PIIMRESEARCH

THE NEW SCHOOL

PIIM IS A RESEARCH AND DEVELOPMENT
FACILITY AT THE NEW SCHOOL

Big Data Visualization and Knowledge Discovery through Metapictorial Modeling

AWARD NO. FA8750-12-2-0325

WILLIAM M BEVINGTON
MARCH 14 , 2014

This work was supported by the Defense Advanced Research Projects Agency (DARPA) Award No. FA8750-12-2-0325. Any design guidance, opinions, findings, conclusions, or recommendations expressed herein are those of the author and do not necessary reflect the views of DARPA. (This is a newly illustrated and heavily revised version of an earlier submission)

INTRODUCTION

This paper presents a visualization theory concerning *Big Data.* One high-level objective for processing big data is to render models that will facilitate *Knowledge Discovery.* This requires investigation into the *characteristics* of what we term *Big Data,* as well as *considerations* into various types of *visual representations* for big data. I argue that a promising modeling option toward achieving knowledge discovery from big data is through what may be termed *(Synthetic) Metapictorial* renderings.

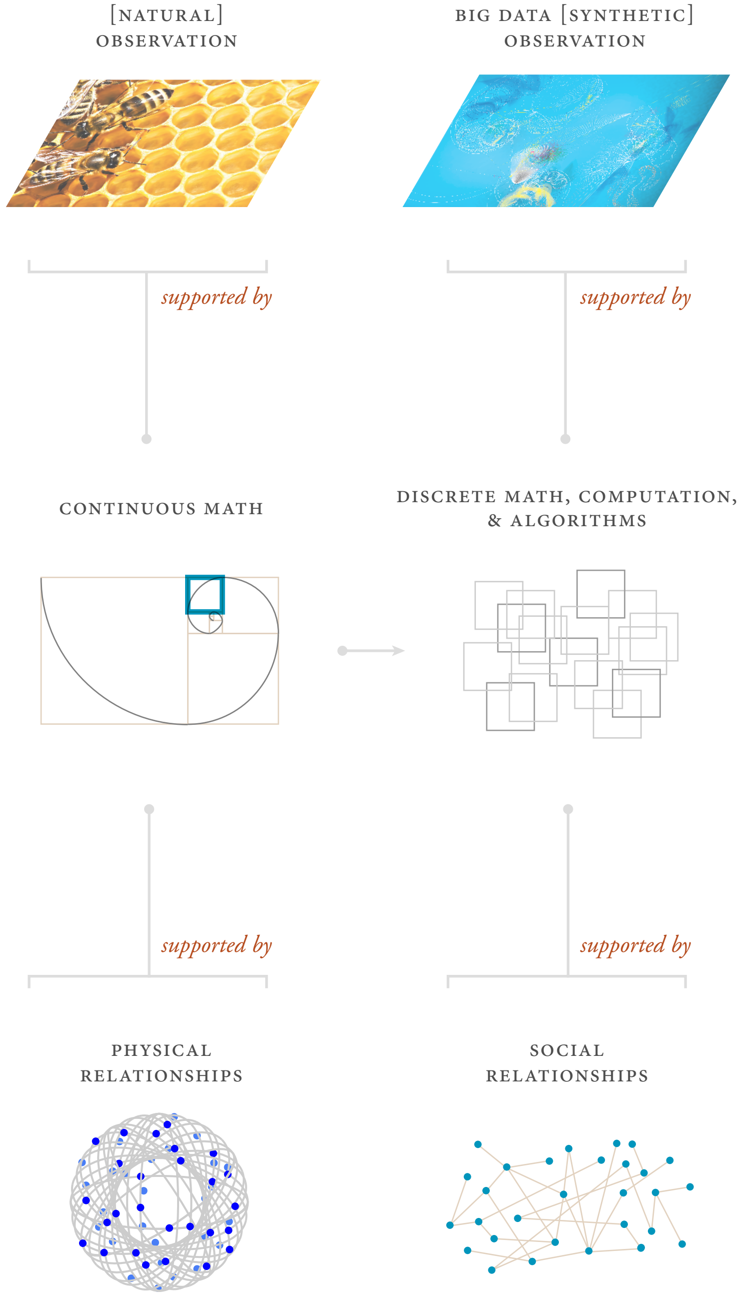


FIGURE 01: *This diagram presents a parallel between natural observation and metapictorial modeling: both are "pictures." From the natural picture we deduct from what is seen; for the metapictorial we induct toward what may be seen. Natural observation serves as a "found-gateway" leading to patterns revealing scientific knowledge. Metapictorial modeling is a "rendered-gateway" revealing "dissimilated" knowledge patterns (primarily, relational patterns) that have much higher rates of dissonance than found in nature. However, once generated, they perform retrospectively, offering insight into the "data-nature" of the whole: exposing critical anomalies, and offering predictive analysis. Therefore, they offer deductive opportunities via navigation back through their own landscape generated by big data — toward essential discovery.*

Natural images "behave" and render according to the dictates of the laws of physics and a preexistent, "continuous math": discoverable and resultant. Metapictorial images are rendered via invented and applied math that is fundamentally computational. The underlying driving force are myriad relational entities and the relational characteristics between the entities.

BACKGROUND

Stephen G. Eick, in his introduction for Fayyad's and Grinstein's *Information Visualization in Data Mining and Knowledge Discovery* states that, "Visualization is the link between the two most powerful information processing systems: humans and the modern computer." [Humans are] "easily overwhelmed by volumes of data that are now routinely connected. Data mining…is a natural reduction technique that complements human capabilities."

I would not necessarily agree that we are "easily overwhelmed," yet I would concur that persons who deal specifically with these kinds of massive datasets are certainly overtaxed to the point where any improvement in cognitive gain — efficiencies of knowledge discovery are to be very welcome. This is particularly the case in workflows that undertake the task of deriving new insight, intelligence, or what is now termed *anticipatory analysis* from the data through the process of visualization. The origins of knowledge were (and are still being) derived from real imagery. Big data is entering a magnitude of scope that in many ways mimics nature's larger systems.

Perhaps then, a new kind of naturalesque/synthetic imagery — here termed *metapictorial* — may be the ideal way to render big data, particularly to support knowledge discovery. The most simplified schematic comparing real imagery to metapictorial imagery would consider these three building blocks: 1) images of reality compared to synthetically generated "reality-styled" imagery (metapictorial), 2) derived, continuous mathematics against applied formulation (algorithms and computational efforts), and 3) unseen physical models compared to unseen relational models (FIGURE 01).

METAPICTORIAL IMAGERY: FINE, DESIGN, & SYNTHETIC

*The t*erm metapictorial has been applied for differing purposes through its fairly infrequent usage. It sometimes refers to a kind of "intuitive artistic competency" in the image-making process. Jožef Muhovič (Linguistic, Pictorial and Metapictorial Competence, Leonardo, Vol. 30, No. 3, pages 213–219, 1997) assigns a referencing/aesthetic meaning, that, he argues, transcends the figurative. It may be understood this way: when a competent artist renders interpretations of things seen in nature they may be highly realistic or considerably abstracted. Yet, despite this range of potential visual outcome something *additional, yet intrinsic,* becomes apparent through the rendering (FIGURE 02). Conversely a less-competent artist (or non-artist) will simply capture less of what is seen — if attempts are made at reality they will be incompetent, if attempts are made for abstracted insight it will neither be achieved nor conveyed.

FIGURE 02: *Picasso's Guernica* (1937)— *an example of the metapictorial from the viewpoint of fine arts. The painting is not real imagery, yet the abstraction is hyper-informative, transmitting the agony of conflict which embodies the composition of emotion distraught, abstracted, social relationships. Compare this to* FIGURE 07, 11, & 12, *which showcase the metapictorial from the viewpoint of the design arts.*

We can witness this in comparing children's drawings done by children (which capture the awe of discovery and are never purposely child-like) to an adult's drawing that is *supposed* to be drawn as a child would draw. These false children's drawings are so often dull and insipid, they lack the awe of the commensurate capabilities of the child.

In comparison to the metapictorial within the fine arts is the metapictorial within the design arts. Here the impulse is to modify the externally-real toward a more specific, and generally reductive representation. Please jump ahead to FIGURE 07 which illustrates this process of creating continually reductive renderings of bees crawling about their hives as an example of several design art renderings.

This paper concerns neither fine arts nor design arts metapictorial renderings, instead considering the modeling of *real kinds* of pictorial images through computational means: synthetic metapictorial images. Images that approach the naturalistic but are not. Also, images that are drawn as a by-product of computation and not with subjective and objective intent, so to speak, of the artist. Regardless, they may be analyzed as one would analyze the natural world. In essence these *synthetic metapictorial* images may permit deep and immersive investigations of big data generated models. They may enable whole new realms of knowledge discovery and other beneficial utility in working with big data. Their essential and characteristic value may be, at first, counter-intuitive; in their initial

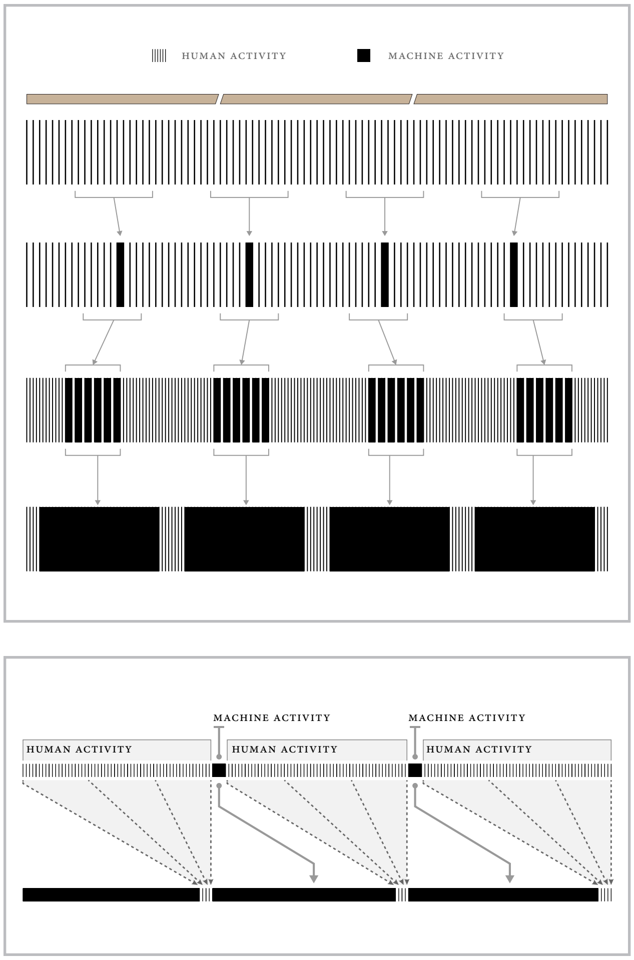


FIGURE 03: *Humans within machine interfaces: The simplest kind of interface is the use of a tool to effect some change against an outside system, such as a sledgehammer leveraging and concentrating human strength to shatter a rock, or as a means can to extract something from a system that evades typical human hand-to-eye coordination; such as fish being "fished out" of water by line or net (which is a matrix of interwoven lines).*

A series of tools, when assembled into an interoperable framework and powered by some means, become a "machine." In the diagram above the thick single lines represent tools, collected together in sequence they become machines. The fourth row contains integrated dense bars, these represent machine systems within a human environment. The base image diagrams vastly increased machine activity, supplementing human interface activity. For disinterested users, the simplest and least complex interfaces are desirable. For knowledge extraction we need to reveal the entire functional and relational "goings-on" within the entire system.

manifestations they would appear similar to experimental computer art.

Graphs, charts, and diagrams rely on the whitespace within or around their "elements areas" to convey informativeness. Photographs function another way — every smallest element (grains or pixels) has a corresponding positional and specific characteristic. If one imagines moving across a deep oceanic space the water is contiguous and displaced by creatures within the greater whole. Yet, the water is informative as well, under magnification it is breathtakingly rich in information itself — yet of relative consistency. Every molecule of water relates to its neighbor, and distance-wise, through ever-less decipherable means, to every other molecule within the ocean. Other organisms within this great body of near-infinite region are themselves infinitely related. Unlike graphical imagery, where details of the whole image provide less information, details of synthetic metapictorial imagery, generated via big data, would provide more-and-more information.

PROPORTION OF HUMAN-TO-MACHINE INTERFACE

Machine processing and machine activity is ever increasing proportionate to human activity (assuming continually developed areas of human activity). The use of hand tools serve as direct extensions and augmentation of human capability; the interface is tactile and physical. At the next level a series of tools may be utilized whereby, as part of the process, tool acts upon neighboring tool to effect a desired process. Through the use of correctly applied energy within a far more complex interdependency of tools a human may act "remotely" to the system. This requires an interface of navigation and control. The tactile may be supplemented by voice or eye movement, or through increasingly non-apparent means. Ultimately the interface becomes as non-intrusive and intuitive as possible. Extending this scenario further, it is easy to comprehend that systems are working continually on our behalf, at extremely sophisticated levels, and fully unbeknownst to us. FIGURE 03 illustrates this idea in simplified linear manner, depicting as well, a sense of scale with the human activity as an ever-decreasing *proportion of the whole.*

It is easy to see why user-centric design is so much a part of toolset development today. It is correspondingly fascinating to consider that the very ease and non-intrusiveness of the actualization of intention renders the human somewhat helpless if the systems do not "behave" as desired or expected. This can happen through the lack of proper energy, or misapplied energy to the tools (disruptiveness), or merely poor design, maintenance, or usage. Interface design is generally focused on the ease by which desires and decisions are carried out. For knowledge discovery we must unveil why these desires and decisions are being made. The metapictorial approach is to reveal the workings and interconnectedness of the human and the machine by analyzing the meta-information that identifies the (increasingly) total activity of the machines and the human activity within them.

INFORMATION, INTERPRETATION, RESOURCE APPLICATION

In order to render change to our environment physically; or to impact others through representations that encourage them to do so, consider this simple model. The subject, armed with whatever quality of information they possess, interprets and thus develops a commensurate capability or desire to act, and then summons the resources to do so. When students enquire of me, "what is good design?" I am inclined to answer (not smugly but with an aim toward discussing this model), "Good design is what good designers make." So the ideal somewhere near the center of the three vectors: information, interpretation, and resource application (FIGURE 04).

For the human-generated model the central area is populated with polygons of capability, expertise, or even a more desirable potentiality amongst all the possibilities. The better outcomes are in the more centric polygons and the lesser outcomes are somewhat easier to achieve because the "area of lesser results" are larger. We use default ideas of good and less good, or go no-go attributes as well. There is no such thing as a "perfect" home run — a major league home run swing is no easy matter, yet the polygon of good includes everything that goes over the fence. Another, overlapping polygon would include inside-the-park home runs (which are the lesser but oftimes more thrilling alternatives). But no one needs to hit a ball 365 yards through a hole that is only a micron in diameter larger than the ball itself — that would be a pin dot polygon in the center of the three vectors. Along these lines we can consider the way humans create either fine art or design art based metapictorial images. The results reach from not-so-good to very good indeed: ever moving into more talented, and increasingly smaller and more centrally boundaried polygons in the model described. Capability means we move toward the center. Conversely, for computationally generated outcomes we move out from the center.

