Bamboo in a Sandpile

Methodological Considerations for Leveraging Data to Enhance Infrastructural Resilience

Robert Spousta III, Steve Chan
Sensemaking Fellowship
San Diego, California, The United States of America
spousta@mit.edu, stevechan@post.harvard.edu

Stef van den Elzen, Jan-Kees Buenen
SynerScope BV
Helvoirt, The Netherlands
s.j.v.d.elzen@tue.nl, jan-kees.buenen@synerscope.com

***Abstract*—The onset of the Big Data phenomenon presents significant technological challenges in managing massive amounts of information, yet it also presents tremendous opportunities for enhancing societal resilience and directly serving the public good. The Internet of Everything, which is driving such massive connectivity and growth in data generation is a highly complex system, continuously giving rise to new communication capabilities, yet also becoming increasingly vulnerable to destabilizing forces and malicious threats. Creating systems that are truly intelligent and capable of balancing these interrelated dynamics in the management of data demands a deliberate approach that is scalable, adaptive, and extensible. In this paper, we discuss three primary considerations for conducting Collaborative Big Data Analytics, including data acquisition, layered analytics, and visualization in order to grow resilient cyber-physical infrastructures that are capable of withstanding significant destabilization. With regard to data acquisition, we present the basic characteristics of so-called Big Data, namely the Six Vs of data variety, volume, velocity, veracity, volatility, and value. In addition, we outline the development of analytical tools and techniques for processing data, as well as methods for effectively visualizing the products of a layered analytic approach. In order to illustrate the utility of such an approach, we summarize findings from our participation in Orange Telecom's Data for Development Challenges in the Republic of Côte d'Ivoire and Senegal, as well as introduce initial findings from our ongoing study of infrastructural resilience in archipelagos. We conclude that while Collaborative Big Data Analytics hold great promise, forums for the open development and validation of methodologies for its conduct are needed to generate more and better uses of the Big Data that have come to dominate our world.**

Keywords—Collaborative Big Data Analytics; Decision Engineering; Infrastructural Resilience; Sensemaking Methodology

I. INTRODUCTION

This paper is an extension of work presented at the 2015 IARIA Data Analytics Conference [1]. Whereas the dot-com boom of the late 1990s and early 2000s ushered in a wholly novel industry, replete with information-based products and virtual services marketed via the Internet, collaborative approaches for conducting civil-centric and public service-oriented data analytics have taken longer to develop [2]. This fact notwithstanding, the rise of the Internet of Everything (IoE) has introduced unprecedented levels of artificial complexity within many cyber-physical systems, which demand constant attention, lest areas of brittleness and blind spots compromise the delivery of essential services through infrastructures that are the backbone of modern civilization. In bridging this gap, we present three basic layers that comprise a framework for gaining insight from data. In previous work [3], we posed the question of whether the protection architectures of critical infrastructure are improving or deteriorating with age; in other words, are they more like milk or like wine? Our investigation suggests that in the case of electric grid systems, infrastructures become more vulnerable with age, particularly as new threats evolve more quickly than existing protective measures are able to adapt. To ameliorate such a circumstance and improve the security and stability of critical infrastructural systems like the grid, we advocate for increased data collection, and more robust analytic capability employed in a Big Data Paradigm. This is illustrated by our juxtaposition of bamboo and the sandpile. Whereas the Abelian Sandpile or Bak-Tang-Wiesenfeld model serves as an effective metaphor for self-organized criticality and cascading effects in complex systems [4], the structural flexibility and dynamic responsiveness of bamboo [5] characterizes a system in which adaptation facilitates resilience. In turn, adaptation is facilitated by a timely and precise leveraging of data. The **Sensemaking Methodology** addresses three primary concerns, namely, the capturing of data, the processing and refinement of data into insight, and the visualization of insight to guide Decision Engineering endeavors. In this manuscript, we briefly outline the system of methods that comprise our three-layer framework, as illustrated by ongoing research focused on the resilience of critical infrastructures such as electric power systems.

The remainder of the paper is organized as follows. Section II discusses some of the primary considerations for data identification and capture, including the variety of sensor platforms that are responsible for producing data. Section III describes the basic categories of analytic tools and techniques that have been developed for data processing, and argues the importance of counterpoising heuristic and algorithmic analytics, which is a core component of our methodology. Section IV addresses primary considerations for effective data visualization. Section V summarizes major findings and lessons learned from our participation in the first two Data for Development (D4D) Challenges as an exemplar of the Sensemaking Methodology for **Collaborative Big Data Analytics**. In Section VI, we articulate how such an approach stands to enhance resilience in complex systems, in particular the employment of data-driven isomorphic and biomimetic applications to critical infrastructures such as the electric grid. Specifically, we explore the context of power grid resilience, and discuss our investigation of a synchrophasor analytics system for archipelagos. We conclude in Section VII with general thoughts on the state of the art with regard to Collaborative Big Data Analytics, and identify areas for future advancement of our Sensemaking Methodology.

