravers and lithographers recorded and displayed subtle nuances and radical discontinuities in variation amongst biological specimens. Analogous to the ongoing accrual of massive data, the great scientific expeditions, such as the voyage of the H.M.S. Challenger, resulted in collections of unprecedented numbers of specimens. And just as Darwin faced the challenge of representing natural selection in a 19th century visual culture invested in the concepts of species fixity [5], we in the 21st century face representational challenges in engaging with vast and abstract data. Often massive data, such as metagenomics collections that disrupt our organism-centric view of nature and reorient it towards communities of sequences, abstract phenomena outside the human perceptual range. Framing narratives then arise from our choice of algorithms, statistics, representational schemas, displays, interaction technologies, and metaphors used in processing and representing the data. These narratives reflect culturally prevailing ways of seeing and knowing that lack representations for data that engender concepts outside their scope. One particular shift in metaphor is occurring as we move from thinking of knowledge organized through a tree of knowledge to, within our networked society, knowledge as a network structure. The shift in metaphor necessarily entails shifts in systems of representation.

The narratives framing data creation and representation circumscribe what we can see and know, and how we see and know. In an era of data richness, we are convinced that we are faced with a crisis of representation. How do we see and know beyond what our instruments, algorithms, representational schemas and prevailing culture enable us to see and know? How do we make explicit the implicit assumptions in our systems of representation and their blind spots? In other words, how do we look for what we don't know we're looking for? Artists and scientists each bring their own experimental approaches to making sense and meaning. ArtScience is emerging as one approach for creating novel ways of seeing and new ways of knowing and exploring "hybrid" strategies [6]. Our work asks: Can experiential, artistic or aesthetic approaches transcend these challenges and enhance our encounters with vast data collections in which the subject of study is high-dimensional, invisible or otherwise abstract?

DataRemix: A Working Definition

We define DataRemix as: the reappropriation and recombination of data in any and all states along the continuum of its transformation from raw to processed [7]. We are interested in the use of additional sensory modalities to make the abstract experiential, in combination with the cultural practices of reappropriation and recombination, the externalization of intuition by making, and the creation and use of metaphor. Grounded in the view of remix as a form of literacy for the 21st Century, it posits a mechanism to destabilize the narratives framing data creation and representation that contribute to the crisis of representation [8].

In the context of DataRemix, data is conceptualized as having multiple transition states rather than as discrete objects. Transitions between raw, partially processed and processed states are triggered by algorithmic, analytic, statistical or other manipulations of the data and decisions by the designers of the system of representation. Novel representational schemas and interactive modalities can be created and explored by externalizing this continuum and making it available to be remixed. Partially processed data and intermediary states along the continuum can be remixed to create different outcomes from those created by the original domain or problem-specific processing pipeline. For example, if noise is being removed from raw data with the final end state being "clean" data, then data remixing options would include analysis of the noise and re-incorporation of noise into different intermediate stages to produce multiple outcomes and representations. DataRemix as

 doi:10.1162/LEON_a_01060

a practice and a way to collaborate are linked and made possible by this concept of data as fluid.

Our concept of DataRemix aligns with, and potentially extends, Eduardo Navas's concept of Regenerative Remixes: "juxtaposing two or more elements that are constantly updated, meaning that they are designed to change according to data flow" [9]. Navas distinguishes regenerative remixes that require continual updating from three basic types that include extended (remixing to extend the duration in time), selective (remixing through the process of addition or subtraction of content from the original while retaining its essence), and reflexive (creates an allegory: material is added and subtracted; the remixed work stands on its own yet is dependent on the original). It is likely that our work in developing DataRemix Engines for specific application domains will incorporate remix types that fit multiple, if not all, of Navas's classifications.

DataRemix is also an approach to interdisciplinary collaboration characterized by the existence of multiple inputs and objectives, each relevant to one or more disciplines and collaborators, that result in a variety of outputs with different finalities. These types of ArtScience collaborations are facilitated by the view of data as including transition states that can be accessed and utilized for various purposes, including multiple analyses of the same data. In contrast to ArtScience collaborations that culminate in one overarching hybrid outcome, such as a single artistic installation or publication, these collaborations are characterized by joint group goals being created collaboratively alongside work to progress individual goals as part of an explicit strategy. These may include creating works of art, new technologies, aesthetically impelled scientific tools, education outreach, and multiple forms of disciplinary and interdisciplinary knowledge, simultaneously and at various temporal rates.

The collaborative process functions as a path to multiple finalities, each of them valid within some contexts and not within others. Looked at from a "30,000 foot view" this type of collaboration appears somewhat chaotic due to the multiplicity of intersecting trajectories coming in and out of the collaborative flow. But the underlying stream linking multiple inputs, outputs and processes gains coherence at different levels of resolution.