Computer generated outcomes are expected to be accurate and desired (and though this is not always the case, it is the *expected and planned for* result). Practical math demands a precise home run every time. Here, however, is a limitation; if a machine is built to provide the same precise home run time-after-time, not only will no fans come out to the ballfield, nothing new will be discovered. So discovery and "goodness" along synthetic lines means an ever widening scope of possibility by direct vectors to parallel "discoverables" — or spirals out that cover significant comparative and contextual results — or crawls through vast amounts of data that build better and better capabilities (FIGURE 05). This kind of approach will also generate quite differing kinds of metapictorials, as would be expected due to the applications (algorithms and computations that generate results). This would mean, however, that the algorithms and computations would need to be design (interpretation) with the intent of allowing users to discover new knowledge and not merely slam out home runs all day long.

DECISION CLUSTERS

There are easily many hundreds, if not literally thousands, of and/or choices — forks in the road — that a team of engineers/designers need to thoroughly investigate, analyze, and consider toward a final choice of visual representation(s) respecting any particular kind of extremely large datasets or data systems. (Legacy issues often mean that large areas are unfortunately structurally or computationally pre-ordained in terms of their design.) Each consideration that can be undertaken will play its part in the algorithm set that will process data toward revealing tangible, or otherwise "visible" compositions. These compositions, once viewed and manipulated by users, might continually produce re-renderings. Some re-renderings would be extremely subtle in their change status, however, the possibility for major alterations could certainly result from any request based upon the composition of algorithms that act upon the data. For this paper I wish to focus on a very limited set of options — those decision trees that might most rapidly bring a viewer to the kind of "next-generation" visualization for big data; visualization schema in order to undertake intensely-effective modeling of synthetic knowledge. Essentially we will be driving toward a marriage of processing where machines and humans are ascertaining fabulous levels of insight — insight that causes the *rebuilding of decision-making tools,* and in turn, modifications to the tools that support decision support (FIGURE 06).

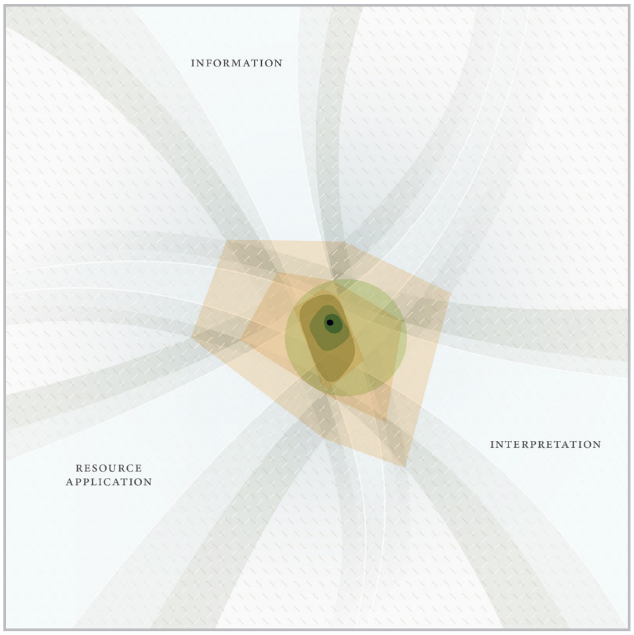
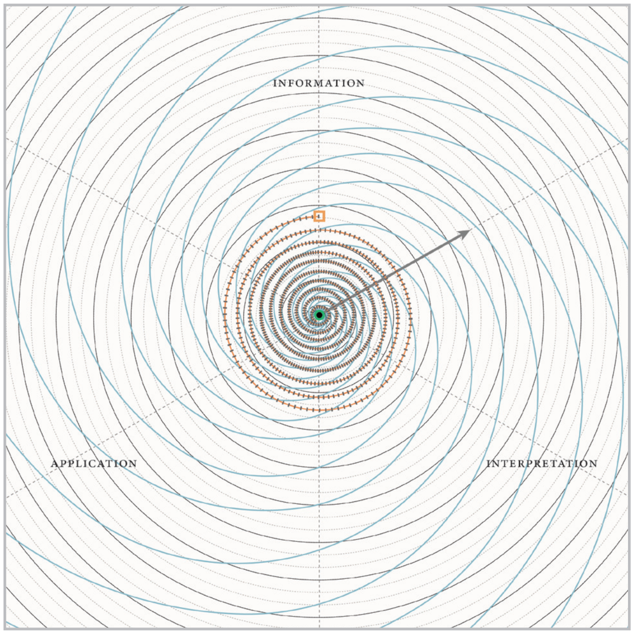


FIGURE 04: *Human-based model — Comprehensive knowledge (information), processed through interpretive skill, and supported via resources and physical application, effect change. Higher capability is depicted here by the denser centrality of the polygons— each of the lighter polygons reflecting lessening effectiveness and capability. The fading gray areas represent a complete fall-off of effectiveness. The physical results of such change can be seen and recorded, replenishing the information field toward the next cycle of activity*



Design theory argues for this ideal in the decision making process: *that each decision cluster addresses the larger conditional split prior to the next largest conditional split.* A decision cluster is any number of *and/or* decisions that appear to be of similar type, or would lead down a similar path of variance, or would be similar in the short-run, *i.e.,* decisions that are small and related enough to each other in the current state of the program and won't be of significance until later in the program. For example, making the choice of using blue now, and leaving the hue or value of that color until later (qualitative); or the choice of a future date in May two years from now with the actual date within that month to be later determined (quantitative). A decision *cluster* could involve hundreds of future decisions, yet these decisions do not impact a point of no-return, nor the elimination of options until some future point in the decision making process. A decision *split* eliminates all the other options. So the argument is simply that the most critical, the most impacting, and the most far-reaching decision clusters are tackled in order of magnitude — that the forks in the road are very similar to reaching a far-away destination by using the interstate federal highways, then the state highways, then the county highways, and then the local roads in the hierarchy of driving. And though this generally is a default logic for driving, for designing this is not often the case — design legacies, tendencies, policies, and politics often thwart the seemingly "logical" approach. When this can be done the 80/20 rule, in reverse, results. Each conditional split

FIGURE 05: *Processing-based model — Information (considered through human-interpretive skills and organized through varied taxonomies and ontologies, but potentially re-oriented through artificial intelligence, etc.), is rendered into "actionable representations" via applications. Precise and specific outcomes are expected (central point). Additional insight is thus supported through data visualization as a result of: "crawling" out in a spiraling manner from the targeted results (the orange cross-ticked spiral), or by vectoring out into other informative "conjectures," (arrow), or processing through polygon-like arcs that move more randomly across fields of data. These latter kinds of fluid crawls through the data will more likely yield metapictorial results. The concentric rings represent fields of varying data, with formal (solid, essentially quantitative) or informal (dotted, potentially qualitative) boundaries. Unlike the human activity which may move with great fluidity across ill-defined kinds of information, the data within a processing model is defined, even if predominately qualitative.*

(if in proper sequence) covers a rapidly decreasing quantity of the entirety of the potential success of the outcomes. Therefore, the first ten major decision clusters, if well conceived, applied against the first ten major conditional splits, will be of the magnitude of the next thousand or so increasingly minor decisions — so things will become logical and easier as the process continues.

AREAS OF CONSIDERATION FOR VISUALIZING BIG DATA

Following are ten areas of consideration presented in order of decreasing magnitude of importance (though all are important), which support those kinds of visual outcomes generated through computational modeling (algorithms) of big data: *specifically, the kind computational modeling decisions that should also support knowledge discovery.* The list is generalist because no specific project is herein specified — however, my generalist directive is toward effecting what the title of this paper indicates, *viz., "Big Data Visualization and Knowledge Discovery through Metapictorial Modeling."* Therefore, I apologize in advance that some of these recommendations seem atypical for best practice development cycles, but that is precisely the point.

1) **Data emphasis:** The bias toward favoring computational methods vs. infrastructural directives, and how this impacts high level visualization models.

2) **Knowledge discovery vs. decision making and decision support:** How legacies of visual model building impact subsequent visualizations, and why innovation is stymied.

3) **Specificity in design-centricity vs. user-centric universality :** The rise in the emphasis upon user-experience issues and heuristics —how this approach creates an info-visualization bias of compromise and "chasing the users."

4) **Tangible vs. interstitial:** as big data "fills" interstitial space why the need for "soft-data modeling" is worthy of consideration.

5) **Taxonomic vs. ontological:** how the notion of naming and categorization supports, or fails to corroborate with, the nature of the data being processes.

6) **Pictorial vs. diagrammatic:** Where real, synthetic, virtual, or quasi-realistic kinds of images — those which are more cognitively "direct" — *regain* primacy over diagram, network, graphical, and symbolic imagery.

7) **Physically pictorial vs. Metapictorial:** Where pictorial imagery, generally understood to observe laws of physics (even in representational modeling), give way to represen-

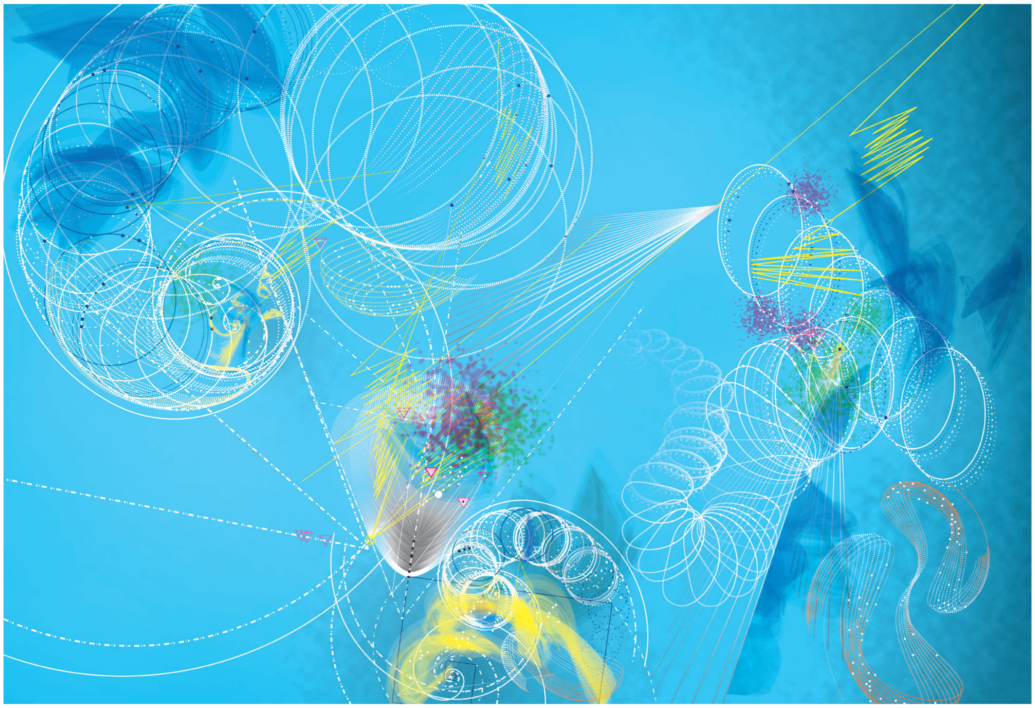


FIGURE 06: *A synthetically generated metapictorial image — differing kinds of patterns are emerging from the blue "ocean" of big data. The blue ground is not and absence of data, nor a basemap for referencing the emergent patterns. Instead, is softly represents a massive amount of essentially consistent data at this moment of viewing the data visualization — the tiny triangles are avatars assisting in searching through the data.*

tational images that defy such standards as they become more abstracted or "surreal" in representation.

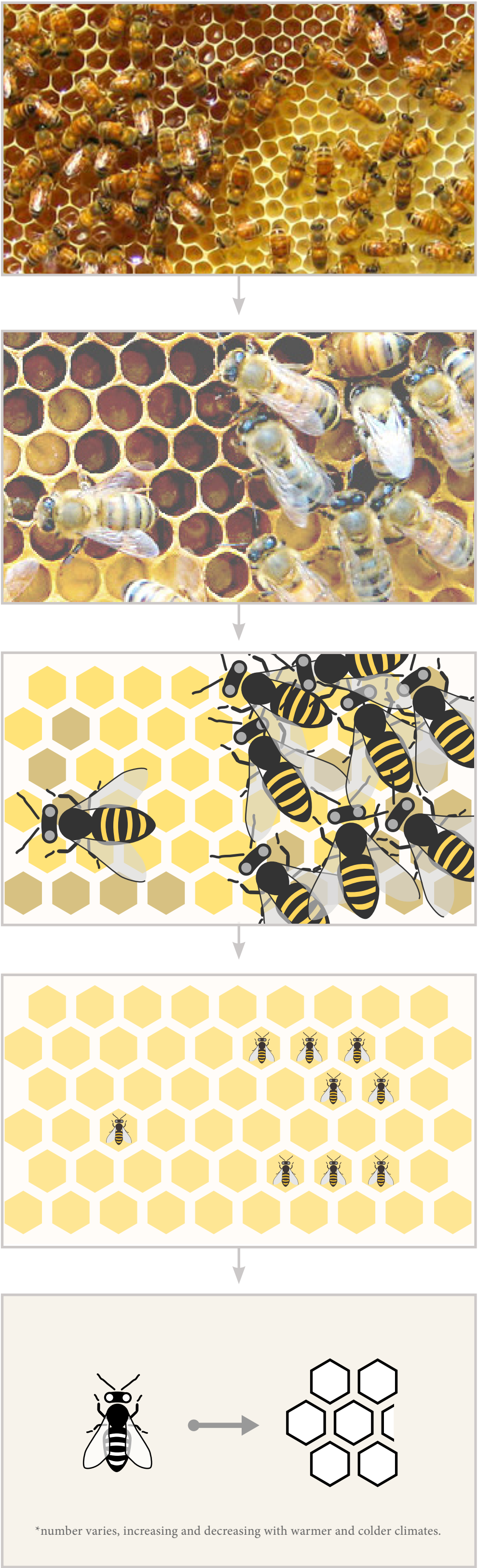
8) **Systems composition vs. engine composition:** the character of renderings that are "map-like" and composed of continuous fields of display, versus compositions that are compact, concise, self-reflective, and of closed contextual reference; and how these latter types can be "dispersed" through the former (*compare* FIGURE 06 TO 07, *and* 12 TO 11).

9) **Control field vs. immersive field:** how controls of the views can be modified by ostensibly external control methods, or through gesture based, immersive methods: relating to how we move through real worlds (intrinsic) versus libraries of knowledge (derived).

10) **Symbolic vs. signified:** understanding the distance from the core signified thing and how cultures share, or create new meanings, as distance of time or space move the viewer away from the signified elements — and the use of symbols as compacted elements of pictorial things.

DATA EMPHASIS: *the bias toward favoring computational methods vs. infrastructural directives and how this impacts high level visualization models.*

Every professional develops a bias respecting the nature of data on one hand and the way "raw" data is best composed as "content" on the other. The nature/content/display (FIGURE 07) composition from



*number varies, increasing and decreasing with warmer and colder climates.

