Reconsidering Database Form: Input, Structure, Mapping

By Matthew LeMay

No. 35 – 2005

Abstract

In this essay, I argue against Lev Manovich's theorization of the database, and resultant critiques of mapping art. Suggesting that database form necessarily involves intricate interrelations of data in a rigid, predetermined structure, I propose that the general divide between content and form proposed by Manovich is at least an oversimplification, and at most erroneous. I take issue with Manovich's designation of mapping art as "anti-sublime," suggesting that it is instead the inputting of data into a database that can be considered "anti-sublime" in Manovich's terms.

In his essay "The Anti-Sublime Ideal in New Media," Lev Manovich suggests that mapping art suffers from a fundamental divide between form and content. Proposing, as do many new media theorists, that data contained within a database is almost infinitely malleable and manipulable, Manovich posits that "since there are endless ways to map one data set onto another, the particular mapping chosen by the artist often is not motivated, and as a result the work feels arbitrary"[1]. While Manovich is correct in suggesting that a computer can map data from one form to another, his theorization of the database does not account for the complex interrelations between data and the database it inhabits. Instead, I suggest a more complicated theorization of the database that accounts for the necessary and interrelated coherency between data and the structure into which it is inputted. Furthermore, I suggest that such systems of coherency alter, limit, and direct any user's experience of the database. Following such a theorization, the question of mapping art becomes not one of impartially mapping data, but of translating entire systems of data. What is called for here is a critique of mapping art in which the divide between content and form is not assumed, but rather reimagined in terms of reworking and translating specific relations and systems of content and form.

In his essay "Database as a Genre of New Media," Manovich posits the database as the quintessential postmodern cultural form. After a fairly detailed technical

discussion of specific database architectures, Manovich suggests that the structural complexity of the database is not manifest in the user's experience:

> New media objects may or may not employ these highly structured database models; however, from the point of view of the user's experience a large proportion of them are databases in a more basic sense. They appear as a collections [sic] of items on which the user can perform various operations: view, navigate, search. The user experience of such computerized collections is therefore quite distinct from reading a narrative or watching a film or navigating an architectural site.[2]

Clearly, Manovich is correct in suggesting that a user's experience of a database differs from his/her experience of a narrative or a film or an architectural site. However, Manovich's suggestion that the complex structural form of the database is not present in the user's experiences it is quite problematic. In suggesting that the user simply experiences the database as "a collections [sic] of items," he does not account for the complex structures by which database form necessarily orders, categorizes, and directs these user's experience of these collections. More importantly, he does not consider the extremely complicated, often codependent and/or causal, relationship between structure and data in a database. In arguing against Manovich's theorization of the database, my goal is not simply to suggest that Manovich does not take into an account an "expert" or "system engineer" viewpoint – rather, I wish to suggest simply that Manovich is erroneous in assuming that the complexity of database structures does not inform the experience of the average user.

One key conceptual omission in Manovich's text is that of data input. Manovich, in his reduction of multiple and complex database forms into simple collections of items, does not address the fact that any database must be planned, programmed, and inhabited by compatible data. Here, Manovich's brief discussion of database complexity, summarized in the statement "a database is anything but a simple collection of items," proves insufficient[3]. Databases are not only complex as abstract concepts – they are specifically complex in the ways that they relate, combine, and organize data.

Discussing the prevalence of database form, Manovich addresses the Internet as a primary example of the centrality of database logic:

> Where database form really flourished, however, is on the Internet. As defined by original HTML, a Web page is a sequential list of separate elements: text blocks, images, digital video clips, and links to other pages. It is always possible to add a new element to the list – all you have to do is open a file and add a new line. As a result, most Web pages are collections of separate elements: texts, images, links to other pages or sites.[4]

Though Manovich is correct in identifying a web page as a "sequential list of separate elements," he ostensibly does not directly define it as a database. In fact, as a largely unorganized collection of information, a web page is antithetical to database form. In order for a database to function as Manovich suggests – for its data to be coherent, searchable, and manipulable – the information in the database must be meticulously organized in a manner befitting the data that is to be input. As Manovich himself suggests, information is input into a website simply as "a new line," a linear and nondifferentiated mode of input that has more to do with a multimedia narrative than it does with a database.