II. DATA ACQUISITION: PROSPECTING FOR DIAMONDS OF THE INFORMATION AGE

Just as various phases of the Industrial Revolution were fueled by human ability to derive value from the planet's natural resources though technology, the ongoing Information Revolution is being fueled by our ability to derive value from data through technology [6]. However, this derivation results in more than the generation of wealth. Data management impacts nearly every facet of society, from economic development and education, to international security and environmental stewardship. In formulating a data-centric approach to complex problem-solving of any sort, it is prudent to first observe basic characteristics of data that impact its selection and acquisition. Whereas multivariate criteria exist for evaluating the quality of natural mineral resources such as diamonds (e.g., the "Four Cs" of color, clarity, cut, and carat weight) [7], so too must data be evaluated from various perspectives. At the most basic level, the phenomenon of Big Data is being propelled by the so-called "Three Vs" of volume, variety, and velocity, which concern objectively quantifiable aspects of how much and how quickly different kinds of information are being communicated. However, in order to make any sense out of this torrent of data, we argue that an additional three qualitative aspects of veracity, value, and volatility are of equal importance, as depicted in Table I below.

TABLE I. CHARACTERISTICS OF DATA

V	**The 6 Vs of Big Data**	
	Description	***Units of measure / Dimensions***
Volume	Massive amounts of data	Bytes => Petabytes
Variety	Multiple forms / formats	video, sms, .pdf, .doc, .jpg, .xls, .rtf, .tif, PMU, etc
Velocity	Speed of data feeds	Event-driven / Streaming
Veracity	Trustworthiness of data	Provenance / Pedigree
Volatility	Shelf-life of data	Time-Sensitive / Static
Value	Usefulness of data	Ambiguity / Uncertainty; Correlation / Causation

a. An alternate V of Viability has also been proposed in [2], which we believe is subsumed above

Size does matter. The Big Data phenomenon is perhaps most commonly linked with the sheer volume of data being generated by a host of remote sensors, household appliances, mobile communication devices, and human content generators worldwide that totals over 2.5 quintillion bytes of data per day [8]. Although difficult to comprehend quantitatively, these reams of data come in many forms, from the millions of photos and videos shared daily from smart phones through applications like Facebook, Sine Weibo, and Snapchat, to telemetric data and raw system measurements recorded by a multitude of sensor types and fed into industrial control systems (ICS), transportation management networks, meteorological forecasting services, and other information management systems [9]. Whereas human beings have historically been the primary generators and collectors of data contributing to knowledge development, with the number of devices connected to the Internet surpassing the global human population in 2008 [10], machines are now responsible for an increasing preponderance of the world's data. Indeed, the growing linkage of people, data, things, and processes is central to the so-called IoE [11], and the driving force behind change in myriad interdependent complex systems. This massive increase in IOE-generated data is both a significant challenge and a promising opportunity. On the one hand, conventional mechanisms for capturing and analyzing data cannot scale up effectively to accommodate the explosive growth in data generation. On the other hand, such abundance can enable us to use data to guide our decision-making and problem-solving in ways that have not been possible until now. A key to unlocking this potential is the ability to rapidly assimilate huge volumes of data and accurately identify useful pieces of information.

Integrating a large variety of data stands to yield the most robust insights. In order to achieve quantitative exactitude in identifying insightful information, a maximally inclusive variety of data types and sources is essential. To generate a complete picture of a system, we must be able to view it from multiple perspectives. In this regard, a critical determinant for perspicacity is the incorporation of diverse data that each relate to a given system or problem set through unique angles that allow for cross-referencing and comparison. This includes the acquisition of both structured and unstructured forms of data. By way of example, when someone is interested in a particular topic or event, they may initially hear a broadcast about it on the radio, then read an article about it, download images, or watch a video on the Internet. As with the parable of the blind men and the elephant; each source of information takes a different form, yet contributes to a more complete understanding when taken together. Similarly, in researching issues of infrastructural resilience, we are striving to utilize a host of data gathering mechanisms, including the collection of electric power signals from monitoring equipment such as Phasor Measurement Units (PMU) and Digital Fault Recorders (DFR), to visual and geospatial data from Unmanned Aircraft Systems (UAS), Ocean Data Acquisition Systems (ODAS), Synthetic Aperture Radar (SAR) and other weather observation tools, to human sensor networks in the form of crowdsourced event observation and reporting.

Data velocity is a determining factor for agile systems. In addition to harvesting a large variety of data, the speed with which data are gathered and communicated is another significant variable, as time-critical operations from financial management and news reporting, to emergency response, law enforcement, and national defense all must be able to quickly sense the occurrence of anomalous events in order to operate effectively [12]. Several factors influence the velocity of data, including the capacity of communication channels, as well as the granularity of observations. The continuous expansion of fiber optic and wireless communication networks enable many individuals and sensors to rapidly exchange data. In addition, sensors are capable of recording measurements at increasingly precise spatial and temporal scales, resulting in more frequently observed change and data generation. At the same time, the increase in data velocity challenges our ability to keep pace. As information is communicated at greater speed, decision cycles are compressed, and we have less time to assimilate more information. To illustrate this point, consider that over 100 hundred hours of video are uploaded every minute to the video sharing site YouTube [13], which accounts for over 400 million years' worth of viewing time in the 11 years since the site's creation in 2005 [14].

Whereas adapting to these quantitative aspects of data are sufficiently challenging, simply capturing a large amount of fast-moving information from different places is not enough to generate improved insight. We must also consider the more qualitative aspects of data, which ultimately determine how useful it can be. In managing both emergency responses and routine system operations, all data consumers rely on the authenticity or veracity of data in order to gain actionable insight. The consistency of data taxonomy is an important aspect of veracity, and, in this regard, discovery standards for electronic resources such as the Dublin Core standards for Metadata are essential for datasets held by diverse curators to remain compatible with one another [15].