FIGURE 07:
The design arts approach to metapictorial rendering — in this diagram the generally reductionist logic of design-based representation is shown. Unlike the fine arts approach to the metapictorial , which tends toward expressiveness, the design approach focuses on the informative. The upper image in the sequence is a photographic rendering; the design process might begin by isolating the area of interest within the photograph (second from top). The center diagram is simply a 1:1 reductive interpretation of the photographic image in question — the complexity of the hive is reduced to simple, albeit multi-shaded, hexagonal cells, the bees are highly simplified but of similar number. The fourth image depicts an additional layer of logic applied to the reductionist endeavor: the logic being to show the number of worker bees relative to the number of cells within the honeycomb. The final diagram is more reductionist still — the logic being to show the ratio of worker bee to number of cells to which it attends: one bee to 6.5 cells. Note that accuracy gives way to the quest for more rapid cognitive assimilation of the data.

data determines a great deal in regard to what may later be derived from it. In the simplest model some application of "knowledge tools" upon this nature/content composition renders representations. The representation embodies the communicative, tangible outcome that may then lead the user to useful interpretation. In discussing the first bias of practitioner's approach to big data I again turn to Fayyad, et al, "…two distinct camps working on two fundamental aspects of data mining have emerged…the first is focused on data storage and retrieval terminology as related to database theory and practice. The second is centered on the notion of algorithmic principles that enable the detection or extraction of patterns and statistical models from data. This latter branch evolved [from]…pattern recognition, and later under artificial intelligence (AI) and machine learning (ML)…[and now] knowledge discovery in databases (KDD)…"

So we will first consider the distinction between that "camp" which is primarily focused on the technical side of collecting, storing, databasing, and delivering data versus (an admittedly soft "versus") the side that looks to exploit the collection with pattern extraction and, ultimately, visualization. I shall focus on the latter, with an interest in the former, as it drives the logic of what kind of patterns result from big data collections which will most benefit the workflow. Before we move to workflow, however, it behooves us to consider, even at the "raw" level exactly what big data is. Before moving to the content level we are left to hover over a series of definitions of big data that all point in the right direction with no definitive opportunity to arrive at any collectively-agreed-upon location. This is irony of the nature of big data — once it can be clearly and universally defined it will not, per se, exist anymore — because everyone will know exactly how to exploit it. When it is somehow "complete" it no longer is big data as we *should define it* (FIGURE 08). For our purposes we'll list some aspects of big data that help set a soft parameter around what big data is — this will suffice as a jumping-off point. With each definition I will say something about a kind of visualization that addresses that kind of big data parameter. In all cases, though, our paths to visualization and knowledge discovery will bias toward the modeling, not the collecting (and storing) of the data. Although a bias, it does not mean that we do not place less emphasis on the critical aspect of data collection, storage, and infrastructure; *we want that aspect of our design process to be fully effective too, of course.* We merely desire that the infrastructural practice follow the lead of front-end mod-

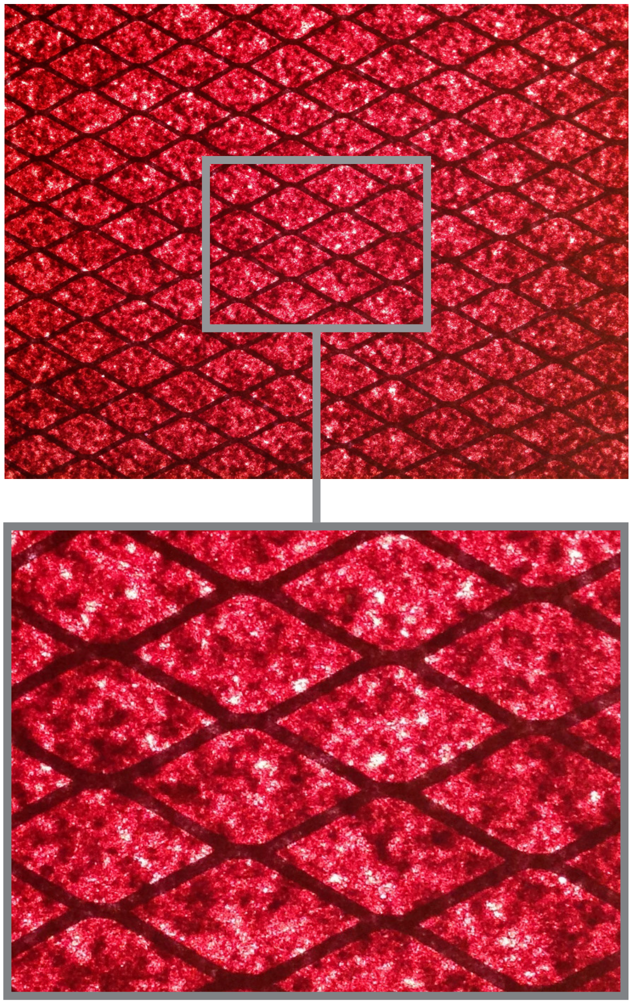


FIGURE 08:
A challenge of rendering massive synthetic metapictorial imagery is one of not knowing "where" to search within the expanded amorphous field. The renderings by definition would be abstract, and essentially without defined boundary. Big data would seem to indicate, by definition, that there is always a "hard problem." Two approaches to this challenge include the use of search avatars — code that provides for moving through the field of the imagery (not initially the data) to suggest areas of interest. Another method is simulated here through the use of applying a grid to the renderings and scoring discrete, yet similarly proportioned, areas within the entire matrix. In such a way a re-assemblage of the representation can be created that contains only those cells that possess anomalies above some threshold of interest or uniqueness.

eling aspects. We do not desire that collections lead the engineering effort or we will not get the type of visualization outcomes that best support knowledge discovery (this argument is too involved to justify here through example, suffice it to say that that which beforehand can be *envisioned in the mind,* which is often the case with collection and library science, will not render the unexpected, and we want to render the unexpected). So our bias is to have the computational camp set the requirements and desirements for data infrastructure side; though this, on the surface, may seem a bit counter-intuitive. Here is a brief list of some big data properties:

a) Usually large sets of data, in Terabytes, Petabytes, Exabytes and in future, beyond these — therefore, we are faced with the problem of modeling datasets that are too massive in ***visual scale*** — we are faced with designing models that can be rapidly scaled and re-scaled (as is now possible in GIS type systems) where one can drill-down and blow-back to maintain context — we are challenged by the need to simultaneously arriving at small, discrete points within the entire picture and opening these "small points" when the data set may not be nimble enough to facilitate this.

b) "…data so large it does not fit main memory." (Rajaraman and Ultman) — Therefore, we have a *de facto* challenge to the visual modeling problem, namely, ***visual incompleteness.*** We are faced with the prospect that our rendering machines cannot process all the data into any one model at any given time; at the most challenging level the models are, therefore, always in a state of being constructed, and our contextual surround is increasingly less defined (or resolved). If we move too quickly, or with too much resolution through the data, we automatically have less data.

c) "Here's the big truth about big data in traditional databases: It's easier to get the data in than out" (*The Pathologies of Big Data By Adam Jacobs Communications of the ACM, Vol. 52 No. 8, Pages 36-44*). Although this is very much an infrastructure problem, it is also a fundamental design modeling challenge — ***visual lag-time.*** Even if we are successful at building very comprehensive models of the data we are pursuing within the greater dataset, what we need to know might still be in a pipeline toward our already built model, or awaiting processing due to an inadequate rendering (not collecting) model (This is probably the case with most visual rendering modes that are being used to render big data today).

d) Big data can be replete with "touchpoint" problems, that is, there may be categories that have no coherent way to discretely render — ***visual non-compatibility.*** Here we would be concerned with taxonomy problems as well as inconsistent rendering "maximals," so that, say twenty points of coherent meaning on the front end are attempting to convey an incompatible number of "meanings" from the dataset. This would be akin to an outlet in an electrical box that had no incoming power source, or an electrical box for intended power output but which possessed no outlet in which to plug into.

e) Purge difficulty — Another challenge of big data is the ability to effectively delete information that is intrinsically redundant, obsolete, or legally non-collectable (these are mostly time-based or policy issues); this can result in —***visual redundancy*** — the default position of visual design and visual communication has always been one of real estate. Gutenberg based his epochal work on the blackletter hands of the exemplary scribes of the day — blackletter being a highly compressed script that saved valuable parchment; classified ad space was sold by agate lines (small depth-measures of less than 2mm width) in Newspapers; and telegrams created a language of "short-speak" nearly two centuries before SMS, or "text-speak." Concision, through analytical logic, or simply via scale (e.g., microprinting and microfiche generated through photographic reduction) has been a handmaiden to communications efforts from the outset. Visual redundancy can be said to exist through both unnecessary duplication of rendered data; non-required rendering of data; and unnecessary scale — all factors of real estate inefficiency.

f) "What we are seeing is the ability to have economies form around data — and that to me is the big change at a societal and even macroeconomic level." (Craig Mundle) Mundle places data on the level with labor and capital itself; bitcoin (a self described crypto-currency) is an example of inherent, distributed (in theory) data value, less stable than capital, and far less stable than labor, there is an inherent idea that data may be more inherently stable and quasi-natural, *more*

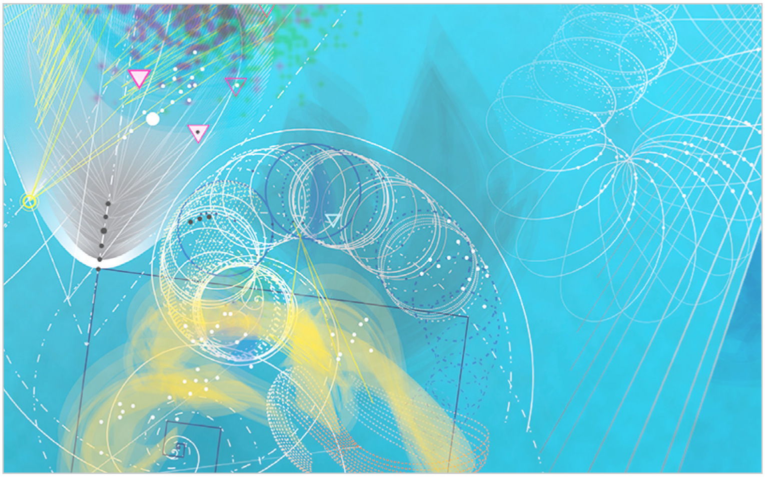


FIGURE 9A:

A simulated metapictorial rendering shown at the highest level of magnitude. This is a detail of FIGURE 06. *Recapping: the blue "ground" is a massive amount of temporarily consistently rendered data at for this moment of viewing. Conceptually, this consistency allows the representation to unveil complex relationships that are emerging from the full collective of data — each pattern is overlapped with other patterns which will reveal complex interdependencies. Moving through this immersive environment will immediately yield similar differences of revealed interconnectedness (layer-by-layer). The small triangular devices are search avatars that bring viewers to potential displays of interest.*

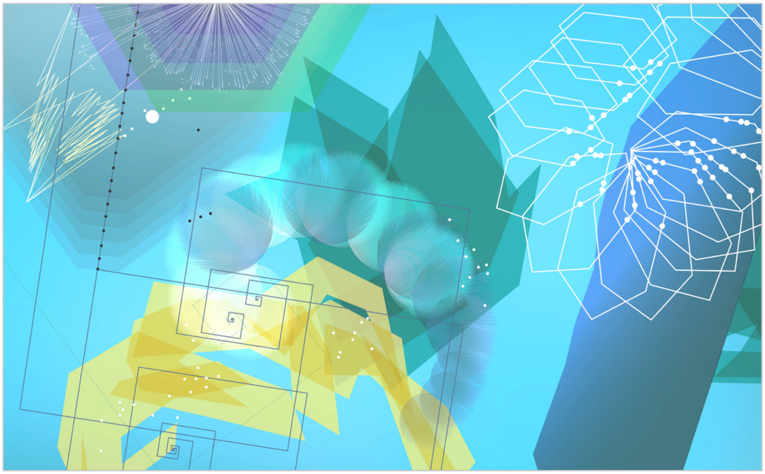


FIGURE 9B:

The model has been stepped-down from the higher level of magnitude to a less granular and less diffused rendering. (This type of rendering might also be displayed if users are moving more rapidly through the model.) As the metapictorial renderings become more formalized the taxonomic aspects become simultaneously more easily recognized — so certain shapes or line or dot elements are consistent and indicate similar aspects or collections of data. The background field also begins to reform and "coagulate" into representative kinds of amalgamated data — in this manner subtleties of visual distinction are sacrificed for more rapidly coherent contextual renderings.

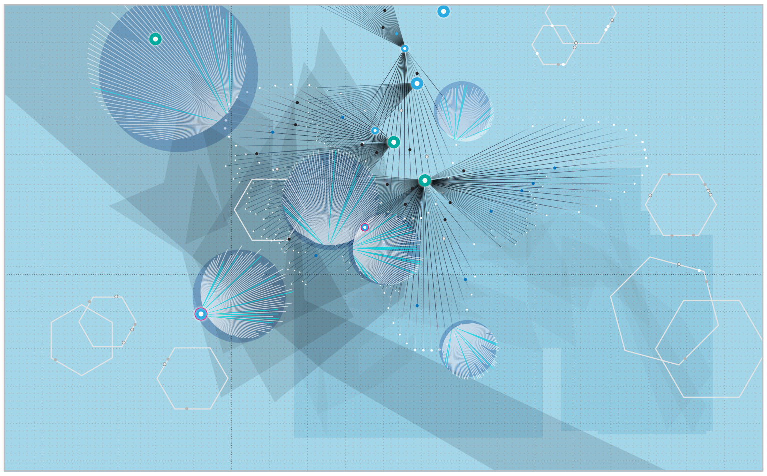


FIGURE 9C:

This is the lowest order of metapictorial magnitude illustrated in this sequence: the model is on the verge of moving from a pictorial rendering into a diagrammatic one. As a hybrid of pictorial and relational imagery the diagrammatic sections function as kinds of "fingerprints" of data that, though generalist in type, might possess higher level characteristics (due to curvature, line density, and consistency, etc.) than might be realized in typical, albeit highly complex, node-and-link diagrams. In such modeling, background context also plays a more taxonomically subtle role.

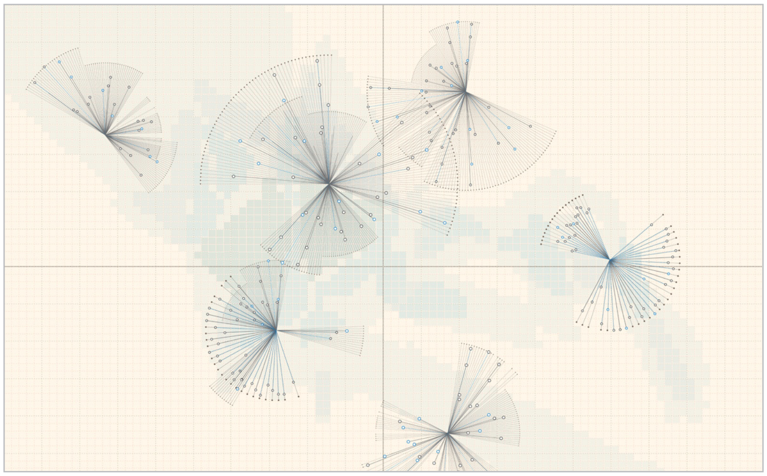


FIGURE 9D:

Here, the metapictorial rendering has been stepped down to a diagrammatic level. There are still rich subtleties within these node-and-link models as a result of the compression of the higher levels of rendering — the node-and-link elements possess aspects of "continuous tone" as opposed to purely digital, discrete models. The ability to withhold aspects of the analog modeling helps to define bridges across more discrete taxonomic divisions and reveal potential areas of unexpected overlap. This example is beginning to formulate into multi-tiered relational network divisions.