Using Manovich's example of a CD-ROM encyclopedia, importance of data input and organization in the user's experience becomes clear. If, for example, one wishes to search for an image of George Washington, one would likely enter the text "George Washington" into a text box, which would in turn prompt the encyclopedia program to seek out articles and images containing the term "George Washington." However, if an image of George Washington appears in the encyclopedia's search results, it is not because the text "George Washington" is in any way being mapped or translated into a visual pattern that can be searched in the encyclopedia's collection of images. Instead, such an image would be returned if and only if the text "George Washington" had been manually connected to the image of George Washington in a database. Furthermore, such a database would have been necessarily designed in such a way as to connect a text field to an image, so that the search program would know how to both search, retrieve, and display the relevant data.[5]

This more directed and limited theorization of the search function flatly disproves Manovich's suggestion that the website, as a "collection of separate elements," exemplifies database logic. Any functional multimedia Internet search engine such as Google Image Search or Yahoo! Video Search must compile a database that relates a website's multimedia content to its textual content. For example, a Google image search of the phrase "George Washington" returns not only images of George Washington, but also an unrelated political cartoon from an anti-Semitic website that happens to include a quotation from George Washington in its text. Taking this example further, one could place the image of an apple alongside text reading "orange," and the image would be searchable only by the term "orange." One could even go so far as to place a digital image of the word "apple" alongside the text "orange," and the result would be exactly the same. Thus, functional database logic can be seen as antithetical to the "collection of separate elements" described by Manovich – it is, in fact, located primarily in the complex connections between elements.

In her chapter "Databases, Data Visualization and Mapping," Christiane Paul echoes Manovich in her discussion of the database as "an essentially dull affair, consisting of discrete units that aren't necessarily related"[6]. However, Paul goes on to suggest

that "the power of databases consist in their relational potential, the possibility of establishing multiple connections between different sets of data" [7]. While both Paul and Manovich seem to suggest that such connections are somehow innate to the database, these connections generally have to be either input by a sentient user, or generated by some kind of data analysis carried out by the computer. This is an important distinction, as it speaks to two different perceptual modes of data analysis – one in which connections are made in terms of human perception and knowledge (for example, the association of the picture of George Washington with the text "George Washington") and one in which connections are made by similarities that are understandable to a computer but might seem irrelevant or impossible to human reason (respectively, proximity of image and text in a website and packet analysis such as that executed by Galloway's *Carnivore*).

These two categories not only construct very different organizations of data, but also bear very different relations to temporality. The inputting and interrelation of data by a sentient user generally requires a good amount of time, and usually results in more static data sets (for example, a CD-ROM encyclopedia). The inputting and interrelation of data by a computer, according to specific criteria of data analysis or proximity, generally creates more dynamic data sets (for example, search engines that use "robots" to constantly update their databases). Contrary to Manovich's suggestion, I would argue that the distinction between these two modes of data input is extremely important and rather straightforwardly manifest to the user. The frustration experienced by most users searching the Internet, compared to the relative ease of use of a CD-ROM encyclopedia's search function, speaks to the important distinctions between these two modes of data inputting in relation to a user's experience of the database.

Such a distinction further complicates Manovich's suggestion that the information within a database can be infinitely manipulated and rearranged by a user. Indeed, in most cases, the information in a database can only be searched, modified, and rearranged in the terms set forth by the structure of the database itself. If, for instance, I wanted to use a CD-ROM encyclopedia to search for an image of George Washington, I likely could not do so by inputting a text search for physical characteristics ("tall," "white hair," etc). Furthermore, I could not use such a database to look for similar images based on any kind of perceptible analysis of the image itself. While expensive, time-intensive feature recognition software can somewhat reliably calculate similarities between images, even this software obeys different rules from those of human perception. In fact, most computer programs that input and organize data on a large scale, such as the aforementioned internet search engines, work by extrapolating some trace of human input, such as the proximity of elements in a web page, in a way that is readable to a computer.

