REFORMATTING INFORMATIC HIERARCHIES:

glitch art excavations with the national institute for computational sciences

by Amy Szczepanski and Evan Meaney

introduction

Let us, for a moment, separate data from information. How can digital communication address this divide? In his 1896 text, *Matter and Memory,* Henri Bergson makes specific reference to the development of data and its ability to be transformed into something we might call viable. "Little by little it comes into view like a condensing cloud; from the virtual state it passes into the actual; ... to imitate perception." [1] The dual evolution of data (virtual) and information (actual), has become so entangled that most people consider the terms to be synonymous, and therein exists the informatic problem that the art community can help answer. Data is not information, and to confuse these terms is to support a hierarchy, detrimental to our cultural future.

Sean Kanuck, National Intelligence Officer for Cyber Issues (USA) explores the link between data and information as it pertains to national security. In July 2011 he gave an address where he put forth the following, "As we enter the era of an 'internet of things' which will all communicate with each other—increasingly without any human mediation—technical standards and governance issues will take on new importance too." He makes this specific when he says, "the electromagnetic phenomena that are stored in hard drives, transmitted over fiber-optic cables, or relayed between microwave towers have no inherent value to human welfare. They must be first collected and then processed to become information." [6] For him, the purpose of data is to bring empirical, non-contextual, bytes closer to the realm of, specifically human, understanding. Thus, information is data when made readily useful for us; really viable information allows us to mentally edit out that it was ever data to begin with. It becomes epistemological shorthand—a fact without a predicate.

Opening up a JPEG file and viewing it asks the computer to interpret pixel data (bytes), developing those bytes into an informational rendering (the picture) for users to interact with as they desire (Figure 1). Most of the time this workflow happens without incident; we receive our information, disregard the data, and carry on satiated with knowledge. However, sometimes a glitch occurs, and when that happens, the workflow's nested hierarchy (knowledge > information > data) is immediately scrutinized. So, it would seem that in an informatic system predicated on a knowledge-giving interface, the only way to become acquainted with the predicate data is by breaking the interface itself through a glitch. For the JPEG, this looks like a visual encoding error, or an errant height/width value—more abstractly, however, a glitch can be

seen as the missing piece of a jigsaw puzzle which reminds you that the constructed image is, in fact, a puzzle and not just an image.

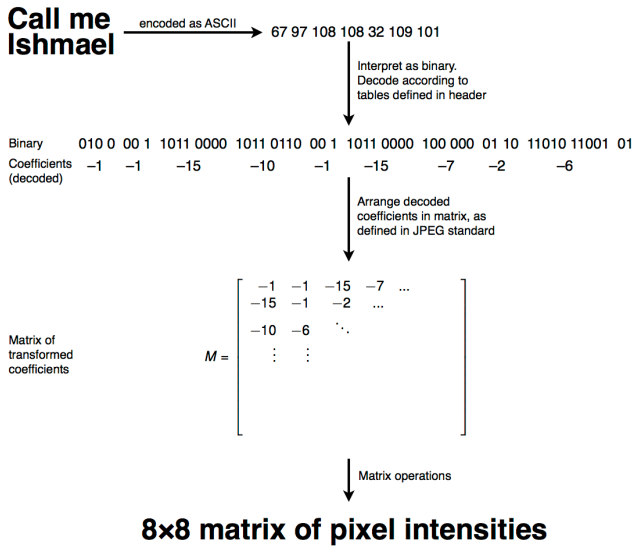

Fig. 1 /
Text, ASCII, Pixel Workflow

New media artists and venues are, as of late, exploring 'glitch' as both a product and a process. This ranges from practitioners like Cory Arcangel, Jon Cates, and JODI to conferences like *GLI.TC/H*, which this year had events in the United States, the Netherlands and the UK. Glitch art is coming into its own relevance; unfortunately, this evolution has gone somewhat unaided by the informatic/archival sciences, even though both communities share common research methods and exigencies. We aim to bridge this gap in the hopes of producing work which will not only be interesting to both camps, but will hopefully unite them in discussion. In many ways glitch art is a common space where the arts and sciences may find productive, yet deconstructive, overlap.

actively_introducing_problems

In the summer of 2011, our team received support from the eXtreme Science and Engineering Discovery Environment to develop art-based inquiries into the data/information divide. This afforded us time on the Nautilus supercomputer, operated by the Remote Data Analysis and Visualization Center of the National Institute for Computational Sciences, housed at the Oak Ridge National Laboratory. Soon thereafter, we began to discuss basic interfaces which we might interrupt. Specifically, we were looking for a media architecture that would accept specific encoding errors, thereby cultivating a glitch, which would separate the user from the sought-after information and confront him or her with less mediated (and less optimized) data structures. We explored graphic encoding standards, identifying the JPEG standard as a system

susceptible to our intended processes. We then began to conduct small scale excavations on JPEG images (Figure 2) to understand exactly how this encoding system mediates abstract data into familiar, visual information.