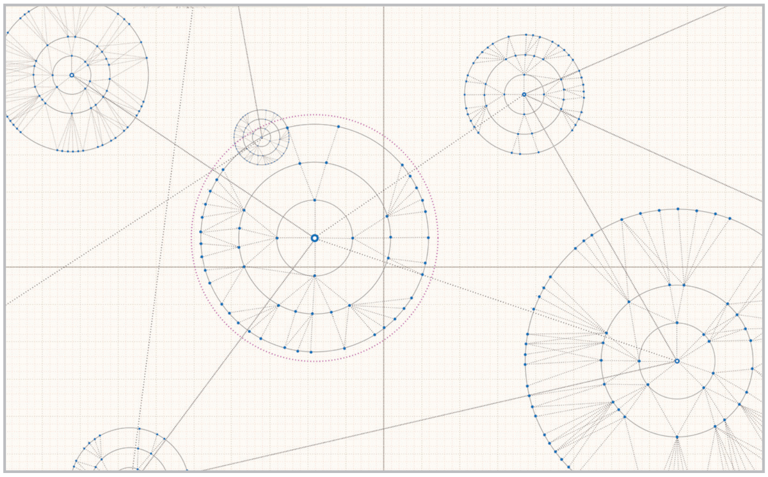


FIGURE 9E:

More highly diagrammatic with great clarity of distinction between link-and-nodes, hierarchy, and interconnectivity. As the level of magnitude is reduced absolute clarity between taxonomic characteristics are readily apparent. Logically, the potentially expanded network of node-and-links are captured within tiers that allow for a rapid cross-comparison of types — each cluster is composed of four levels and each cluster is further modified by size. The background begins to be shed of data reference, per se, it is now establishing reference values to the elements placed upon it — the total "size" of this rendering would be hundreds (or more) times larger in area than the metapictorial renderings.

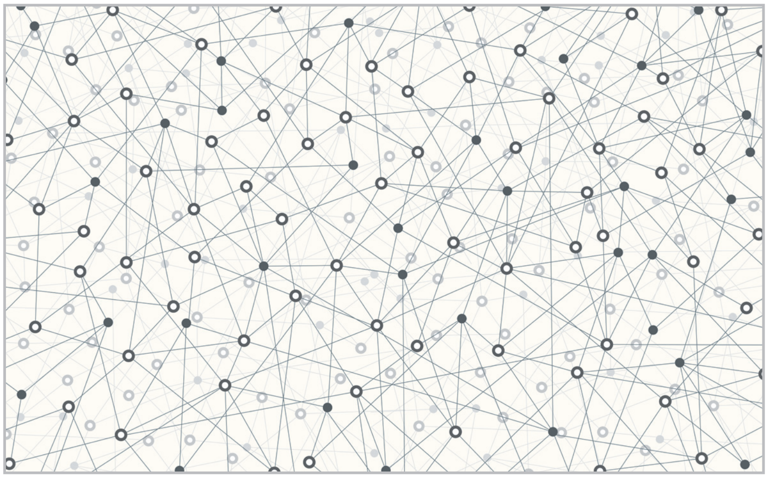


FIGURE 9F:

Although there remains a hierarchy of types, the diagram has been opened into an "unconstrained" relational network. The diagram still maintains qualities of an z-axis. This level of dimensionality is indicated by the use of black in foreground and greys in background. There is no longer any contextual referencing and no basemap to support extra-notation regarding the value of the nodes. Although still touchscreen based and immersive, the user would need to be selecting aspects of connectivity, the diagram would become extremely complex and very large without such filters and limitations.

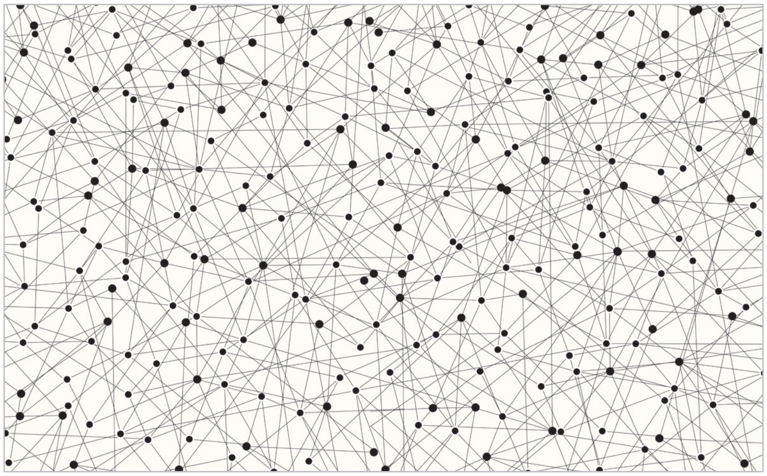
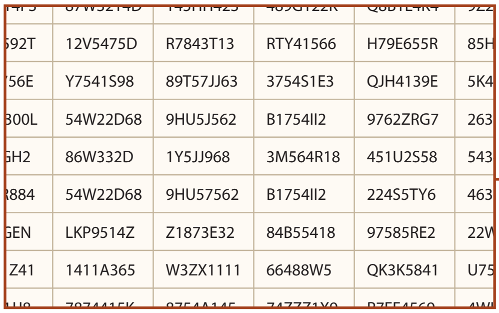
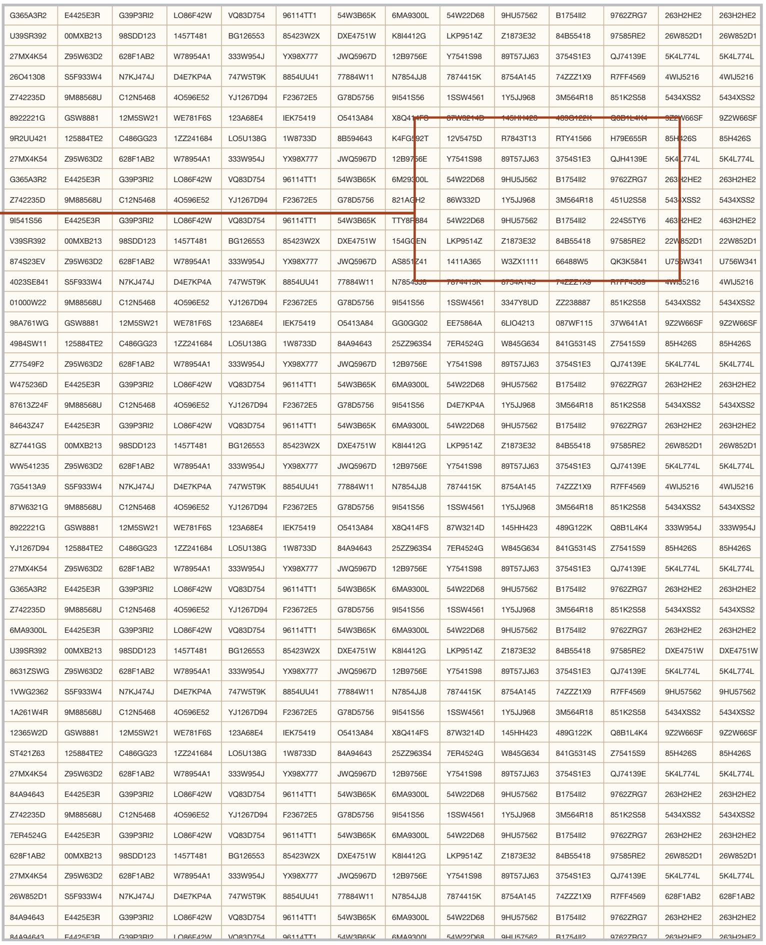


FIGURE 9G:
Stepping the diagram down another step the image is truly 2D, the node-and-link elements are on a single plane. There remains a slight hierarchy in nodal size; links would be selected from a large menu of interconnectedness options. This model, if spread out and compared to the highest order of magnitude, would be thousands of times large on a 2D basis — this would, however, be the desirable view if the users are down to the near-last-mile of their search. The orientation of the node angles provide context, otherwise there is minimal contextual intelligence.



592T	12V5475D	R7843T13	RTY41566	H79E655R	85H
56E	Y7541S98	89T57JJ63	3754S1E3	QJH4139E	5K4
300L	54W22D68	9HU5J562	B1754II2	9762ZRG7	263
GH2	86W332D	1Y5JJ968	3M564R18	451U2S58	543
884	54W22D68	9HU57562	B1754II2	224S5TY6	463
GEN	LKP9514Z	Z1873E32	84B55418	97585RE2	22W
Z41	1411A365	W3ZX1111	66488W5	QK3K5841	U75



G365A3R2	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
U39SR392	00MXB213	98SDD123	1457T481	BG126553	85423W2X	DXE4751W	K8I4412G	LKP9514Z	Z1873E32	84B55418	97585RE2	26W852D1	26W852D1
27MX4K54	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
26O41308	S5F933W4	N7KJ474J	D4E7KP4A	747W5T9K	8854UU41	77884W11	N7854JJ8	7874415K	8754A145	74ZZZ1X9	R7FF4569	4WIJ5216	4WIJ5216
Z742235D	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
8922221G	GSW8881	12M5SW21	WE781F6S	123A68E4	IEK75419	O5413A84	X8Q414FS	87W3214D	145HH423	489G122K	Q8B1L4K4	9Z2W66SF	9Z2W66SF
9R2UU421	125884TE2	C486GG23	1ZZ241684	LO5U138G	1W8733D	8B594643	K4FG592T	12V5475D	R7843T13	RTY41566	H79E655R	85H426S	85H426S
27MX4K54	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJH4139E	5K4L774L	5K4L774L
G365A3R2	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6M29300L	54W22D68	9HU5J562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
Z742235D	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	821AGH2	86W332D	1Y5JJ968	3M564R18	451U2S58	5434XSS2	5434XSS2
9I541S56	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	TTY8R884	54W22D68	9HU57562	B1754II2	224S5TY6	463H2HE2	463H2HE2
V39SR392	00MXB213	98SDD123	1457T481	BG126553	85423W2X	DXE4751W	154GGEN	LKP9514Z	Z1873E32	84B55418	97585RE2	22W852D1	22W852D1
874S23EV	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	AS851Z41	1411A365	W3ZX1111	66488W5	QK3K5841	U756W341	U756W341
4023SE841	S5F933W4	N7KJ474J	D4E7KP4A	747W5T9K	8854UU41	77884W11	N7854JJ8	7874415K	8754A145	74ZZZ1X9	R7FF4569	4WIJ5216	4WIJ5216
01000W22	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	3347Y8UD	ZZ238887	851K2S58	5434XSS2	5434XSS2
98A761WG	GSW8881	12M5SW21	WE781F6S	123A68E4	IEK75419	O5413A84	GG0GG02	EE75864A	6LIO4213	087WF115	37W641A1	9Z2W66SF	9Z2W66SF
4984SW11	125884TE2	C486GG23	1ZZ241684	LO5U138G	1W8733D	84A94643	25ZZ963S4	7ER4524G	W845G634	841G5314S	Z75415S9	85H426S	85H426S
Z77549F2	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
W475236D	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
87613Z24F	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	D4E7KP4A	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
84643Z47	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
8Z7441GS	00MXB213	98SDD123	1457T481	BG126553	85423W2X	DXE4751W	K8I4412G	LKP9514Z	Z1873E32	84B55418	97585RE2	26W852D1	26W852D1
WW541235	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
7G5413A9	S5F933W4	N7KJ474J	D4E7KP4A	747W5T9K	8854UU41	77884W11	N7854JJ8	7874415K	8754A145	74ZZZ1X9	R7FF4569	4WIJ5216	4WIJ5216
87W6321G	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
8922221G	GSW8881	12M5SW21	WE781F6S	123A68E4	IEK75419	O5413A84	X8Q414FS	87W3214D	145HH423	489G122K	Q8B1L4K4	333W954J	333W954J
YJ1267D94	125884TE2	C486GG23	1ZZ241684	LO5U138G	1W8733D	84A94643	25ZZ963S4	7ER4524G	W845G634	841G5314S	Z75415S9	85H426S	85H426S
27MX4K54	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
G365A3R2	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
Z742235D	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
6MA9300L	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
U39SR392	00MXB213	98SDD123	1457T481	BG126553	85423W2X	DXE4751W	K8I4412G	LKP9514Z	Z1873E32	84B55418	97585RE2	DXE4751W	DXE4751W
8631ZSWG	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
1VWG2362	S5F933W4	N7KJ474J	D4E7KP4A	747W5T9K	8854UU41	77884W11	N7854JJ8	7874415K	8754A145	74ZZZ1X9	R7FF4569	9HU57562	9HU57562
1A261W4R	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
12365W2D	GSW8881	12M5SW21	WE781F6S	123A68E4	IEK75419	O5413A84	X8Q414FS	87W3214D	145HH423	489G122K	Q8B1L4K4	9Z2W66SF	9Z2W66SF
ST421Z63	125884TE2	C486GG23	1ZZ241684	LO5U138G	1W8733D	84A94643	25ZZ963S4	7ER4524G	W845G634	841G5314S	Z75415S9	85H426S	85H426S
27MX4K54	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
84A94643	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
Z742235D	9M88568U	C12N5468	4O596E52	YJ1267D94	F23672E5	G78D5756	9I541S56	1SSW4561	1Y5JJ968	3M564R18	851K2S58	5434XSS2	5434XSS2
7ER4524G	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
628F1AB2	00MXB213	98SDD123	1457T481	BG126553	85423W2X	DXE4751W	K8I4412G	LKP9514Z	Z1873E32	84B55418	97585RE2	26W852D1	26W852D1
27MX4K54	Z95W63D2	628F1AB2	W78954A1	333W954J	YX98X777	JWQ5967D	12B9756E	Y7541S98	89T57JJ63	3754S1E3	QJ74139E	5K4L774L	5K4L774L
26W852D1	S5F933W4	N7KJ474J	D4E7KP4A	747W5T9K	8854UU41	77884W11	N7854JJ8	7874415K	8754A145	74ZZZ1X9	R7FF4569	628F1AB2	628F1AB2
84A94643	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2
84A94643	E4425E3R	G39P3RI2	LO86F42W	VQ83D754	96114TT1	54W3B65K	6MA9300L	54W22D68	9HU57562	B1754II2	9762ZRG7	263H2HE2	263H2HE2