This more complicated and rigorous view of database structure necessitates a thorough reevaluation of mapping art, as well as the practice of "mapping" in

general. In her chapter, Paul bases much of her argument about mapping art on the assertion that "information itself to a large extent seems to have lost its body, becoming an abstract 'quality' that can make a fluid transition between different states of materiality"[8]. Manovich, in "The Anti-Sublime Ideal in New Media," suggests that "by representing all data using the same numerical code, computers make it easy to map one representation onto another"[9]. However, taking into account a more detailed theorization of the database complicates the suggestion that mapping as such is "easy." Indeed, one could suggest that mapping art must take into account not just the content of the database it maps, but the organization of that content in the database, and the relationship between content and organization as well.

Once again describing mapping art as an extension of data'a malleability, Manovich suggests that the ease of translation between different mapping forms is the central problematic of such art:

> Since usually there are endless ways to map one data set onto another, the particular mapping chosen by the artist often is not motivated, and as a result the work feels arbitrary. We are always told that in good art form and content form a single whole and that content motivates form. Maybe in a good work of data art the mapping used must somehow relate to the content and context of data – although I am not sure how this would work in general.[10]

Perhaps what is most striking about Manovich's description of "good art" is that it could just as easily apply to "good databases." As I have suggested, database design necessarily involves the interrelation of data and the structure of that data within the database. Furthermore, it is consideration of the nature, amount, and type of data to be input that facilitates the design of a functional database structure. To put it another way, in database design, "content" often does motivate "form."[11]

This interconnection between form and content in database design greatly complicates the idea that, as Simanowski puts it in a discussion of mapping art subgenres and critiques, "mapping preserves the structure of data in an exact way and mirrors it in a new form"[12]. Indeed, if we are to accept that the structure of a database is in fact determined largely by the form and type of the data that inhabits it, we must reconsider the notion that mapping art merely changes form and preserves content. One way to theorize this is to suggest that, rather than mapping the same content onto a different form, what is being altered in mapping art is often the very relations between form and content.

In a discussion of mapping art forms, Simanowski identifies three subgenres of mapping art; those which locate visual pleasure in the "beautification of data," those which locate visual pleasure "not [in] the transformation, but [in] the disclosure of information," and those that map the "sensualization of an idea"[13]. I suggest that the first two of these subgenres may also be described in terms of their relationship to

database structure and input. The former, in which data is aestheticized, is not simply an impartial translation of content, but rather a reworking and/or reinterpretation of the entire system of relations between form and content. One could suggest that such a reworking is necessary to any form of mapping in which data as presented in a form greatly different from that in which is was input. The latter subgenre, conversely, preserves the structure of the database, and simply represents the data in more or less the same form and structure in which it was input. To put it another way, a distinction can be made in mapping art that maintains the structural relations between the data as it is input and as it is displayed, and mapping art that ruptures or reconfigures these relations.

John On and Futurefarmers' *They Rule* is a primary example of mapping art in which the data is output more or less precisely as it is input. Building on a static collection of data harvested in 2002, *They Rule* visualizes data using what one can safely assume is exactly the same connections and the same forms present in the original database. While a visual "interface" is provided for examining the data, this interface simply renders with graphics what might be just as visible in a tabular representation of the database. *They Rule*, like many such pieces of mapping art, relies upon data that has been structured input by a human user in order to facilitate certain connections. In this particular case, these connections are mobilized to make a political statement.