A more persistent challenge regarding veracity is the ability to establish the provenance and pedigree of data, particularly in the context of data manipulation and spoofing, or counterfeiting in the information supply chain. While gathering redundant data from multiple sources, and cross-referencing particularly specious data are prudent strategies for mitigating the negative impact of false or corrupted data, ensuring data veracity is a challenge that requires vigilance and adaptation. By way of example, the April 2013 issuance of a false tweet from the Associated Press's hacked Twitter account alleging injury to President Obama during an explosion at the White House caused the Dow Jones Industrial Average to plummet 142 points in just two minutes [16]. Such a resourceful, yet malicious use of technology illustrates that data need not be characterized by great volume or variety in order to generate massive impact. In turn, methods to ascertain and safeguard the authenticity of data must be equally resourceful.

Data veracity is particularly significant for operating and safeguarding critical infrastructures. With regard to electric power systems, the lack of a shared standard for grid performance metrics can compromise the value of system-wide measurements, as U.S. grid ownership is fractured between a diverse mix of privately-owned corporations, rural cooperatives, municipalities, the federal government, and a host of independent power providers, with over three thousand organizations distributing power to consumers, each with varying standards of performance monitoring [17]. Internationally, in addition to a lack of shared performance metrics, grid operations are also challenged by the lack of a shared grid event lexicon, which in extreme cases can actually prevent system interoperability [18].

In addition to its veracity, the shelf life of data also has a large impact on its utility. Volatility or duration of relevance depends largely on the nature of the decision which data are serving to inform. Whereas certain digitally preserved historical records maintain their relevance or value in perpetuity, other datasets that pertain to rapidly evolving circumstances may remain relevant for only a matter of days, if not seconds, or less. By way of example, international standards for the operation of power grids dictate that the onset of electrical islanding events amongst distributed power generating sites be detected and addressed in no more than two seconds [19]. Although the granularity of time series data is a vital consideration for decision making on compressed time scales, the continuity of data collection similarly impacts the longitudinal analysis of slower developing patterns. By way of example, a maintenance gap in 2012 of the Tropical Ocean Atmosphere array led to a 70% drop in data collection, thus compromising the consistency of measurements, and potentially skewing the analysis of long-term anthropogenic climate change and global warming studies [20].

Whereas an evaluation of each of the Four Cs are combined to determine a diamond's overall quality, the aforementioned Vs can be similarly combined to determine the utility of data. Data value loosely correlates to how much of any given decision can be engineered from it. In other words, can we decide a course of action based on a single dataset? If so, then that dataset is of high value. If many disparate datasets are required in order to engineer a single decision, then each of those datasets is of comparatively lower value, taken in isolation. Determinants of value include such factors as the level of ambiguity with regard to data meaning, as well as discernibility between correlation and causation (i.e., whether multiple variables simply change together, or whether a particular variable directly catalyzes change in others). Amongst the sea of data corresponding to myriad variables, a principal challenge is determining which pieces of data are the most significant indicators of change or phenomena of interest. Although no formalized schema yet exists for evaluating these Six Vs of data, recognizing the significance and employing methods for addressing each is a fundamental aspect of operating in a Big Data Paradigm.

The actual task of data acquisition is no less complex. For all organizations - public, private, and any permutation in between – how best to gather and disseminate data remain open questions [21]. With the United Nations (UN) having asserted that information in itself is a life-saving need for people in crisis, just as important as water, food, and shelter, the necessity of publicly-accessible data is clearly a global one that now transcends the realm of scholarly open access [22]. Yet, there is no comprehensive, authoritative single source for information, and so we must get data from as many places as we can, in as many ways as we can. By extension, an intelligent system is ideally capable of continuously ingesting data from multiple sources, through diverse media. However, there are a variety of practical limitations on such a capability, including human controls over data accessibility (e.g., personal privacy, political sensitivity, national security, commercial ownership, etc.), as well as technological challenges with data capture and curation [23]. Moore's Law for semiannually doubling transistor capacity, Gilder's Law for annually doubling communication bandwidth capacity, and Koomey's Law for annually doubling computational energy efficiency have each held steady for years to yield the current explosion of Big Data. Yet, at the same time, system input/output (I/O), memory, and storage capacities have each increased at a much slower rate [24], creating a dynamic whereby data is generated faster than it can be consumed, as pictured below in Figure 1.

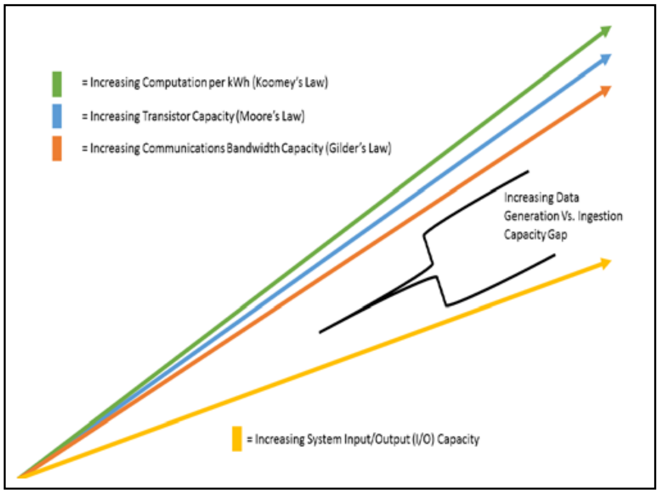

Figure 1. Technology Trends Impacting Big Data

Such limitations notwithstanding, an important component of our Sensemaking Methodology involves prioritizing the most critical data requirements, and where these cannot be met directly, identifying suitable proxies. Just as the UN employs rates of electrification as a proxy to gauge a nation's level of development, we can employ statistics such as the penetration of renewable energy sources to infer probability of electrical islanding and other disturbance events. As a commodity, data in and of itself is not particularly valuable. However, the more of it that we can gather, then the greater the chances are to yield valuable insights through intelligent refinement. Generally speaking, the onset of the Big Data phenomenon is not an automatic boon to the generation of deep insight into complex problem sets. On the contrary, Big Data itself presents unique challenges that necessitate new ways of working with information, through robust analytic techniques.