Fig. 2 /
JPEG Test 15
2011

The question remained, however, as to how we could create something which not only produced a technical glitch, but contextualized it beyond a dismissible "failure." This had two outcomes:

We developed a body of 2D work: artistic studies of intentionally 'mis'-mediated textual data which we have titled, *NULL_SETS* (*N_S*, henceforth). Using a decoding/encoding process that will be illustrated in the next section, *N_S* takes on the project of mediating canonical/essential corpuses of human-language text (data) into visual information by rendering them, non-symbolically, as JPEG images. The computer-to-human translation, which normally expects computer-oriented code (binary/hex/ASCII) to produce human-oriented visuals, explodes at the introduction of human language. In other words, we used the text of *Hamlet* as byte data—rendering a JPEG image that visualizes all of the words in Shakespeare's play, but none of the information.

Conceptually, we came to realize that the same hierarchy which promotes information while editing out data is inherent in the reception of any break, schism, or hack. A glitch is, traditionally, a pejorative term—glitches need to be overcome, not enjoyed. Glitches show the system's failings. *N_S* challenges that assumption; demonstrating that, as any researcher will tell you, systemic failure is a large part of systemic understanding. As Isaac Asimov wrote, "The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny….'" Thus, a glitch demonstrates a system's operation (in our case, language) in a way that expands the system's parametric discourse; suggesting, perhaps, new metrics, methodologies, and workflows. In this way we can see a glitch not as a failure, but as an new, discursive result, like fitting a round peg into a square hole. In a way, it is an achievement.

The JPEG standard defines a system for encoding color images with compression [5]. These files can typically be viewed as having three parts: a header, the data that describes the pixels in the image, and a two-byte footer.

The header contains structural data about the image, including the dimensions, the forms of encoding and compression being used, and a series of look-up tables for mediating the compression algorithms. Information in the header is flagged with two-byte markers that begin with the hexadecimal quantity 0xFF and whose second byte indicates the feature being defined. The marker is followed by two bytes that give the length of the table or feature being described, followed by the pertinent information. A typical JPEG image header will have a length of hundreds of bytes, encoding roughly a half-dozen attributes of the method. The footer, on the other hand, is always the two-byte quantity 0xFFD9, which represents the end of the image. For most JPEG-encoded images, the bulk of the content falls between the header and the footer and is encoded as a series of bits (Figure 3). The image data interpreted by the JPEG decoding algorithm includes the average intensity and average color over blocks of 8-by-8 pixels as well as encoded coefficients that describe how the intensity and color vary over each block.

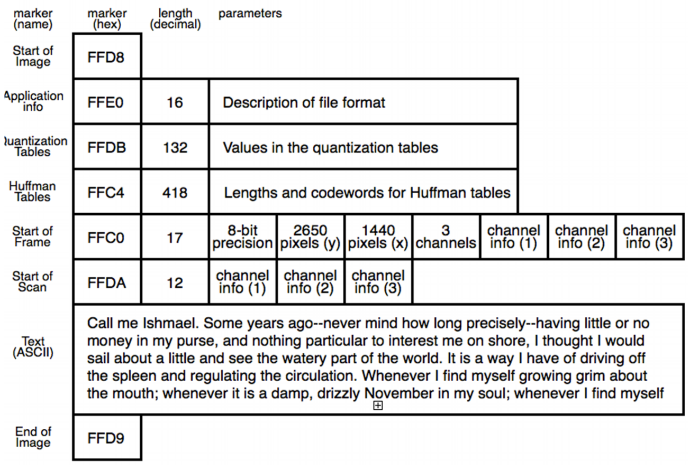

Fig. 3 /
JPEG Structure

We constructed a 607-byte JPEG header for an 2560-by-1440 pixel image encoded with the discrete cosine transform and non-differential, Huffman coding. We also specified the default look-up tables for image quantization and run-length encoding. The two-byte footer is the quantity 0xFFD9, representing the end of the image. Next, we used the Unix utility cat to concatenate our header file, a text file in rich text format, and our footer. Thus, data that had previously been intended to be decoded as ASCII characters will now be read as if it represents quantities defining the

pixels in an image. We gave the resulting file a name with a .jpg file extension and then viewed it with standard image viewing tools. Due to our original texts being encoded in rich text format and consisting of ASCII-encoded text (which never contains 0xFF), we did not need to worry about the byte 0xFF erroneously triggering a marker in the data segment; this would be a real concern if using an arbitrary binary file instead of text.