It is what Simanowski describes as data beautification that Manovich primarily addresses in his discussion of mapping art as anti-sublime. Manovich suggests that such art is anti-sublime since it transforms "invisible and 'messy' phenomena [...] into ordered and harmonious geometric figures"[14]. However a database is, by design, neither 'invisible' nor 'messy.' Certainly, the dynamic described by Manovich can be located in new media – but I would suggest that it is not so much in mapping as it is in the construction of databases[15]. Given that a database necessitates the designation of specific fields, criteria, and interrelations, the inputting of information into a database can be described very much in the terms Manovich uses to describe mapping. Certainly, new media does facilitate certain objects and systems, like the Internet, that are "messy" and seemingly infinite. However, by the time most users are navigating the Internet, they are already navigating an organized database, structured to facilitate certain connections. A compelling argument could be made, then, that database logic is, itself, anti-sublime.

What, then, do we make of mapping art that attempts to sensualize or aestheticize data that is, itself, enmeshed in an anti-sublime form? For starters, we must reconsider the notion that such mapping impartially translates content, as the content it is translating is largely inseparable from the form in which it is stored and organized. The content of a database is coherent in relation to its form, and vice versa. Far from suggesting, as Manovich does, that the choice of data translation is unmotivated or irrelevant, we must examine how processes of translation

reorganize data, and how this process relates to the organization of the data that is being translated.

Calling for such an examination might account for why several pieces of data beautification mapping art seem so distant and opaque. Greyworld's *The Source*, which "translates" stock market information into a series of floating balls, is a primary example of mapping art in which the organizing principles of the work does not engage with the organizing principles of the data. Rather than responding to the complex and interrelated ways in which stock market data is often translated, the motion of the floating balls is based exclusively on the degree of trading activity. Thus, it is not so much the data that is discarded by the piece, but the organizational systems that structure the data.

This is not, however, to say that we should strive for art that simply recreates the organizational principles of the data it translates. Golan Levin's *The Secret Lives of Numbers*, a piece that seemingly has more in common with *They Rule* than it does with *The Source*, presents a graph of every number from zero to one million. However, the graph is not arranged by numeric value, but rather by the "popularity" of each number, a statistic gathered from a Google search of each individual number. By culling such data from a Google search, *The Secret Lives of Numbers* exploits the ability of computer-generated databases (such as Google search) to rapidly seek out traces of human intention and causality on a scale impossible for manual data input. While this data is, obviously, not fully reliable, this is largely irrelevant to the piece. *The Secret Lives of Numbers* is compelling not so much because of the data it presents, but rather because of the way it reconfigures relationships between data. Thus, *The Secret Lives of Numbers* can be seen as engaging with both the transmission of information and the sensualization of an idea, as the systems of organization being engaged and reconfigured are far from impartial, but rather highly directed by the artist.

Mark Napier's *Black and White* utilizes a similar reconfiguration of data relations, but does so in a way that engages with both the aestheticization of data and the sensualization of ideas. Ostensibly, *Black and White* uses the complex *Carnivore* system designed by Alexander Galloway to read the binary code of CNN.com and construct a dynamic black and white graphic based on its reading. However, such a reading does not, in fact, require the *Carnivore* system at all. The data that is being culled, much like the image being presented, is not particularly impressive in and of itself. What is interesting, instead, is the way that this piece reconfigures the relations of data from the text of CNN.com, to binary code, to black and white visualization. While one could critique Napier for ineffectively using the *Carnivore* technology (as was my initial reaction to the piece), one can also read this ineffective use of technology as a rejection of data-as-data, insistently shifting emphasis onto the relationships between data and the systems that organize them. That is to say, one could suggest that an effective piece of data aestheticization art must

ultimately reject the importance of data-as-data, as the translation of data from the form in which it was input (here, the text of CNN.com) to another form (here, the black and white visualization) necessarily involves a loss of the system's original coherency.