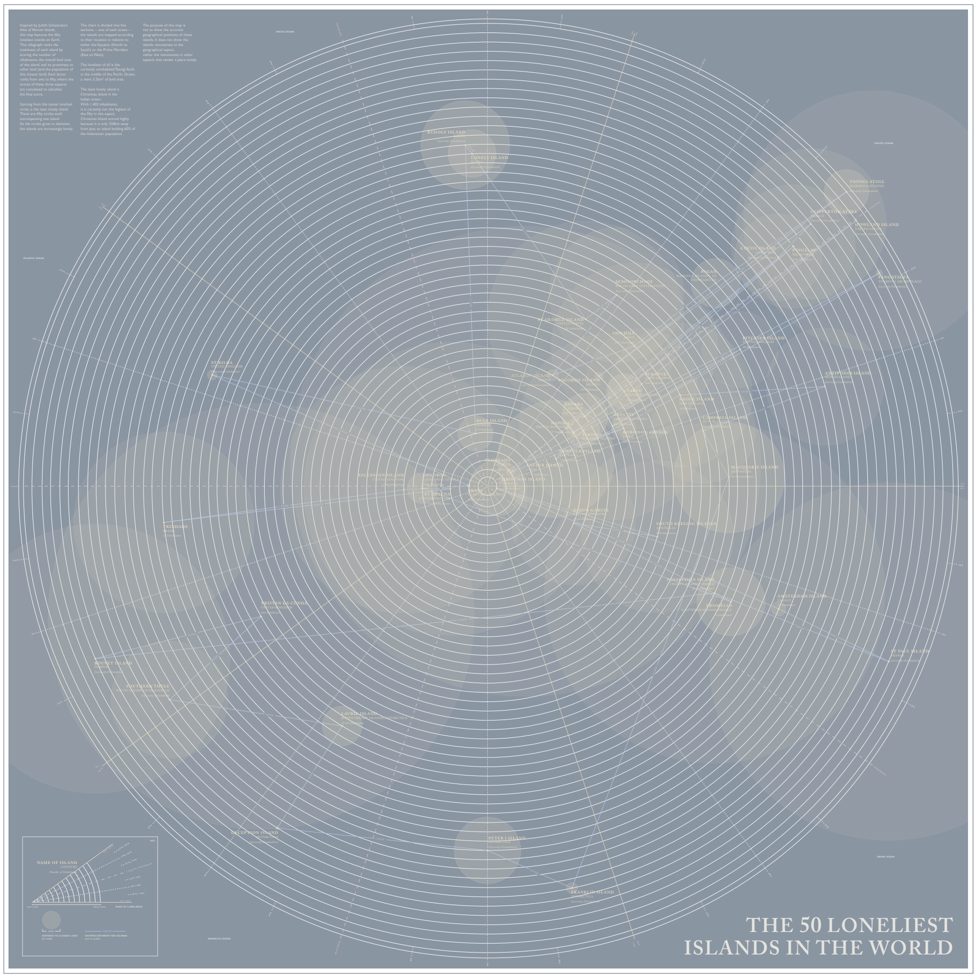
FIGURE 10:
This last example in the sequence illustrates the lowest level of magnitude: a relationally constrained spreadsheet. Millions of pages of such spreadsheets (here depicting USA automotive license plates as an example) would be continuously updated; and each of the spreadsheets can only depict certain levels of reference and interconnectedness. Therefore, many plate numbers would need to appear in a multitude of positions in order to express such interconnectedness. This rendering would most likely be navigated by more traditional control sets as well.

FIGURE 10 has served as a brief tour of informative visual representation from the metapictorial to the diagrammatic; within the diagrammatic it has moved from clustered to de-clustered unconstrained relational imagery to constrained relational imagery — the spreadsheet.

structure, than superfice. (But this awaits real world finance and government weariness of non-fiat currency to bear out.) This would lead to an idea in visualization of *"status ordinarius"* — in such interfaces there would only be excruciatingly slow, minimal change, a kind of reverse of time-lapse photography. Change detection is an example in image analysis that conveys this idea — if you have all the data over time it actually said to be moving "slowly."

g) Overlapping dimensions of big data, an example as defined by IBM: *Volume, Velocity, Variety, Veracity* — this useful 4V mnemonic speaks to the fact that there are vectors of overlapping qualities, or logical dimensions in bag data that provide a challenge to — ***consistent visual logic*** — in most cases this would be a fairly insignificant problem. One could use a consistent logic within one kind of view, a map for example, and then select another view, a photograph or diagram, toward elucidation of a point on a map. Therefore, the user simply disengages from one form of visual logic and calls up another. However, the potential richness of big data might mean that the scale of the data, the speed of its availability and change, the sheer kinds within the whole system, and the unknown factors of its reliability are not discrete, but merged. In such a case one would not move through types effectively, or if one moved through types the extractable knowledge might be ineffective. Consistent visual logic is a significant challenge when categories are effectively merging; graphical approaches such as rubber-sheet graphs have been used to address a kind of flow-through across typically discrete logics — big data permits logical flow-throughs which might be a challenge to model-

FIGURE 11:
Lonely Islands (48 x 48 inch), information design problems involve careful considerations of non-data space. Although this is not always a major consideration, and is often addressed through the intangible "talent" of the designer, it plays a major role in the effectiveness of a presentation, particularly in larger displays. This example included a fair degree of classification and "scoring logic" as well — these kind of design art metapictorial models would need to be coded to the specific content of that which is to be communicated. Only one kind of visualization — real imagery — generates its own naturally rendered model with requisite interstitial space. Synthetic metapictorial imagery can emulate these kinds of logic.

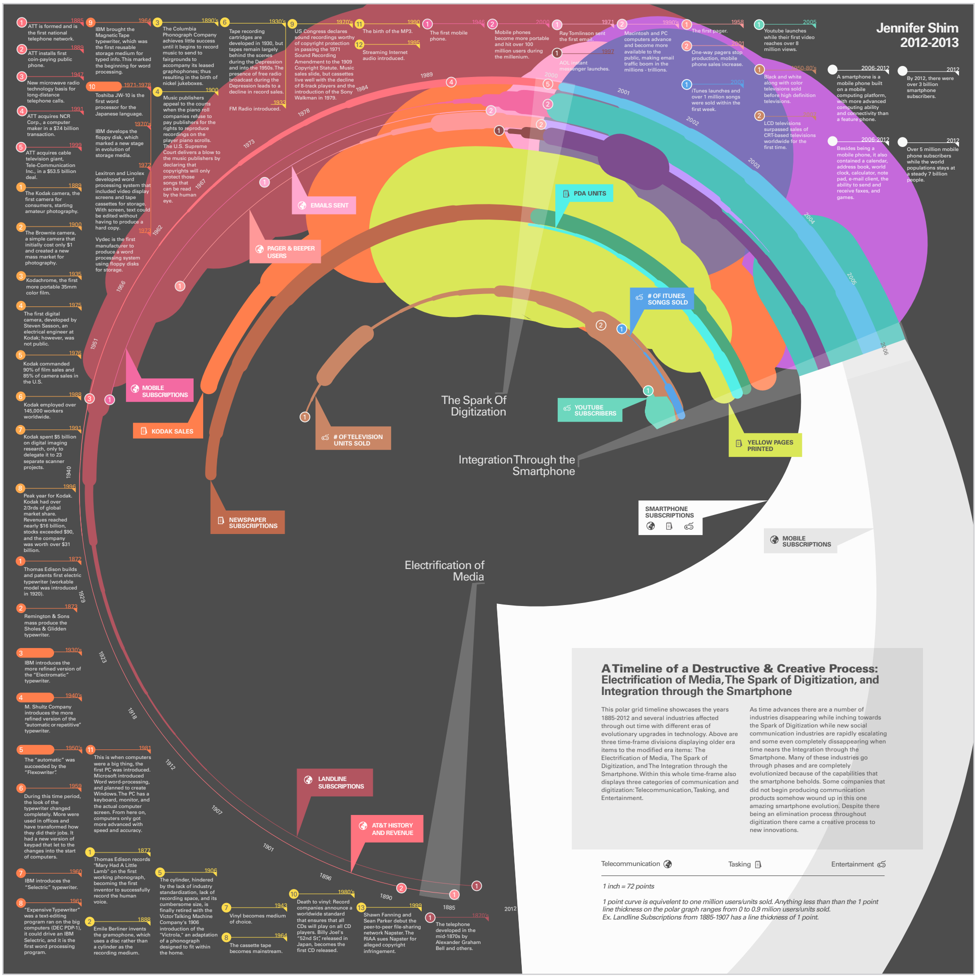


KANNY YEUNG 2013, USED WITH PERMISSION

ing data in a cognitively ascertainable way. From these definitional approaches to big data it can be seen that, as we are constantly looking to what the source material may be rendered, we are leaning toward the computational versus the infrastructural as the driver of some potential toolset. Having gone through these descriptions of the nature of big data I would ask that they be considered in the notion of the metapictorial that is being constructed here, so this is a good point in the document to look at how visualizations for data *through to* big data may have commensurate logic for visual representation — this is shown from the highest to (nearly) lowest levels of magnitude in FIGURES 9A through 9G and FIGURE 10. Having done this consider the second area in our list of significant conditional splits respecting:

KNOWLEDGE DISCOVERY VS. DECISION MAKING AND DECISION SUPPORT: *how legacies of visual model building impact subsequent visualizations and why innovation is stymied.*

The former CEO of In-Q-Tel, Gilman Louie, made this statement, "A tool that presents me with new ways of looking at new data is not nearly as useful as a tool that presents me with new ways of looking at the data I have to deal with every day." FIGURE 11 and FIGURE 12 address this kind of concern from a decision-maker's perspective. Though they are static images they are comprehensive and integrated (what we might term "info-engines" as they produce an energy of knowledge from a closed system). Louie was in position as a decision-maker and such integrated findings from knowledge one uses everyday would be beneficial. His directive reveals an underlying concern



A Timeline of a Destructive & Creative Process: Electrification of Media, The Spark of Digitization, and Integration through the Smartphone

This polar grid timeline showcases the years 1885-2012 and several industries affected through out time with different eras of evolutionary upgrades in technology. Above are three time-frame divisions displaying older era items to the modified era items: The Electrification of Media, The Spark of Digitization, and The Integration through the Smartphone. Within this whole time-frame also displays three categories of communication and digitization: Telecommunication, Tasking, and Entertainment.

As time advances there are a number of industries disappearing while inching towards the Spark of Digitization while new social communication industries are rapidly escalating and some even completely dissappearing when time nears the Integration through the Smartphone. Many of these industries go through phases and are completely evolutionized because of the capabilities that the smartphone beholds. Some companies that did not begin producing communication products somehow wound up in this one amazing smartphone evolution. Despite there being an elimination process throughout digitization there came a creative process to new innovations.

1 inch = 72 points

1 point curve is equivelent to one million users/units sold. Anything less than than the 1 point line thickness on the polar graph ranges from 0 to 0.9 milion users/units sold.
Ex. Landline Subscriptions from 1885-1907 has a line thickness of 1 point.

FIGURE 12: *The former example was a abstracted map positioned over an actual, albeit highly distorted, geospatial surface. This example is a simple quantitative and relational diagram rather than a design arts metapictorial image, although it does possess many aspects of the metapictorial from the point of view of its whole composition. The diagram illustrates the rise of multiple kinds of media and their subsequent number of users. It illustrates how digitization collapsed the distinction between media, and finally, how smartphones full immersed there into a collective (white area). One can see the unique developments; but the composite pattern renders a "gestalt" like insight — a picture of the entirety. (Therefore somewhat metapictorial.)*

JENNIFER SHIM 2013, USED WITH PERMISSION regarding data analysis in the general sense; that there is something lacking in our understanding of what we collect, we are simply not getting out of what "we put in". The very notion of big data magnifies this rift to a level of near absurdity. If we are getting ever more data are we at the same time falling more behind on the ability to parse this content? If we break the argument down we might consider a number of scenarios:

> *1) the standard sets of data viewed through*
> *the standard tools;*
> *2) new sets of data viewed through the standard tools;*
> *3) the standard sets of data viewed through new tools; or*
> *4) new sets of data viewed through new tools.*

In the first instance (the standard sets of data viewed through the standard tools) there would be no knowledge discovery, *per se,* although there would be an every growing collection of new knowledge along the same kinds of taxonomies and collections of the past. The second example (new sets of data viewed through the standard tools) will either be a kind of compromise of analysis, or more desirably, enrich the context of the existing collections. Both of these scenarios probably follow protocols of decision support. Users are familiar with the toolsets, and the toolsets have been designed with these users in mind. The process is very much putting square pegs into square holes; nothing much risked and nothing much gained. (But lots of billable hours[+], or conversely, hours billed[-].) These processes, admittedly, are critical in their own right for communicating to decision-makers (an ever increasing number who are not subject experts having arrived at positions of authority more from lateral, than through hierarchical means). The decision-maker relies on findings from standardized tools for the majority of actions taken; and these reports are derived from familiarized users of the data sets in question. Alternatively the decision-maker intends that directives, derived from supplied intelligence, are carried out — this is done by a system outside the loop in some cases; or through the same tools that generated the reviewed intelligence in all the others. Either way, the decision-maker is a kind of Janus: looking in one direction for actionable input and looking the other direction to see that desired actions. Then, resulting from all those sources of input, are expected actions generated from tools that fulfill the decision-maker's directives. If the decision-maker wanted to receive intelligence with ever higher levels of insight this might mean that the tools for decision support would be incommensurate to that objective. Again, as most toolsets are (understandably) built to specifications that allow the carrying out of the decision making processing it would follow that these would be somewhat antithetical to knowledge discovery.

To go further, let us consider the scenario where our opening quote — "…a tool that presents me with new ways of looking at the data I have to deal with every day" comes into play. This would be our third tier; the standard sets of data viewed through new tools. Here we see Mr. Louie's focus: the desire to discover the new or essential from the data collection as it stands and as it grows from the current method of collecting. Such next-generation toolsets were, of course, one of the whole points of In-Q-Tel. Big data unapologetically pushes us to the fourth tier; new sets of data viewed through new tools, truly a potentially chaotic point, and the very concern that the statement elicits. This though, is the *exact chaos that we need to tame,* and the very reason why knowledge discovery must be the "design-lead" in the effort. With big data coming at us (even though it is doing so through our own invitation and technical prowess) we cannot use standard methods of either computation nor visualization or we enter the fray already "behind the curve." Building tools for knowledge discovery requires a dedication to some creative methodology. (I prefer to just say *creative method,* but methodology seems to be the default term.)

This is a challenge because, with the exception of the most rarefied R&D, contracting is modeled around outcomes (read: requirements) that flowed from decision-maker's directives and program execution. Contacts rarely require, "do something novel" without then expressing exactly what novel is. Subordinates arc toward compliance; contractors fear losing current or future task orders. Workflows around data are generally biased toward decision support and program execution (discovering more of the known) than they are aimed at knowledge discovery (discovering more of the unknown). Very different kinds of visualizations and visual control methods are actually necessary for these two tasks. In the broadest sense, informative visualization supports areas for: Decision Making, Decision Support, or all the areas of "New Discovery." Of course, this kind of discovery/'uncovery' is often a major challenge that is undertaken as part of an objective expressed under decision making; yet the visualization for the former (discovery) are often significantly different than those directed toward decision-makers (read: "make it simple"). Once a decision is made the visualizations for supporting such decisions usually require quite different visualization and interface toolsets than one would require at the "start" of the process.

Allow me to express the challenge through the business model of a (typically) bureaucratic pyramid structure. Many workers at the broad base of the knowledge tools pyramid might serve only as data entry specialists; moving up from here are the analysts looking at such data in composite or "50,000 foot" views using visualization toolsets that support a fairly well stated objectives. Toolsets (and their visualizations) would be expected to be user-centric; well-considered in terms of clarity, minimalism, match, jargon-reduction, etc. In short, toolsets that were built very much with heuristic considerations. Here the users, though tasked with a purpose not of their invention, are carefully considered and accommodated for.