Once this workflow had been stabilized, we began to introduce and compare text-bodies from different disciplines within the arts and sciences. These bodies, when translated through the *N_S* / JPEG process, became both the subjects of, and the titles for the individual images. Using the text from the Qur'an produced our image, *Qur'an Test*(2011) (Figure 4). Appropriating the HTML from a Google search produced *Google Search Results for Google Test* (2011). Using this process, any text may be both preserved (as byte data) and glitch-mediated (as an image) in a JPEG translation; creating a new means of comparison and aesthetic exploration.

Fig. 4 /
Qur'an Test
2011

name_a_glitch_and_own_it

Not surprisingly, the vocabulary of several fields comes to use for describing the process of creating the *N_S* images. Archaeology lends us the term "excavation," which connotes the practice of unearthing layers to make knowable what was previously unrevealed. This word, however, is an imperfect match for this exploration. The most crucial difference is that the excavation of the archaeologist is, by its very nature, a destructive process. Digging into a site changes it in ways that can not be reversed. Even with the most careful records of the excavation process, one could not return all of the artifacts, stones, and soil to the ground in the exact way that they were before the dig began. On the other hand, *N_S* is not based on destruction of files but rather on their translation and remediative creation. In stitching together the header, text, and footer a new object is created. Unlike the archaeological dig, this process is completely reversible. This generalizes to other digital techniques that rely on fully deterministic processes—as long as the workflow is sufficiently documented. So, in this way, *N_S* is, descriptively, more like a fibers project. Following a pattern with the correct yarn, one can recreate a stitch. Following an algorithm with the correct bytes, one can retranslate a glitch.

In *N_S*, our users are, for perhaps the first time, not editing out their limited interaction with the predicate data, but asked to contemplate said data as information. The forms created by the *N_S* glitches are akin to the work of abstract expressionist painters (Pollack, Hofmann, Still, Rothko) who remind their audiences that art isn't always about accurately representing a subject, but exposing the means by which that representation interfaces with human senses. The individual 'mis'-mediations of our project pose questions of user expectation. For example, our remediation of Melville's Moby Dick, *Moby Dick Test* (2011), contains the same characters, chapters, scenes, and narrative as a leather-bound first edition—but, now encoded as an image (Figure 5). The user is kept at a distance from contextual, 'leather-bound' information, removed (and not supported) by any humanly recognizable graphical interface. *N_S* demonstrates a data-as-information system where the unrefined, inhuman predicate is presented as finished work. To better understand the evolving compromises of computing and computational art, we should be willing to appreciate data as an unmediated companion to, and not simple subject of, information.