III. STACKING THE DECK: TOOLS AND TECHNIQUES FOR LAYERED DATA ANALYTICS

The analytic phase of the Sensemaking process is designed to extract actionable insights from the raw material of massive datasets. It is constituted by layers of algorithmic and heuristic techniques, high performance, distributed, and cloud computing, machine learning, signal resolution, video analytics, and natural language processing, nested under the Unstructured Information Management Architecture (UIMA) [25]. Whereas the significance of structured information such as that contained in relational databases is by nature unambiguous, the various components comprising UIMA are geared towards discerning meaning and significance from a variety of unstructured information sources, such as sensors, video, and natural language. Many of these components are rooted in mathematical concepts that have developed over centuries, and therefore some brief accounting of their evolutionary pathway is instructive.

Mathematicians and logicians of the 17th century realized that the painstaking work of numerical calculation could be conducted automatically in order to free up the mind to focus on higher level analytical work. First the Pascaline, then the Leibniz Computer represent some of the earliest automated calculating machines, later advanced in the 19th century by Charles Babbage in the form of the Analytical Engine, which was the first programmable computer designed to solve a variety of mathematical and logical problems [26]. Although Babbage's inventions are certainly significant achievements, the detailed instructions that his protégé and peer, Augusta Ada Byron Lovelace wrote for using the Engine are arguably even more lasting for being the first computer programming code [27]. Indeed, a case can be made that modern computer programming evolved from Lovelace's early conceptual programming work, as indicated by the fact that one of the U.S. military's first high level programming languages was eponymously called "Ada" [28]. In addition, George Boole's work on representing the process of logical thought in binary mathematical form paved the way for digital computer logic and the electromechanical switching processes that remain foundational to the operation of computers today [29].

While providing early functional models for computational intelligence, the work of early mathematicians and philosophers also served to inform the ontological basis on which many modern platforms have been constructed. Specifically, the theories of knowledge and logic proposed by the American Triumvirate of Pragmatism in the late 19th century have played a significant role in influencing the trajectory of the Information Revolution in the 20th and 21st centuries. The Triumvirate, comprised of William James, Charles Pierce, and John Dewey, established three postulates that would later serve as the philosophical underpinning for modern information retrieval and search engine systems, namely applications of the semblance of indeterminacy, order in chaos, and long-run convergence [30]. When combined with such an ontological orientation and the invention of semi-conductive transistors, the Shannon-Weaver model of communication opened the door to modern information and communication technology, by establishing a theory of information that conceptually integrated disparate elements of data source, message, transmitter, signal, channel, noise, and receiver into a coherent system [31]. Finally, the works of mathematician and cryptanalyst Alan Turing and others at Bletchley Park to unlock the signals intelligence encrypted by Nazi Germany's Enigma Machine, as well as John Mauchly and Presper Eckert's Electronic Numerical Integrator and Computer (ENIAC) [32] not only helped to turn the tide of World War II, but also gave birth to the field of computer science [33].

From this shared lineage, the modern analytical toolkit of computation has evolved into far too many instruments to concisely summarize here. However, there are fundamental components of the analytic process, which we will strive to articulate. Upon acquiring data, the initial step in the analytic layer of our framework is data ingestion and cleansing, which can actually account for up to 80% of work in data science [34]. By way of example, satellite imagery is unfortunately not as simple as an "eye in the sky" beaming down neat pictures to a computer console for analysis and distribution. The many 0's and 1's that make up the digital representation of a physical object must first be processed and translated into an intelligible picture. Once raw data are refined into a malleable commodity, that commodity can then be annealed into meaningful insight through a systematic layering of Analytics on Analytics (A2O). The most critical aspect of A20 is the contrasting or counterpoising of diverse datasets. This process begins with a foundational geospatial and or temporal matrix of data points, and proceeds through a set of systematic organizational steps that include data clamping, normalization, and hierarchical clustering, in order to reveal patterns and detect aberrations.

Given the importance of data veracity, a significant aspect of analytics in a Big Data Paradigm is the ability to recognize latent relationships in seemingly unrelated phenomenon. The case of the Boston Marathon Bombing well illustrates this point, as minor human error inputting one of the perpetrator's names into a federal database prevented law enforcement and intelligence officials from recognizing a critically predictive event in the month's leading up to the attack [35]. In a Big Data Paradigm, the mistaken addition of an extra "y" included in the Treasury Enforcement Communication System (TECS) record for Tamerlan Tsarnaev would not have prevented the issuance of an alert to detain him during re-entry to the United States from Chechnya, upon the recommendation of the Russian State Security Service (FSB). The application of fuzzy logic [36] and similar techniques in an A20 approach accommodates uncertainty in information granulation, which recognizes approximate relations in data instead of relying solely on exact similarity and total certainty. In general, morphological analysis and aberration detection serve to evaluate the role of myriad variables in the dynamics of complex systems and networks over time. This is depicted below in Figure 2, which is a snapshot from the SynerScope visualization suite showing change in international communication network connectivity.