As we move up to the top of the pyramid there is a much smaller workforce, perhaps these are those tasked with knowledge discovery. Regardless they will be expected to contribute *innovatively.* Here I would argue, an overtly dedicated concern of heuristics and user-consideration will necessarily diminish the potential of true knowledge discovery. Knowledge discovery involves a kind of pain that comes form *not being comfortable* with what is before one, but greatly desirable to derive value at whatever cost is necessary to extract, or pay. How could these two visualization approaches be the same? It is highly unlikely that they should be.

SPECIFICITY IN DESIGN-CENTRICITY VS. USER-CENTRIC UNIVERSALITY: *the rise in Emphasis upon user-experience issues and heuristics — how this approach creates an info-visualization bias of compromise and "chasing the users."*

Our bias of knowledge discovery modeling (over decision making modeling and decision support modeling) rightly elicits a parallel interest in innovation and creativity. Any design trajectory toward a elevated, yet ill-defined, target should rightly do so. This means we need to look very carefully at our next general decision cluster: heuristic considerations. Here we see an area where the consensus of the professional is toward increasingly advanced means to deploy user testing, understand the psychology of users, and creating for users some kind of highly intuitive workflow for those who will be looking at, and working with, whatever is generated from our big data system.

A book that I have recommended for over fifteen years to all students in classes of Information Design is David Bohm's *On Creativity* (Rutledge). Bohm, a scientist, has a lengthy philosophical investment in thinking about the creative as it applies to the scientific. In the opening chapter he discusses the state of mind and the singular drive of the creative endeavor. After discussing the fact that most go along with the system, or react against the system in an uncreative way, he continues (paragraphs 53 and 54) "What, then, is the creative state of mind, which so few have been able to be in? …it is, first of all, one whose interest is what is being done is wholehearted and total, like that of a young child. With this spirit, it is always open to learning what is new, to perceiving new differences and new similarities, leading to new orders and structures, rather than always tending to impose familiar orders and structures in the field of what is seen.

"This kind of action of state of mind is impossible if one is limited by narrow and petty aims, such as security, furthering of personal ambition, glorification of the individual or the state, getting 'kicks' and other satisfying experiences out of one's work, and so forth. Although such motives may permit occasional flashes of penetrating insight, they evidently tend to hold the mind a prisoner of its old and familiar structure of thought and perception. Indeed, merely to inquire into what is unknown must invariably lead one into a situation in which all that is done may well constitute a threat to the successful achievement of those narrow and limited goals. A genuinely new and untried step may either fail altogether or else, even if it succeeds, lead to ideas not recognized until after one is dead."

This is an obvious inversion to the kind of logic found in the "Design for Dummies" books near the check-out counters of Barnes & Noble and more akin to the type of book written by Bob Gill, *Graphic design made difficult.* [sic] Design-centric logic is a type of *extraction* while user-centric logic is a type of *provision*. Consider, for example, the progress of science and the pains taken to find forms and patterns — applicable reliability — from nature. The search was essentially that of *design from nature* that is *design in nature.* Once knowledge is derived through considerable effort (and the kind of thinking extolled by Bohm above) human societies could choose one of two major paths: adapt (through a kind of provision) or modify through a new kind of design, post-natural design. This is what Western culture did. It will be seen that big data is forming a new nature (and a fast revising culture).

This is a nature on *the other side of post-natural design.* The data sets are becoming large enough to form natural systems of their own, and the same types of investigations that extracted patterns form nature will need to be applied to these conglomerates to access their values and in turn, act upon them.

In our search for next-generation tools for the display and analysis of big data visual renderings we should look toward the golden ages of our investigations into nature; what kind of mind took the challenges, risks, and dedication upon themselves to pursue such endeavors? And, who were those who then dedicated themselves to understand their findings? The students were not coddled, they too had to invest in the difficulty of the thing. Our bias then, is to look to multiple kinds of sacrifice on the part of engineering/design in development, and a similar (but different because its increasingly non-tangible) kind of sacrifice on the part of intended users whence deriving useful knowledge discovery from the toolsets.

TANGIBLE VS. INTERSTITIAL: *as big data "fills" interstitial space why the need for "non-data modeling" is worthy of consideration.*

One paradoxically-laden question we may ask of the big data milieu is this, "As data grows is it expanding outward or filling inward?" Logically the answer would seem to be both — in some cases we are building out the system with volumes of new data that is additive, in others we are collecting ever-refined data which can be said to exist between two points.

In mathematics this idea is captured in the notion of "infinity a" and "infinity b". In one case we have whole numbers in an ever expanding continuance of integers. We can assume that the interstitial space between these integers is the same; most diagrams equally space such points (provided the integers are uniform and sequential). Diagrams are conceptual. When one looks about in a natural world all that is seen is real imagery. Only foreknowledge allows one to look at a number of lakes and recall a map of the area, or hold up a chunk of quartz and see a periodic table of elements and the values for electron arrangements and atomic numbers for silicon and oxygen. The interstitial spaces in these diagrams are nominally equidistant. With the common exception of long period timelines (where the past is often compressed) most conceptual diagrams are fairly true to uniform interstitial space. Turning to infinity b, which is theorized to be larger than infinity a, we can understand that the interstitial spaces are of varying size, as an ever decreasing fraction would have similarly ever decreasing "realms" of interstitial "quantities".

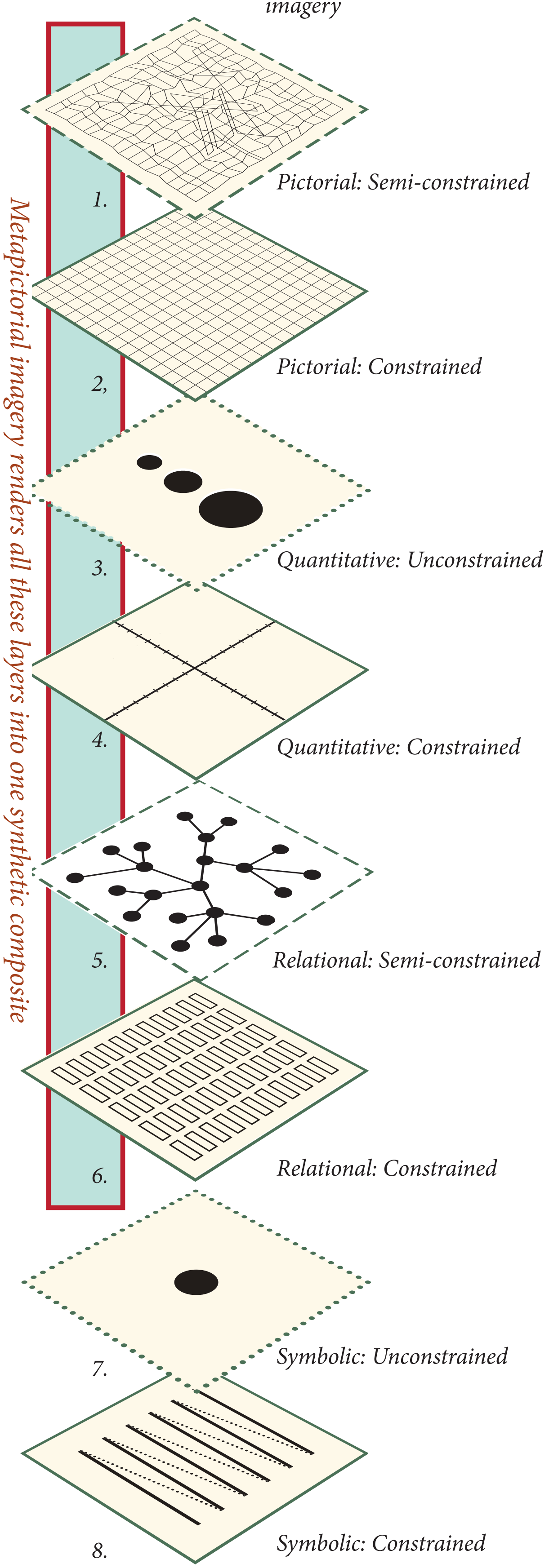
Importantly, this may provide us with another defining nature for big data, namely that the added dimensions in big data would arguably create a larger infinity of interstitial space. For, if all those data sets that are "not big data" have structures that provide any number of linear possibilities, even if those possibilities provided ostensible multi-dimensionality, there would still be mostly non-possible areas within the mass of intersecting infinity a-like datasets. Conversely, big data would be more like infinity b, permitting all those areas between the lines to possess more data. Therefore the interstitial space in an infinity b paradigm would be a higher infinity. And, this cannot be captured by graphs, diagrams, charts, or any 2D diagramming. It would be more likely captured by 3D imagery such as foldable rubber-sheet graphs, or distorted maps; but it would be most likely expressible through *a new kind of thing* (to purloin a bit from Wolfram) — parallel natural imagery. As real imagery is seen in our normal psychophysical context so could big data be seen through such *new-natural* imagery. However, it would not be natural but synthetic — it would capture the interstitial space. In the same manner by which a figurative artist, filling the canvas with a portrait, may choose to *not include some detail.* Yet, in the finished work something must be there. In a diagram we can leave things out, in real imagery we can falsify but there is no "blank" space.

Of course our whole objective is to model the data, (FIGURES 11 and 12 again) but we also need to model the non-data — we need to model the interstitial space — and this can only be done *automatically* with contiguous imagery; otherwise it involves extremely careful design investigation and spatial resolution. We need to consider non-data space from a computational level and aesthetic level with additional care for every rendering that is not in the higher state of pictorial imagery — for pictorial imagery the issue is merely determining boundaries — this can be a real advantage in working with big data.

TAXONOMIC VS. ONTOLOGICAL: *how the notion of naming and categorization supports, or fails to corroborate with, the nature of the data being processes.*

One of the intellectually rewarding challenges in the process of designing is establishing taxonomies and ontolo-

Figure 07: the eight patterns and four kinds of informative imagery


Metapictorial imagery renders all these layers into one synthetic composite
1.
Pictorial: Semi-constrained
2,
Pictorial: Constrained
3.
Quantitative: Unconstrained
4.
Quantitative: Constrained
5.
Relational: Semi-constrained
6.
Relational: Constrained
7.
Symbolic: Unconstrained
8.
Symbolic: Constrained

gies which lead to the logical structuring and hierarchies of types within the "allness" of the content being considered. In this paper I use the word "taxonomy" to describe a logical division of kind within the full set; added to this is usually the conditional (but not intrinsic) requirement that a name or label is attached to this inclusive grouping.

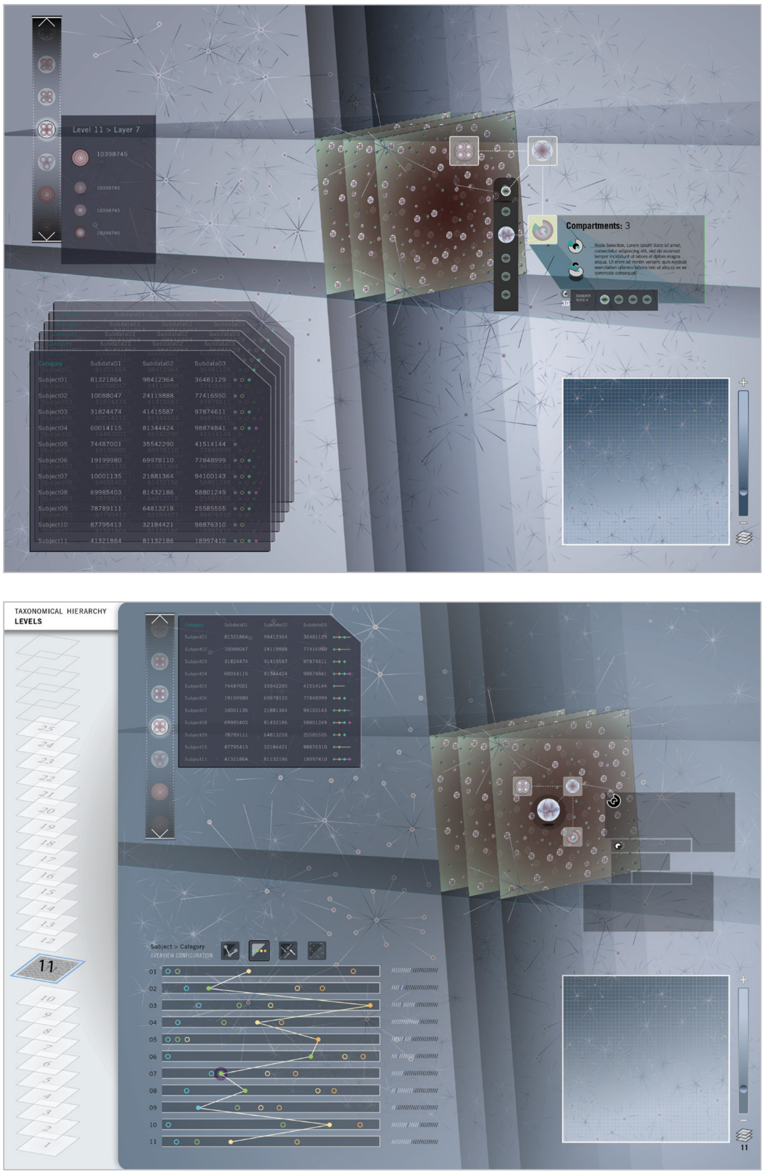
Such naming of patterns provides essential reference. By ontology I (here) refer to the nature, or the description, of any of these kinds. We might say that naming puts "an edge on things." By creating a semantic border around a particular class of data it allows us to consider how that data can then be desirably linked and scored to other data and how our computational modeling can render useful models across our greater taxonomy. In a way it can be said that taxonomic part of the process limits the perceptive opportunities, while the ontological process expands them. Why? Because the naming is intentionally a shorthand — the whole purpose is often to provide a mile-marker, stepping stone, arrow, or accepted agreement-point before proceeding to "what matters." The taxonomic process is one of concluding. Conversely, the ontological process is one of opening. The study of the nature of things is essentially open-ended — it can never be fully resolved and very often requires one to regenerate the taxonomy. If we carry this to *reductio ad absurdum* we would simply apply a label to everything and communicate no essential knowledge thereby, or study the nature of everything to the level of deep knowledge, but, absent any compact descriptive linguistic devices, be unable to com-

FIGURE 13: *Former work by the author identified four areas of principal patterns that underscore all informative visual representation. These include: Pictorial, Quantitative, Relational, and Symbolic types (P,Q,R,S). Each of these, in turn, have two levels of rendering — either high-constraint (and therefore structurally identifying visual elements according to valued positions of a basemap), or semi-constrained, or unconstrained (where elements carry intrinsic value). Fine art based metapictorial rendering would always be of the Pictorial semi-constraining type. Design art based metapictorial imagery would be the same except for informative, not expressive, purpose. Design art metapictorial imagery could also include aspects of quantitative and relational aspects as well. Synthetic metapictorial images, driving data through computational and algorithmic engines generate metapictorial imagery due to the sheer intensity of quantitative and relational densities. All three types utilize symbolic imagery in an annotated or navigational sense.*

municate any of that knowledge.