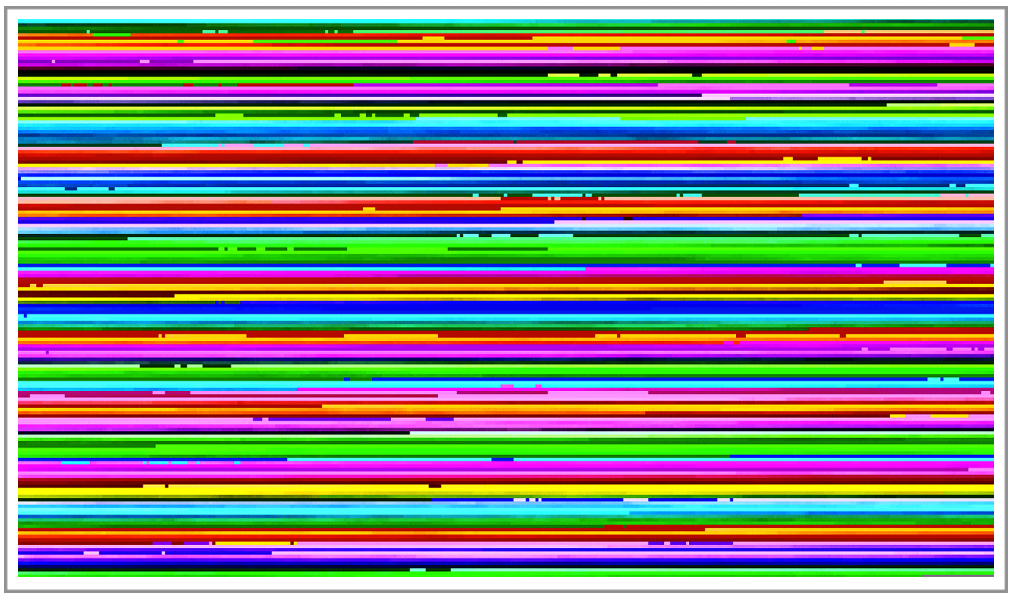

Fig. 5 /
Moby Dick Test
2011

These issues of understanding data, information, and knowledge through hierarchical change raise questions about digital provenance. While the notion of provenance has, historically, been defined as tracing the integrity of the ownership of an object, there is also a counterpart in the notion of digital integrity. This question cuts across the divide between data and information: Which process has a scientist gone through to generate data that was "born digital" rather than collected from experiments or sensors, and how can the scientific community examine these processes to repeat the experiment? What stops a database administrator from changing records about parking tickets? How do we react to a retailer using photo editing software to stitch models' heads to digitally constructed bodies in its advertising? There is no inherent reason to privilege a certain collection of zeros and ones or to frown at its

manipulation. It is only when we examine the information represented by these bits that we enter the ethical realm of asking what it means to change the representation of knowledge. Immediate mediations of data, like *N_S*, allows us to, not only make an effort to preserve and present data-as-information, but to remind others that, while information and knowledge can be very useful, they are always positioned, by humans, in a biased context.

failing_checksums

As we saw in the last section, the divide between data and information, between fact and fabrication, as problematized by the *N_S* images, functions as a cross-disciplinary tent-pole. Something under which we can all gather, bringing our lexicons and specific understandings of the general problem of digital epistemology. For example, while the actual USA Patriot Act has been informationally positioned by artists, politicians, activists, and hackers alike, our *N_S* remediation, *The Patriot Act Test* (2011), makes an effort to espouse a modicum of empiricism when dealing with its subject matter. We position the work only as far as the bytes will allow us. This tactic allows others, no-matter what their discipline's attachment to the subject matter, to be momentarily removed from the biases of hierarchical information, appreciating data apolitically —returning to the well. Unfortunately, the well itself is not a completely renewable resource.

To get a better understanding of the abstract scarcity, and subsequent hierarchical privilege of information, we can turn to a pretty big example: the Compact Muon Solenoid at CERN's Large Hadron Collider. "The data volume recorded per second in CMS corresponds to 10,000 sets of the Encyclopedia Britannica. The data recorded during the 10-to-20 years of the LHC lifespan will be equivalent to all the words ever spoken by humankind." [3] According to one of the researchers on the project [4], the LHC produces so much data that there is no way to store it all—let alone examine it all. Physicists have written software so that the LHC will only save data that is likely to be "interesting."

Thus, the divide between data and information will be stretched and, ultimately, severed. Ironically, there will be so much archived information (so many Encyclopedia Britannicas) that information which had previously been privileged will, once again, be relegated to data that must

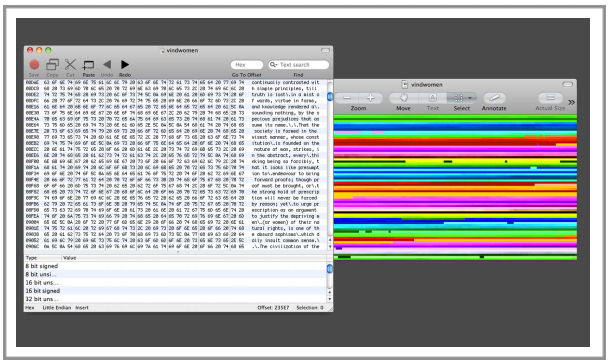

Fig. 6 /
A Vindication of the Rights of Woman
Comparison

be rescued from a noisy channel. The excavations in *N_S* give body to this reminder, visually reverse engineering information back into data (Figure 6), and affording its users the option to momentarily escape the social biases of perspectival knowledge.

reverse_engineering_data

There is a serious difficulty in objectively discussing the mediation of data and the subsequent archival hierarchy of information, as this workflow is both hegemonic and not localized to any one discipline. Researchers, artists, designers, politicians, chefs; everyone needs to derive knowledge from information, and have faith in that knowledge. For the sake of our non-infinite memory, we need to omit data to function. Projects like *N_S* are not meant to collapse the hierarchical system—but rather to serve as artful reminders of this systemic need. Individual works from *N_S* (Figure 7) are interested in reflecting the multiple levels of inability derived from the privileged state of what we (think we) know. Ultimately, the *N_S* images performed on the JPEG standard are, in themselves, hackable analogs for the verification process of any new experience, showing that we must take apart the structures in our epistemological workflow just as we dismantle the structures in our encoding paradigms. Only then can we see how both the interface and the predicate mirror our responsibility (even an abstract responsibility) to the understanding and faith of communication and archivism.