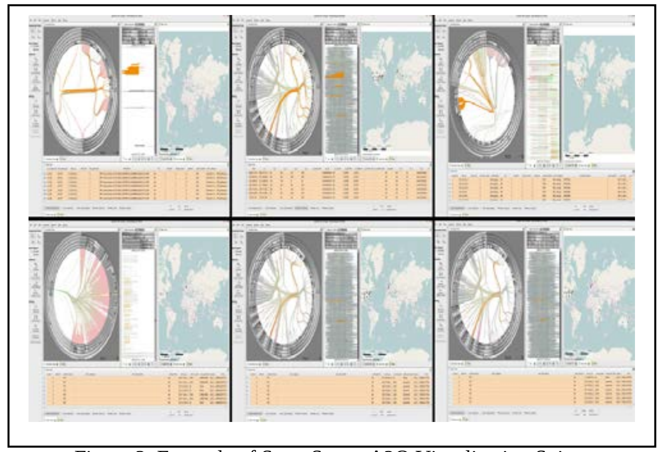

Figure 2. Example of SynerScope A2O Visualization Suite

Whereas many high performance computing applications rely on synthesized data (i.e., Monte Carlo Simulations, etc.), our approach is predicated on the acquisition and analysis of empirical data. However, a fundamental prerequisite for effective A2O is the storage and management of massive datasets. In this regard, distributed computing architectures and parallel processing are critical capabilities [37].

Impressive though they may be, machine capabilities comprise but one half of the analytic layer of our methodological framework. The remaining half relies on the inherently human capabilities of contextual orientation and intuitive leaping [38]. Whereas machines are capable of generating, processing, and storing massive quantities of data, the human mind remains unique in its ability to superimpose context over data in order to discern relevance and meaning. Hence, the Sensemaking Methodology is characterized by its fusion of technological and sociological perspectives. On the one hand, we leverage the technical advantages of machine capability in an algorithmically inductive pathway accumulating specific observations to build generalizable insights. On the other hand, we also leverage intuitive capacity in a heuristically deductive pathway applying general knowledge in particular circumstances to yield precision insights.

This socio-techno unification is at the heart of our methodology for pattern recognition and Decision Engineering. Going back to the example of satellite imagery, let us consider the case of the Global Earth Observing System of Systems (GEOSS) and the view of Somali villages at night as an illustration of counterpoising algorithmic versus heuristic insight. With the rise of both maritime piracy off the Horn of Africa, and the violent extremist organization Al-Shabaab in Somalia, international security organizations were keen to establish a link between the two groups [39]. As assets in the GEOSS satellite constellation observed significant variances in the night-time illumination of various towns along the Somali Coast and provincial capitals, analysts sought to employ the algorithmic insight of seeing more lights at night as evidence for a correlation between the dispensation of pirate ransoms and the buildup of jihadi strongholds [40]. However, heuristic insight suggested that the ideological and religiously-motivated nature of Al-Shabaab was incompatible with the financially-driven motives of the criminal piracy network, and therefore a link was unlikely. The truth of this insight would later be established through data gathered by the International Criminal Police Organization (INTERPOL) and the United Nations Office on Drugs and Crime (UNODC) [41], which verified that although pirates invest in infrastructure improvements (e.g., electrification and lights), Al-Shabaab degrades local infrastructure to fund arms purchases and maintain secrecy. Such an example shows us that while technology and algorithms are more than capable of identifying patterns of interest, we still need heuristic insight to decipher what those patterns actually mean.

IV. A PICTURE TELLS A THOUSAND WORDS: IMPARTING INSIGHT THROUGH DATA VISUALIZATION

Upon recognizing patterns of interest through an analytic process, relevant insight must be visualized in a way that directly informs the engineering of decisions. The primary aim of the data visualization phase is to establish the relevance of insight gained through the A20 process, and help to guide the actions of decision makers by parsing out critical points of useful information from massive amounts of data. Figure 2, above, displays output from one of our visualization platforms, the SynerScope. SynerScope and other similar tools use a coordinated multi-view approach with a scalable and flexible visual matrix in order to visualize key morphological insights into how complex systems and networks change over time. However, before we progress into any further detail with regard to contemporary visualization techniques, let us briefly consider data visualization in its broader context.

Efforts to visualize and impart insight are as old as human knowledge and communication; from cave paintings, to pictographs, hieroglyphs, numerology, symbolic logic, and language. In order to understand what methods have been developed over time for effectively conveying knowledge, it is instructive to visit certain historical examples. One case in point is the work of the Mixtec civilization of Oaxaca, Mexico [42], depicted above in Figure 3. Although the figure above depicts the Mixtec's primordial cosmology and creation mythology, it is an early example of how human insights gained through observation of natural phenomenon (i.e., data analysis) were preserved for distribution and posterity. This and other similar precedents from early civilization remain germane to many data-related fields, including Education, the Arts, Public Information, Manufacturing, Product Advertisement, Device Instruction Manuals, Traffic Signage, Emergency Management, and Information Technology (IT) [43]. With the advent of the Internet, and eventually the World Wide Web, the tradition of data visualization has continued to evolve. Today, such professional disciplines as Cognitive Science, Behavioral Psychology, Computer-Assisted Design (CAD), and Strategic Communication all build on the work of early visualization specialists by combining machine capability with human insight to generate socio-techno innovations in how the brain senses and interprets information. In turn, our interpretation and assimilation of information drives our ability to engineer decisions and determine appropriate courses of action, as individuals in daily life, as agents in organizations, and as members of the global citizenry.