Expanding the taxonomy through refinement of types allows the ontological characteristics to move deeply into the labels. As an example I will refer to an article by Ole Henrik Magga in which he discusses the Saami culture and through their "long, intimate relationship with Arctic environment" their ontological familiarity with reindeer. *(Diversity in Saami Terminology for Reindeer, Snow, and Ice, International Social Science Journal, March 2006).* When this knowledge is expressed through the taxonomic labels of language it yields over one thousand ways to identify a specific reindeer quality. This is because factors of sex, age, shape, nature of coat, antlers, etc. are built into fairly compact expressions. The taxonomic effort is *moving toward* the ontological in this respect (until it would have a unique description for *each* reindeer). Big data from the collection side is an issue of taxonomic emphasis if only because the connecting points or sensors possesses technical conditions that, at least in the outset, are directed to certain kinds of collections. Once, these flow in volumetrically and cross-collectively they become more "ontological" in feel if not actuality.

The disadvantage for knowledge discovery, then, is that the analysts bring with them expertise-focused labels. The systems interfaces they look at are imbued with taxonomic divisions that absolutely pre-categorize. (What seems to occur whenever we fill out a form with multiple choices in which we cannot seem to answer with full truth, or accuracy, due to directed answers. So we are forced to compromise accuracy). The solution is to design displays that allow the full dimensions of the data to be "fluid" and uncategorized in the first state. These can then be clustered by the users through the discovery process for the purpose of comparison. Then, in order to facilitate communications about findings, a taxonomy can be generated that calls upon a "fairly compact expressions" which yield insight into the discoveries.



Category	Subdata01	Subdata02	Subdata03	
Subject01	81321864	98412364	36481129	● ○ ●
Subject02	10088047	24119888	77416950	● ○
Subject03	31824474	41415587	97874611	● ○ ●
Subject04	60014115	81344424	98874841	● ○ ● ●
Subject05	74487001	35542290	41514144	●
Subject06	19199980	69978110	77848999	● ○ ●
Subject07	10001135	21881364	94100143	● ○ ●
Subject08	69985403	81432186	58801249	● ○ ● ●
Subject09	78789111	64813218	25585555	● ○ ●
Subject10	87795413	32184421	98876310	● ○
Subject11	41321864	81132186	18997410	● ○ ● ●

FIGURE 15, 16: *With a purely immersive interface model, as would be most desirable for metapictorial rendering, there are several drawbacks respecting collaboration, information capture, information sharing, and delivering intelligence to decision makers. These hypothetical interfaces address the concerns. In the series of illustrations shown in* FIGURE 10 *the levels of hierarchy are moving along a kind of "y" axis of complexity; in these illustrations the logic is more of a "z" axis. Here, the magnitude of the visualization is being graphically rendered so that the findings can be captured. Once the transfer is made to these graphical captures the control aspects permit formal navigation and sharing. Elements and findings can be extracted and constructed into models similar to* FIGURE 11 *&* 12, *such presentations may be ideal for decision makers to review.*

PICTORIAL VS. DIAGRAMMATIC: *where real, synthetic, virtual, or quasi-realistic kinds of images — those which are more cognitively "direct" — regain primacy over diagram, network, graphical, and symbolic imagery.*

Readers will recognize that the thread of my argument leads to this: that I advocate that "real" imagery serves as an informative visual modeling agent of big data, and that this will strongly support knowledge discovery. Imagery is cognitively direct, but graphical models of all kinds have served the role of isolating findings from nature. This in such a way as to bring powerful clarity through isolating discrete intelligence and displaying these (usually quantitatively) in revealing patterns.

Let me pause here to refer to such a taxonomy, that I developed (with the insightful and never tiring assistance of Dr. Arno Klein) about twelve years ago which is pointedly germane to the idea of metapictorial imagery, (FIGURE 13). This is worth reviewing in terms of the magnitude of density of image needed to extract greater intelligence from big data. I then argued and still do that all *informative visual representations* fall under only four systems. This coarse generalization may raise some objections, "only four?" After the analysis of many thousands of examples of information design, the logic supporting only four core types, or primary classes, was found to be sound. (This system was first developed during an academic-contractual project for the U.S. government, it was entitled the VT-CAD system. This stood for "Visualization Taxonomy for the Classification, Analysis, and Design" of Informative Visualization. The goals of the program were captured in the "CAD" moniker: to effectively and easily *Classify many kinds of images,* to provide a rapid *Analyze collections,* and to assist in *Designing imagery* and toolsets for high performance communication.) Each class is both structurally (through its appearance) and logically (through its nature) defined. These are the four: *Pictorial, Quantitative, Relational, and Symbolic.* Pictorial patterns convey real and

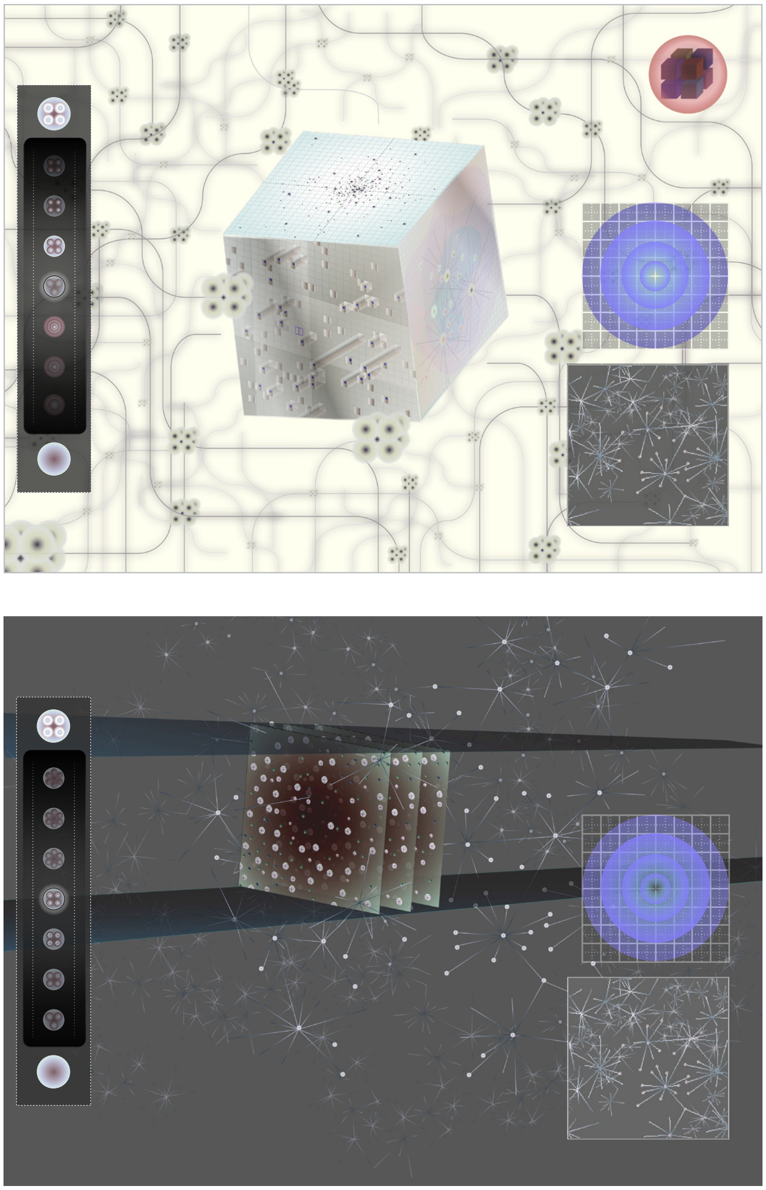


FIGURE 17, 18: *Further to the examples on the previous page, these illustrations address the potential need to move from immersive, gesture based, interactivity to a controlled interface. A parallel idea exists in reality (observations of nature) when it becomes necessary to capture and share findings. Users need to isolate (at true scale, or via magnification) that which they have observed. With synthetic, metapictorial imagery, the user moves though a world that needs to be addressed by layer and composition. Controls would all be carried within the viewing glasses of the user, or through some other augmenting toolset so no customization to the metapictorial model is required — however, as shown above, users could toggle to controllable layers within the metapictorial so that full collaboration and knowledge discovery findings could be rapidly shared.*

imagined imagery. Quantitative patterns are concerned with distances and numerical values. Relational patterns utilize cells and locations for containing and comparing elements. Symbol patterns exploit encoding and syntax. *(please see the paper PIIMPAPER0103 [searchable online] for additional detail on this theory)* Each pattern serves to extol both composition and comprehension of their respective, or applied, content types.

As we work on extensive bottom-up taxonomies of ways of looking at data, specifically big data, we see multiple discrete and hybrid examples of these four principal kinds of visualization. But the quantitative and relational are the yeoman. Graphs, charts, diagrams, tables, node-and-link diagrams are looked to as best-practice ways of rendering big data. The last of these, node-and-link imagery has been particularly called upon for the task. Node-and-link and network diagramming, i.e., relational diagramming has grown in usage to the point of being the near default method for massive views of interconnections amongst data. The logic of "relational" is certainly appropriate to the nature of big data.

By "pictorial, real, and imagery" however, it is not meant that such images are necessarily typical in appearance. Images taken from nature, even discrete and highly subjective magnifications such as Karl Blossfeldt's images in his work *Urformen der Kunst (Art Forms in Nature)* are still, essentially real, as would be fractal images and hyper-spectral images from GIS data — maps may go far into subjectivity and bespeak significant re-scaling, yet these two, are essentially real. The reality exists because of quantitative traceability. There are quantities behind the images and in that sense obey the physical laws behind them: quantities. Much big data adheres to these restrictions as well, but our concern for knowledge discovery shifts to relationals as the underlying driver, not quantities. This will generate images that are "real" but beyond what we would recognize in reality, more unlike real imagery generally, but possibly more like it in tiny areas of highly magnified specificity.

PHYSICALLY PICTORIAL VS. METAPICTORIAL:

Where pictorial imagery, generally understood to observe laws of physics (even in representational modeling), give way to representational images that defy such standards as they become more abstracted or "surreal" in representation.

Let us look a bit more into this idea of projected quantities as reality and the "physically pictorial" vs. projected relationships as "metapictorial." This brings us all the way back to FIGURE 01, which at the beginning of the paper was a good chunk to bite off and is, perhaps, more palatable now. First, a very quick retelling of the former paragraph: real images can be greatly distorted to convey some specific meaning, yet they are still tied to their root quantities (however distorted) and their root symbols (however subjectified). For example, a high-resolution satellite image of the Earth's surface may be said to be real, accurate, and pictorial. If I take a napkin and draw upon it a map from point-a to point-b , showing perhaps a couple of lines representing several streets, I am evoking reality. The napkin sketch is neither accurate nor highly pictorial (it actually becomes network-like). It is still real. If we look at some classic David Hirschfeld caricatures we see horrendous distortion. These distortions provide insightful cognitive gain about the nature of the artist's subject at the cost of distortion from accuracy. The images are still real, and deeply behind those images (and of no concern to the viewers), is a physical reality of blood and tissue and brain and bone. The point is admittedly stretched, but there is a kind of *natural big data* behind the smallest, the most inaccurate, the most subjectified of imagery. Natural big data possesses a true superfice of the real images, supported by quantities, and further framed by physical laws, i.e. nature.

Synthetic big data (on the other hand) can be visualized in any number of ways. Many of these methods are simply derivatives of pictures anyway, yet, as with reality, signal captured "pictures" (or in non-visible examples, other forms of signal captures, audio, and vibratory, for example) render tangible representations. It is the pictorial, or signal captured, images that are at the top of the food chain, value wise. Behind the synthetic big data there will surely be numbers, but these numbers may be more aligned to relationals than quantities. This means that natural big data is supported by quantities underscored by reliable physical "law."

What then underpins the relationals that generate our metapictorial images? Is it akin to a reliable physical law? Yes, it would be in some ways akin, in that it would drive a great deal of the relational data on top of it, but no, it would not be reliable. This is because the underpinning rules for relationals are (at least as we understand them today through economic and gaming theory, et al) unreliable and fairly inconsistent in the singular. They may bubble up to a slightly more reliably decipherable fuller context. The underpinning equivalent to *physically pictorial* would be the ***affinities, dependencies, and exploitations*** that establish the relationals. Therefore we have continual connecting, disconnecting, coming into existence, going out of existence — as well as highly varying temporal and magnitudinal fluidity in the "machine" that drives the relationals within big data. Therefore, the images so projected would not be deeply underwritten by physics law but by a kind of *"law of the emotive."*

Some renderings of big data do nearly build images that appear almost real, almost mimicking fractal imagery. As the volumes increase and we step back from visual renderings of massive data sets we might begin to see clouds (real clouds, not mass-connected computing diagrams), or flowers, and myriad things that look like super-magnification. This, I believe is the right step in terms of next-generation big data visualization. What will need to be added is fluid modeling of interstitial space, fluid taxonomic "gateways" and variable scoring systems for the affinities, dependencies, and exploitations amongst the data. This is the way that we can shift from a quantitative underpinning to a mathematically relational underpinning to the data sets and render next-generation visual representations of big data. These are also the kinds of visual representations that can capture interstitial space.

SYSTEMS COMPOSITION VS. ENGINE COMPOSITION:

the character of renderings that are "map-like" and composed of continuous fields of display, versus compositions that are compact, concise, self-reflective, and of closed contextual reference; and how these latter types can be "dispersed" through the former.

Due to the relational nature of big data there is another factor within our metapictorial world that needs to be considered. This will be an important feature, particularly if we want to turn our "awesome" pictures into usefully clustered things that can then be drilled into to yield expected, or unexpected, intelligence. When we look at a map of the earth we see continuance, every projection is a compromise because we are really looking at a sphere that must be distorted in order to view either by scale or other technology, or limitation.

Systems composition (as defined here) refers to this kind of data continuum — this is exactly the kind of endless surface one would expect from big data, particularly big data that is continually growing in size. So system composition is map-like; it keeps going. Relational imagery is made up of cells that either border one another (as in a spreadsheet) or are connected by node-and-links. Machines can quickly read spreadsheet cells, but more processing power is needed to read (or render) a node-and link framework. The "walls" in a spreadsheet between the entries are merely 1D links; up-down, and left-right. Node-and-link imagery allows for compacting spreadsheets that might minimal data entry for thousands of columns and dense data sets in one corner or another. By allowing the links to connect the nodes with greater subjectivity additional dimensions can be added and space can be saved. This results in a systems composition that can unfold endlessly through multiple dimensions. The same can be used for quantitative displays. If one took a spreadsheet of, say, massive size and reverse-projected it into a sphere, one would have a "engine composition", which is here defined as an enclosed data model (again compared, engine-like, in FIGURES 012 and map-like in FIGURE 11, *and all the metaphysical models are map-like* FIGURE 06, 08, SERIES 09A —09C). When we stand, observe, and move about the Earth we are in a systems composition; but from a great distance we see the planets as unique engine composition within a greater systems composition of the universe. When we look into Tuesday's science section of the *New York Times,* we are met (oftentimes) with lovely information graphics diagrams — these are engine compositions, all the information that needs to be presented is encased in the diagram. Engine compositions are composed within frameworks that in rare, rather brilliant cases (such as, for example the periodic table of elements) the engines are beautifully self-referencing. These exist, again, in terms of certain physical sets in science and multiple symbolic sets in the arts and metaphysical worlds (poems, paintings, wisdom literature).