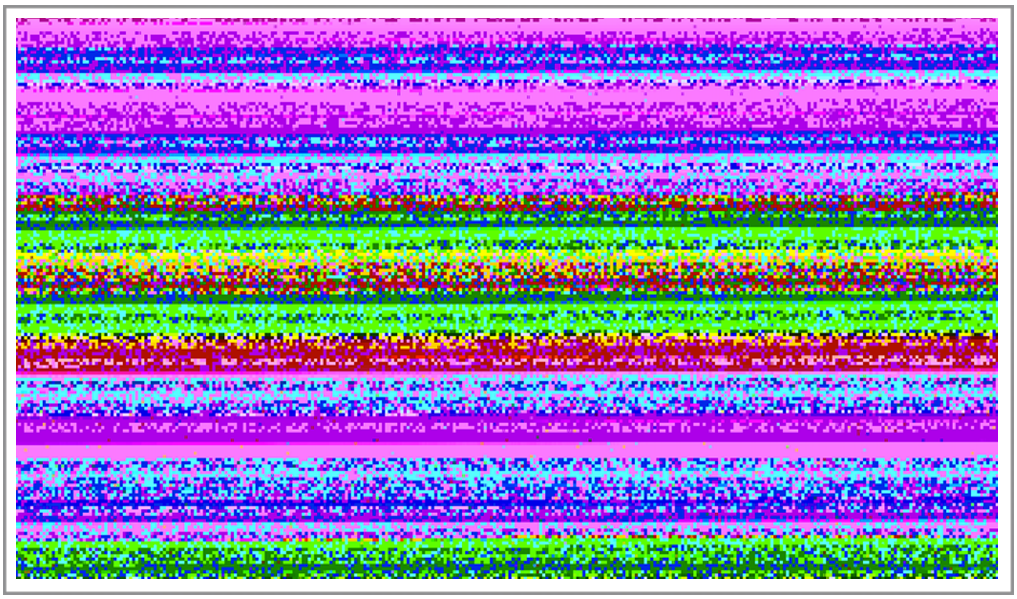

Fig. 7 /
Genome Test
2011

The ever-expanding number of works from *NULL_SETS*, like the Melville remediation, are not targeted to open up an eponymous conversation about, for example, Moby Dick. Instead, they are a bit like Derrida's appropriation of the Pharmakon. The digital archiving of information has, out of necessity, moved into a malleable space. One that we mistake to be immutable, empirical, and unbiased. *N_S* disconnects data from information and offers that glitch as an aesthetic/conceptual countermeasure, demonstrating that digital archivism is both the cure for, and the poison to, our future knowledge—that we are imperfect beings attempting to make a perfect memory, and in so doing, our failures become part of the process, incorporated, valuable, and worthy of our study.

citations.txt

[1] Bergson, Henri, "Matter and Memory" translators N. Margaret Paul and W. Scott Palmer. Dover Publications. 1912.

[2] Derrida, Jacques, "Dissemination", translated by Barbara Johnson, Chicago, University of Chicago Press, 1981.

[3] Hollingsworth, Matt, "Data Acquisition and Computing at the Compact Muon Solenoid Experiment" http://www.ornl.gov/ornlhome/calendar/calendar_full.cfm?Id=4943

[4] "2008 UTK Physics Newsletter" http://www.phys.utk.edu/news/2008/news_09102008_lhc.html

[5] International Telecommunication Union, Terminal Equipment and Protocols for Telematic Services: Information Technology — Digital Compression and Coding of Continuous-tone Still Images — Requirements and Guidelines, *http://www.w3.org/Graphics/JPEG/itu-t81.pdf*

[6] Kanuck, Sean, "Cyber Security Frameworks," Keynote address at CENS-GFF Workshop on Cyber-Security: The Geostrategic Implications of Cyberspace, S. Rajaratnam School of International Studies, Singapore, 18 July 2011.