Nevertheless, this does not mean that modern data visualization is a perfected science. Rather, visualization is a principled art that requires both intelligence and intuition in its composition. In turn, efforts to visualize pseudo-insights that are not informed by robust analytics run the risk of proliferating misinformation, bias, conflict, and spoilage of resources [44]. In addition to these pitfalls, data-informed visualizations also can be subject to information overload, if insights are not concisely crystallized in a digestible form, as depicted below in Figure 4 [45].

Figure 3. Image from the *Codex Vindobonensis Mexicanus*

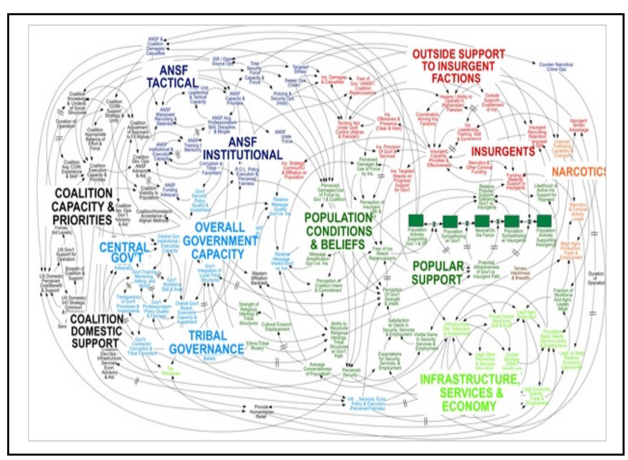

Figure 4. Example of Counterinsurgency Diagram

The design of any given data visualization is driven by two primary factors; the nature of the decision it serves to engineer, and the demographic characteristics of the audience or decision maker. Firstly, is the aim of the visualization simply to impart generally useful information, or is it intended to inform a specific choice? If the aim is the former, then visualizations such as that in Figure 4 may be suitable. However, decision-quality visualizations must clearly depict actionable insight, and inform implementable courses of action in a timely manner. Secondly, how much does the target audience for a given data visualization already know? An audience of laymen will require a significant amount of context in order to make sense out of visualizations depicting complex phenomena. Conversely, too much context will be superfluous (and potentially distracting) to an audience of experts. Therefore, constructing an effective data visualization means striking a delicate balance between sufficient context and specific insight.

With this in mind, we turn to a key consideration regarding the value of data visualization; the depiction of changing dynamics and identification of brittleness in complex systems. In light of the many layers of interdependence that characterize our most critical infrastructural systems (e.g., electric grids, the Internet, etc.), there is significant potential for percolation effects or cascading failure [46]. Therefore, to ensure the resilience of such systems, it is essential to closely track changes in system states over time, to identify areas of brittleness or weak links in the chain, and actuate corrective measures before these weak links fail. With regard to the resilience of the Internet in particular, tools such as the SeeSoft System, pictured below in Figure 5, enable analysts to visualize statistics of interest in software code [47]. In the case of Figure 5, a color-coding scheme displays how recently lines of code have been changed, with red lines having been most recently changed, and green lines having remained unchanged the longest.

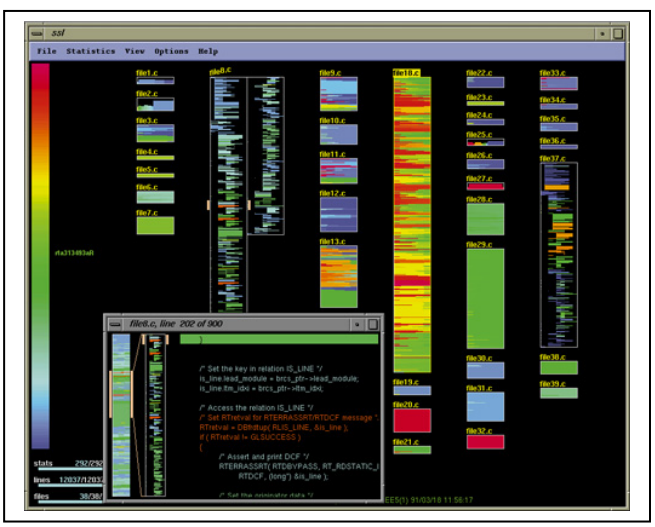

Figure 5. SeeSoft software code visualization system, Lucent Technologies

Visualization tools are invaluable assets that enable us to quickly and clearly see areas of potential brittleness in complex systems. In the case of Figure 5, above, we have a mechanism to visualize answers to questions such as whether software security improves with age, as lines of code not recently updated to address proliferating cyber threat vectors are likely brittle [48]. Therefore, visualization is not only a product of the analytic phase of the Sensemaking Methodology, but can actually be a feedback loop that helps to inform the A2O process. In general, visualization is a fluid endeavor, whereby patterns that reside in large amounts of data can be quickly and easily recognized by a human user, and help to guide their decision making amidst dynamic environments and changing circumstances.