Everything may *somehow* be related to everything, but we are most likely more interested in how certain things are *specifically* or *generally* related to certain other things. Any filter begins to turn a systems composition into an engine composition and when the last element of unresolved connectivity "snaps away" from the greater system successfully that engine can be investigated in totality for the knowledge it may reveal. As an aside this is why information designers often struggle greatly to build representational models that contain (or ostensibly contain, or through some magnitude of scale contain) all of the data. Whatever that thing looks like is a control window into the knowledge. Compare this to an interface designer who often sets immediately about categorizing the data and then setting up links and control methods to retrieve such-and-such a data type. One can see how the engine is a picture of not only the data but, at its superfice, an entryway into the data — whereas a group of controls, no matter how well designed, are merely *symbolic* gateways into the data.

Nature also possess these kinds of engines, although they are dispersed throughout her system. Ant or bee colonies are kinds of engines, and we see the greatest model in our scanning the universe with a life generating engine, the sun, providing energy to a life receiving engine, the planet earth.. Of course, whole sets of algorithms need to be created that might render a homogenous system into a system full of disconnected engines within the system. These then would be reconfigurable to generate new scatter-worlds of revised engines within remodeled systems.

Comparison is one of the most rapidly deployable tools of analysis and visual representations within an engine/systems modeling environment is a logical approach to providing this capability. This completes our tour of the modeling recommendation for big data representation. It mimics our natural world in its first and second and third levels; the universe, to the sun and earth within that space, to the pictures (detailed captures) of the world. It goes a different way at the deeper levels of relational and emotive for synthetic big data (as compared to the quantitative and the physical for natural big data).

CONTROL FIELD VS. IMMERSIVE FIELD: *how controls of the views can be modified by ostensibly external control methods, or through gesture based, immersive methods: relating to how we move through real worlds (intrinsic) versus libraries of knowledge (derived).*

Gesture-based control has already taken the field in consumer products. The idea of clicking here, here, or here to call up the information we are looking for is less desirable then gesturing through fields of information and ferreting out what we are interested in finding. The journey is far from complete — many gesture based controls are still just controls, merely a bit more "hip" in execution. Let us consider three levels of data/knowledge retrieval form potential informative sources. The first is a non-controlled view, the next,field-controlled views (or stage-controlled views), the last immersive. In conjunction with this we should consider what may be called "core" or "adjacent" renderings.

Non-controlled views are mostly 2D renderings, or possibly 3D renderings in some cases (such as the famed dioramas at New York's Museum of Natural History). They may be static "single-shot" renderings, or temporal as with moving-picture imagery — film and video. All the data that is available is at the *same time available*, or *through-time available*. With same time availability the viewer moves him or herself through and around the model like a living cursor (as eye-tracking software might reveal). Information graphics fit into this first tier, but I would more desirably include information maps. Within information maps there is a heightened value to the use of the informative by merely moving in any direction around or across the surface. One's eyes, supported by the ability to physically move and visualize and process the data, are the toolsets that renders the informative representation. This can be particularly well understood through very large scale maps and visualizations. Ben Shedd of Princeton University deals with the idea of "exploding the frame" where one is not confronted with the frame or border of the image because the image is large enough to extend beyond the areas of peripheral vision. So, the viewer, cursor-like, moves about to immerse themselves in what is spread out before them.

Big data could work through such a rendering but issues of technical capability and cost of rendering might be significant. One can see how such renderings bring us back to similitude to how we move about in the actual world; these renderings are in controlled spaces so issues of speed and comprehension (in addition to the physical infrastructure) emerge. Such large renderings do not exclude the idea of engine composition as previously discussed: there is no size limit for an engine composition.

The Eames' film *Powers of Ten* is both a movie and an engine composition (which is something of a feat), still a classic after nearly forty years. With non-controlled engine composition the viewer learns by choice; the intelligence in such models is instilled through a careful consideration of content, through the structural logic of where content will be rendered singularly and in context to other content. This is supported by the visual manifestation respecting design issues of shape, border, color, luminosity and myriad other aesthetic considerations. These are usually fairly hard to make well because all the information must be present through one of two means: comprehensive integration or successful storyboard, but when these are present the results, being well made, are highly informative.

Controlled-views, driven by the art and science of traditional interactivity, permit users to modify renderings through any number of controls external to that rendering. The stage, or field, can be practically endless in its possibilities (if the data is dynamic then of course it is endless in possibilities). Concerns about real estate are a major factor in the design of these interfaces because every rendering is by default a compromise, the viewer is always not seeing something else.

The other issue with the staged view is the user, generally, has to know where they want to go, or even more real estate must be utilized to show them where to go, and how to control this procedure. For this reason the idea of intuitiveness is a crucial factor in the design of most toolsets with stage views. Originally all the controls of the stage were outside the view in question and users worked from their keyboard. Mouse controls permitted areas within the staged view to be clicked; this added a sense of "immersion" but through an off-hand method. With gesture based touch-screen capabilities the users can now render many alternative views by interacting with the current view; this then is a kind of segue into full immersiveness. Almost all the user interfaces today are controlled view systems which, at best, are compromises for generating knowledge discovery views from big data. This is because of the legacy of the kinds of visualizations generated, combined with the ways by which these visualizations were navigated.

Geospatial renderings (i.e. real imagery), network diagrams, graphs, spreadsheets, time series renderings, bar charts, and scatterplots all have navigational languages that are more-or-less specific to the tasks at hand. Geospatial renderings are very well advanced GIS tools have well developed immersive navigation capabilities. Scaling, sliding, accessing annotation, as well as deeper capabilities such as adding and deleting layers of metadata or running time sequences have received a great deal of attention from developers primarily due to US Government interest and funding to advance such capabilities. These kinds of navigational non-intrusive rendering capabilities will be very easily adapted to the types of imagery that could be developed for big data metapictorial imagery as well.

The future is the most direct immersive environment possible; where the concept of "gestures" is extended to hand, voice, eye tracking, facial expression, even thought directives. Such high levels of sensitivity will require very advancing capabilities for correcting the navigational process, reversing, and developing procedure improvements through machine learning and feedback loops. Equally important will be non-intrusive capturing of the relevant findings.

In many ways the entire world of interface design and control can be seen as the middle stage between the: non-navigational, unidirectional world of the non-controlled, yet very high quality informative visual representations of the physically pictorial examples of the past. As well as the and the fully navigational, all spectrum metapictorial imagery that will be possible in future. These conceptually immersive renderings, driven by an "understanding" of the *affinities, dependencies, and exploitations* models of relational synthetic big data could provide the kinds of knowledge discovery that solve mega-problems or prevent harm to humans, ecosystems, and economic stability.

To conclude this section it behooves us to discuss the idea of adjacencies. These are all the elements of information that are not intrinsic to the Systems composition vs. Engine composition, instead they are supplemental yet provide critical guidance to effectively navigating any kind of informative visualization. A full paper is available on this idea from PJIM (Parsons Journal for Information Mapping) publications *(Complications and Adjacencies An Organizing Logic for Information Graphics, Anderson, Bevington; Volume II Issue 3, Summer 2010).* For our purposes the following is worth considering —

"The composing of intelligible patterns from the noise of raw data is a hallmark of a good information designer. The most successful examples extract and present essential relationships in a coherent manner while limiting the obtrusiveness of accessory relationships. Effective results are self-evident whereby the information graphic is absorbed by the mind holistically. Such clarity often belies the intense efforts involved: like a baton race, all the work is concentrated to a point just before being passed on to the next participant in the informational relay. To this end, the designer applies a pattern or grid to position all the interrelational data fields. We call this process stacking: the mechanism for creating a beneficial complication whereby users see and understand holistically, which we consider to be cognitively superior to linear presentations. The success of layered compositions depend on the appropriateness of the basemap (pictorial, relational, quantitative, or symbolic) and the quality of the designer's integration. *What can be correlated should be correlated. What cannot be inter-relationally correlated, such as titles, labels, metadata, etc., should not interfere with the stacking grid since they introduce noise. Any "noisy" element is better brought "outside" the main grid and handled as an adjacency."*

It would probably be useful if such adjacent information was handled by parallel devices (such as say, something like the Google Glass, which displays information through a prism device to the upper right of the right eye's peripheral vision [until one glances up to read the display]) This would take full advantage of other kinds of augmented viewing devices that play their role into a totally device-free capabilities.

Our bias then is full immersive capabilities with augmented systems for adjacent information — it is the formatting of the delivery of the adjacency components that would require customizing user-to-user while the core immersion into the metapictorial would be universal. In this manner collaborative usage to any scale of users would be possible.

SYMBOLIC VS. SIGNIFIED: *understanding the distance from the core signified thing and how cultures share, or create new meanings, as distance of time or space move the viewer away from the signified elements — and the use of symbols as compacted elements of pictorial things.*

For this, the tenth factor discussed in our sequence of bias toward conditional splits we are challenged with the fundamental idea of communicating findings from the results of knowledge discovery. The challenge is well recognized: what might get lost in translation and how do we mitigate this loss of clarity? Again we are going to stand the simplification game, somewhat, upon its head. Information designers are (or should be) obsessed with clarity and simplicity as a mode of reducing noise, focusing attention to key aspects, providing exactly the right hierarchy of informational delivery, and supplementing the presentation with adjacencies that return the subject to the most effective pathway to understanding.

Intriguingly, this is often accomplished with a very high level of concision. Such concision takes the form of symbols and high levels of simplification and "rounding off" of smaller exceptions in favor of making critical comparisons of the larger exceptions. Transit maps are a fair example of this, where the richly complex, dirty, noisy, tactile potency of railway lines, ties, steel, and concrete, are all reduced to uniform, brightly colored lines that run beautifully parallel and occasionally take 90 (or 45)degree turns over maps of minimalist polygons and precision notation. Lovely. In truth, this is radical symbolic simplicity of what the transit systems are. And, it is effective, so much that Harry Beck's London Underground map of the mid 1930s has permeated the design of most subway and transit line maps in use in the world today (although the printed paper versions are in rapid decline.)

The issue of the signifier and the signified grows as the users of increasingly complex systems are more and more distant from the decision-makers and findings must be converted form expert jargon-rich language to actionable, simple intelligence. Another factor that appears to be on the rise in institutions with ever increasing workforces is a propensity for decision-makers that have not moved "up through the ranks" but "across ranks." This means that there is less of shared work-culture and viewpoint variance.

The translation of the "findings medium" into the "reporting medium" also adds complication; standardized reporting mechanisms, designed to assist in the rapid dissemination of findings are often woefully inadequate in accurately capturing the subtleties of knowledge discovery. What results from all these potential misalignments might not be the advantages of simplicity, but what I will call "simplexity." By this I mean ineffective simplification resulting in a loss of the critical message in favor of a message that satisfies the need — but not the condition of knowledge discovery.

Further, the current cultures working across the vast range of disciplines supported by big data require a kind of integration that defies a uniform understanding of symbol, metaphor, and signifiers. The answer here may be the capture of the essential story directly from the source imagery. This builds upon the use of the tools that support the immersive direct user within the metapictorial renderings. Knowledge discovery within this framework can be replayed as would be a film, and as with any film that is viewed in a language that is not one's own it can be played with the appropriate sub titles. In this case the subtitles can possess a very rich symbol set that clearly conveys. This could come as close, say, as a book well translated from the original tongue; not perfect, but far better that re-translations that scrap all that was originally understood through a model that coincides more with reporting than informing.

DEFICIENCIES OF METAPICTORIAL MODELING FOR BIG DATA

The section previous touches upon multiple issues that will be of concern should a successful, highly informative model of synthetic metapictorial imagery be developed for big data informative visual representations. In essence the whole idea of synthetic metapictorial imagery is that it will mimic the experience of moving about, as sentient beings, in the real world.

The same problems that face scientists and designers in the real world, will face the engineers, scientists, designers, analysts, and innovators in the synthetic metapictorial world. Namely, how do I *handle* this information? How do I capture my unfolding findings? share it? how do I work with others in its extraction? how do build simplified models of composite findings? The main concern here is to essentially move back down from a fully immersive world to a controlled one.

By comparing FIGURES 9A – 9C to FIGURES 16, –19 an answer emerges. FIGURES 9A – 9C unfold the potential effectiveness of metapictorial modeling by density and degree across a range. Akin to a melodic process across time. However, the logic of collaboration requires a kind of harmonic requirement, deriving value from within a framework of time. FIGURES 16, –19 show how the granularity of a synthetic model can be immediately "downsampled" to graphical models that can handle all the tasks which those who are tasked with knowledge discovery can turn to, must turn to, when they step outside their singular investigative world and share findings. Additional description here would be fully redundant to the captions on pages 19 and 20.

CONCLUSION

Synthetic, human-centric *qualitative big data* is at the threshold of mimicking the kinds of "natural" data that can be derived from physical *quantitative* phenomena. This is so because the interstitial space of qualitative data is rapidly decreasing as the gaps of incompleteness are just beginning to fill (this parallels the history of modern science as it reveals and explains nature, thus decreasing the unknown and formerly numinous). The advances in scientific knowledge came primarily from direct or augmented observation. These observation are revealed through pictorial, *real imagery*, or signal emission and capture (alternative kinds of "imagery"). These images, in turn, have been found to possess quantities (numbers) within discernible patterns (it may be argued that every other type of map, graph, chart, diagram, table or symbol is a kind of reduction or extraction from these realistic images).

Further, these higher-level, realistic images, as driven by quantities and underpinned by physical laws, are reliable and repeatable (to the extent that this is understood within the limited number of dimensions that can be observed and calculated concurrently). I argue that big data may also understood through exactly the same means: image observation. However, computational formulas, and the rendering algorithms derived from these formulas might, for big data missions, not be developed to directly address quantities and physics (as with natural investigations), but instead address the *connectivity and relations* within the larger scope of big data — thus capturing a softer series of variables as determined by more experimental taxonomies and ontologies. For, unlike quantities which reveal the underlying physical properties in nature, the relationals in human-generated big data may instead be underpinned by forces of affinities, dependencies, and exploitations (a kind of physics of social-interconnectedness).

The images generated from such renderings (which I refer to as metapictorials) might allow significant levels of knowledge discovery. This is particularly the case if the renderings are large-scale and fully immersive and designed without any need for customization. Customization can be handled through augmenting tools that are "outside" these metapictorial renderings (forgoing customization means immediate large scale user-pools and collaboration). The findings from those who explore such metapictorial worlds may be immediately shared with others via the same kinds of augmentation tools which permit translation at the recipients level of expertise and interest. (By toggling between immersive, gesture-based interactivity to control-based graphic renderings.) In essence, a cycle of *Natural Pictorial Imagery,* to *Quantities underpinned by physics;* — to *MetaPictorial Imagery,* to *Relationals underpinned by affinities, dependencies, and exploitations,* can be understood as the visual rendering pathway allowing a far richer exploitation of big data resources. The high potential for knowledge extraction from such metapictorials may support significant problem solving as we continue to amass big data collections. Such metapictorials may also point to areas that are most profitable for further data collection, ever enhancing the cycle of knowledge discovery.

WILLIAM M BEVINGTON is the Senior Information Theorist at PIIM, The Parsons Institute for Information Mapping. *bevingtw@newschool.edu*