The success of *Black and White* as compared to other pieces of data beautification art suggests that, perhaps, the impartial "mapping" of data without the consideration of the systems of organization being translated rarely, if ever, makes for good art. *Black and White* is compelling largely because it refuses to locate meaning entirely in its content (the data) or its form (its visualization of that data), in effect prompting the viewer to consider the relationships between the data being presented and the form in which it is presented. Manovich is correct, then, in suggesting that the divide between content and form is problematic, but he is incorrect in suggesting that this divide is inherent to data and databases. Denaturalizing the database as cultural form, and acknowledging that any functional database must be engineered to accommodate a symbiosis of form and content, means that we can no longer accept Manovich's problematic as central to mapping art. Instead, we must call for mapping art that to some extent employs the "sensualization of ideas" described by Simanowski, and engages with the relationships between form and content. Utilizing this more complex theorization of the database, such an engagement does not seem at all as far-fetched as Manovich suggests.

A reexamination of mapping art and its relationship to the sublime(s) and the anti-sublime is called for on a scale far beyond this paper, but I believe that a reevaluation of the notion of data and the database must be the starting point for any such analysis. The seemingly oppositional currents of comprehensibility and aestheticization in mapping art are not simply issues of how data is mapped, but rather whether the organizational principles inherent in the data are maintained and/or reworked, and how they are maintained and/or reworked. Rather than simply conceding that a divide between form and content is inherent to mapping art, we must insist that, if the database form is as prevalent as Manovich suggests (and I would argue that it is), so too is the interconnectedness between data-as-content and structure-as-form. We similarly must acknowledge that any fairly drastic mapping procedure necessarily involves a loss of the coherency that exists between the data as originally input and the structure into which it was originally input. Acknowledging such a loss of coherency means seeking out meaning not in the data presented by mapping art, but rather the ways in which systems of data and meaning are reconfigured by mapping art.

Notes

1. Manovich, Lev. "The Anti-Sublime Ideal in New Media." 2002. <http://www.chairetmetal.com/cm07/manovich-complet.htm> (12.19.2005).

2. Manovich, Lev. "Database as a Genre of New Media." 1997. <http://time.arts.ucla.edu/AI_Society/manovich.html> (12.19.2005).

3. Manovich, 1997.

4. Manovich, 1997.

5. In using examples such as this, I have attempted to refrain from an elaborate discussion of database engineering and data inputting. Instead, I have attempted to explain, in very simple terms, how the form and content of databases are interrelated and, in many ways, determined and limited by the choices made by the database's designer. Whether or not these choices are directly apparent to the user, I suggest that they greatly affect the user's experience of the database.

6. Paul, Christiane. "Databases, Data Visualization, and Mapping," in: Paul, Christiane. *Digital Art*. Thames and Hudson, 2003. 174-189. <http://dev.stg.brown.edu/projects/netart/Chr-Paul-DigitalArt-Mapping.doc> (12.15.2005).

7. Paul, 2003.

8. Paul, 2003.

9. Manovich, 2002.

10. Manovich, 2002.

11. Commenting on this paper, Francesco Ricardo points out the difficulty of using the terms "form" and "content": "Lemay's use of 'form' is deprecated in database language, in which the official terminology is 'schema', or 'scheme' (for its 'structure') and 'data' (for its 'content'), a database is composed of one schema and any amount of data." The slippage of these terms is, indeed, problematic. What I am calling for, then, could be described as a provocative translation between the schema-data relationship of a database (using Ricardo's terminology) and the form-content relationship of the resultant art/visualization, without necessarily collapsing these terms onto each other. At the very least, I suggest that the acknowledgement of these relationships being altered is necessary.

12. Simanowski, Roberto. "Mapping Art as Cultural Form in Postmodern Times." 2005. <http://www.brown.edu/Courses/GM/GM144-2005/52-lecture-mappingart.doc> (12.15.2005).

13. Simanowski, 2005.

14. Manovich, 2002.

15. Again, I am not suggesting that the user must be knowledgeable of the specifics of database programming and data inputting, but rather that these concepts inform the user's experience of the database.