V. PROOFS OF CONCEPT: SYNERSCOPE AND THE DATA FOR DEVELOPMENT CHALLENGE

To demonstrate the utility of our approach, we come to the shores of West Africa and the Data for Development Challenge (D4D) [49]. Since its inauguration in 2012, the annual D4D Challenge has represented a unique opportunity for Big Data analysts to experiment with diverse tools and techniques for harvesting insight from mobile phone data. For each challenge, international competitors from academia and private industry are given the chance to analyze a multitude of datasets pertaining to mobile phone use in a designated country during a circumscribed portion of the year [50]. We have had the privilege to participate in the first two such challenges, in the Republic of Côte d'Ivoire and Senegal, with a sampling of our results displayed below in Figure 6 [51].

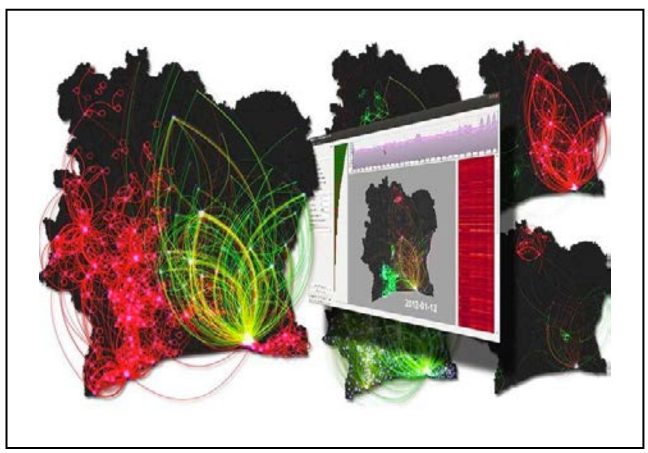

Figure 6. 2013 D4D Best Visualization: *"Exploration and Analysis of Massive Mobile Phone Data: A Layered Visual Analytics Approach"*

In conducting our analysis of the D4D datasets and generating the illustrations sampled above, two lessons became clear to us. First, we needed a large variety of data sources, through which to contrast and correlate mobile phone activity with other significant trends and events. For the first D4D in Côte d'Ivoire, we contrasted the given mobile phone data with UN reports of violent conflict and significant social disturbance, as well as meteorological data for the given timeframe. This helped to reveal regional political affiliations and ethnic enclaves, as violent events targeting certain political and ethnic groups in the capital city, Abidjan, catalyzed notable increases in call activity to specific communities elsewhere in the country. In addition, we observed that abundant rainfall in areas of significant cocoa and yam cultivation correlated with heightened call activity, likely indicating increased agro-business developments at specific points in the growth and harvest cycles in response to favorable weather conditions. Our second key learning was the need to adopt multiple perspectives from which to interrogate the datasets. Our normalization and clustering algorithms produced dendograms, with which we were able to sort items (e.g., cell towers) of similar calling behavior into groups for further investigation. By grouping cell towers of similar call behavior, we were then able to further explore what other commonalities linked these disparate regions.

Although such techniques are still relatively nascent, we believe that the work of our team and fellow D4D participants is a clear demonstration that Collaborative Big Data Analytics can help to increase insight into complex interrelated phenomenon, and thus improve Decision Engineering in a variety of social, political, and economic arenas. However, the implementation of our **Sensemaking Methodology** remains in the early stages, and inevitably there is room for improvement in such an approach. Specifically, increasing the volume and variety of data included in the A2O phase will yield greater insight in future D4D Challenges, and other applications of our methodological framework. In addition, the deliberate articulation of alternate frameworks for Collaborative Big Data Analytics will help to progress the state of the art, by revealing common best practices as well as shortfalls and gaps.

VI. KEEPING THE LIGHTS ON: THE ROLE OF SENSEMAKING IN SMART ELECTRIC GRIDS

In addition to work on the D4D Challenges, our exploration of infrastructural resilience also helps to illustrate the utility of a systematic approach for data acquisition, analytics, and visualization. As society has evolved, technology has advanced, and the complexity of systems has increased. In turn, this rise in complexity and interdependence leads to increased vulnerability in the social and physical systems upon which we rely for essential services [52]. In response, a resilience-oriented approach assumes that unpredictable and destabilizing events will inevitably occur, and accordingly focuses on how flexibility and adaptation can be instilled across systems. Making good use of data is central to such an effort. Therefore, we present initial findings from the application of our Sensemaking Methodology to an investigation of electric grid systems, and the development of a synchrophasor analytics system for archipelagos.

The concept of resilience is a core area of study in a wide variety of disciplines, from human psychology [53] and childhood development [54], to ecosystems [55], economics [56] and disaster preparedness [57]. Generally speaking, resilience refers to the capacity of a system to absorb shocks while maintaining functionality. However, resilience is a highly conditional state and the determinants of system resilience vary depending on the nature of the system and the context of specific shocks or destabilizing forces [58]. Factors that promote resilience in one system do not always translate neatly into other system contexts. By way of example, while a diverse social network can enhance individual resilience, the interdependence of diverse infrastructural components may itself be a source of vulnerability for the collective infrastructural system. In addition, a particular system cannot be broadly characterized as either vulnerable or resilient in perpetuity, because each threat affects a system differently, and threats continually evolve.

Nonetheless, we maintain that insights generated from the study of resilience in social-ecological systems [59] do bear relevance for the promotion of resilience in cyber-physical systems, such as the electric grid. In particular, the so-called "R4" framework is a helpful tool for conceptualizing the key harbingers of resilience, namely robustness, redundancy, resourcefulness, and rapidity [60]. With regard to lifeline critical infrastructure systems like the electric grid, robustness; or the ability to withstand shocks is of particular importance [61]. Although rapidity (i.e., the ability to return to a state of normal functioning in a timely manner) is also a chief concern, it is an outcome of how effectively the redundancy and resourcefulness of contingency measures can augment a system's robustness.