But What Does It Mean?

Remixing and appropriation have a long and rich tradition in the arts. From traditional collage to DJ culture, rearranging pre-existing visual and sonic information in order to create new meaning has been a long-time preoccupation for artists of all backgrounds. Remix, as we now know it, comes from the contemporary cultural phenomenon of recombination and appropriation in popular music and DJ culture. Since the early stages of recorded sound in the late 19th century, technology has allowed individuals to reorganize the normal listening experience. With the arrival of easily editable magnetic tape in the 1940s and 1950s and the subsequent development of multitrack recording, such alterations became more common. Remix exists in a variety of forms with differing structures and emphasis in their relation to the original work. In its most simple form a remix is a song that has been edited to sound different from the original version. This can be achieved by rearranging the multiple tracks that constitute a song or by the juxtaposition of bits and parts of different songs. Similar to the collage, the remix can be made of the combination of original artworks with other materials.

Lessig has argued that digital remix is the 21st century literacy and is as fundamental as reading [8]. To copy and paste data as we surf the web while doing research or just looking for a recipe has become so ubiquitous that we hardly think about it. To copy is to separate a piece of information from any original context or meaning so that it might be free to function otherwise. To paste is to reinscribe, to place the floating sample into a new chain of signification.

The idea of remixing sonic and visual information has given birth to a plethora of sub-entities that inhabit the digital biosphere. From mashups to animated gifs, these events of ephemeral creativity prove that the joy of recombination rests not only in the final result of a specific mix, but in the performative aspect of remixing. The potential of endless reorganization for the emergence of new meaning is at the core of the idea of the remix. Appropriation and recombination are creation, following Paul D. Miller (aka DJ Spooky): "Give me two records and I will make you a universe..." DataRemix proposes that we have all become information DJs.

In both process and outcomes, the "datamades" resulting from DataRemix are envisioned to function analogously to Duchamp's "readymades." Their ultimate objective is to destabilize the framing narratives of data creation and representation in order to generate the possibility for new forms to arise in hopes of allowing us to see and know beyond what our instruments, algorithms, representational schemas and prevailing culture enable us to see and know.

Viegas and Wattenberg reflect that the power of artistic data visualization arises from artists "committing various sins of visual analytics" and directly engaging and guiding an audience towards a point of view. They remind us that even with dispassionate analysis as its goal, creating a visualization that is truly neutral is "generally impossible" and propose further exploration of the value of artistic explorations [10]. In this light, we propose to explore DataRemix as a mechanism for artistic approaches to engage empirical approaches in creating new ways of seeing and knowing.

References and Notes

* Based on a presentation given at VISAP'13, 13-18 October 2013, Atlanta, Georgia. The IEEE VIS Arts Program (VISAP) is a forum that encourages dialog about the relationship between aesthetics and visualization. The theme of VISAP'13 was Art+Experiment.

1. Gantz J.F. and Reinsel D. (2012). "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East – United States," IDC Country Brief. Online: <http://tinyurl.com/opxro4z>.

2. boyd, d. and Crawford, K. (2012) "Critical Questions for Big Data Information," *Communication & Society*, 15:5, pp. 662–679.

3. Boorstin, D.J. (1994) *Cleopatra's Nose: Essays on the Unexpected*, New York: Random House.

4. Naur, P. (1995) *Knowing and the Mystique of Logic and Rules*, Kluwer Academic.

5. Smith, J. (2006) *Charles Darwin and Victorian Visual Culture*, Cambridge University Press.

6. Malina, R.F., Strohecker, S., LaFayette, C. and Ione A. (2013) "Steps to an Ecology of Networked Knowledge and Innovation: Enabling New Forms of Collaboration among Sciences, Engineering, Arts, and Design." Online: <http://tinyurl.com/l3rssd2>.

7. West, R., Malina, R., Lewis, J., Gresham-Lancaster, S., Borsani, A., Merlo, B. and Wang, L. (2013) "DataRemix: Designing The Datamade Through ArtScience Collaboration," in *Proceedings of the IEEE VIS 2013 Arts Program (VISAP)*, Atlanta, Georgia.

8. Lessig, L. (2008) *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, London, Bloomsbury Academic.

9. Navas, E. (2009) "Regressive and Reflexive Mashups in Sampling Culture," in Sonvilla-Weiss, Stefan (Ed.) *Mashup Cultures*, Springer.

10. Viegas, F.B. and Wattenberg, M. (2007) "Artistic Data Visualization: Beyond Visual Analytics," in D. Schuler (Ed.): *Online Communities and Social Computing*, HCII 2007, LNCS 4564, pp. 182–191, 2007. Springer-Verlag Berlin Heidelberg.