Resilience can take many forms. In particular, research in ecological systems has evolved through two fundamental categories of systemic resiliency that differ over the balance between resilience and stability, or the flexibility to operate in multiple states of equilibrium or basins of attraction, as depicted below in Figure 7. Systems of inherent or engineering resilience are characterized by relatively low resilience and high stability, whereby systems operate within a comparatively narrow envelope of equilibrium that are designed for efficiency and productivity and thus do not tolerate instability. Systems of engineering resilience typically operate within a rigid set of parameters, or a single basin of attraction, and therefore have been able to operate effectively on a comparatively sparse data paradigm. In contrast, systems of adaptive or ecological resilience are characterized by relatively high resilience and low stability, whereby systems can function in multiple states of equilibrium or instability in order to persist or remain functional [62]. These multiple states of equilibrium or basins of attraction enable systems of adaptive resilience to remain viable in a more fluid set of parameters.

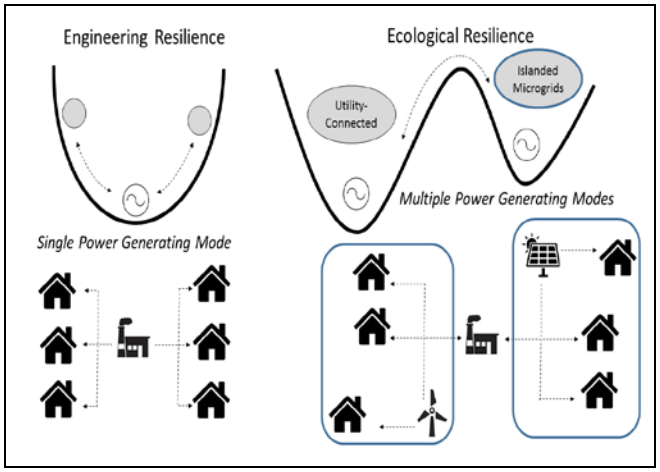

Figure 7. Engineering & Ecological Resilience, adapted from Scheffer et al, 1993

The ecological resilience concept pictured on the right of Figure 7 more closely approximates the aspirational "Smart Grid" of future power delivery. In this conception, resilience is a dynamic process that unfolds within a stability landscape, determined by a system's latitude, resistance, precariousness, and panarchy [63], which necessitates the adoption of a Big Data Paradigm in light of the many variables that must be managed within a given system. A resilience-oriented approach differs from a stability-oriented approach in that it does not require complete knowledge of all possible future events, but rather assumes that the unexpected will occur, and accordingly, it focuses upon devising systems that can respond adaptively to new and changing circumstances through the timely acquisition and analysis of granular data. These divergent conceptions of resilience raise interesting questions as to the nature of increasingly complex cyber-physical systems, such as the electric grid, particularly in light of the electric grid's shift toward a heterogeneous blend of power providers and distributed energy resources.

While engineering resilience characterizes the grid of the past and present, in which stability and efficiency are prioritized, the grid of the future has the potential to operate in an ecological resilience paradigm, whereby continuity is prioritized and power is provided in increasingly varied ways. In light of many technological advancements, perhaps the characterization need not be one of mutual exclusivity, but rather, how systems like the electric grid can be a hybrid enjoying the benefits of both stability and resilience. In this vein, a complementary focus in resilience research explores the effects of long-term change on the functionality of systems. In contrast to quick-onset shocks, the effects of gradual transformation can be equally disruptive and challenging to the resilience of systems, thereby demanding an adaptive capacity on the part of human operators [64]. The move towards a smarter grid represents just such a transformation, with multiple competing priorities that must be balanced in order to maintain infrastructural systems that are both sustainable and secure. This includes the deployment of newly developed hardware and software tools, as well as human capital investment for training operators to work in a more adaptive and data-centric paradigm.

Indeed, as societies become more reliant on an increasingly complex web of infrastructural systems, the need for both resilience and stability cannot be mutually exclusive. The imperative for many infrastructural providers to operate their systems at a profit adequately incentivizes stability in routine operations. However, a comparably salient incentive to invest in measures that enhance resilience to rare — yet devastating — black swan events [65] is lacking. The interdependence of modern critical infrastructure systems is itself a chief vulnerability or blind spot**,** in that the disruption of one critical service or system can potentially catalyze the catastrophic failure of all systems [66]. As evidence of this potential, we need only consider how many devices in our homes and communities rely upon electricity to function, from refrigerators and cash machines, to life-saving medical devices and water treatment facilities. In the event of a power outage, any such device not supported by its own independent power source would cease to function. When disaster events compromise the operation of critical infrastructure systems, the potential arises for situations to quickly transform from emergencies into crises. In such scenarios, societal or community resilience and the ability of citizens to cope effectively with crisis becomes a crucial consideration, albeit with its own set of unique challenges [67].

To hedge against such catastrophe, enhancing the resilience of even a single infrastructural component increases the collective resilience of the entire mosaic of critical infrastructure. Therefore, the focus of our study is on enhancing resilience in the electric grid as a vital component in the broader infrastructural system. In doing so, we aim to provide a blueprint for Decision Engineering that can be translated to other infrastructural sectors, public services, and missions that are challenged with the management of large and complex systems.

The first step in this effort is understanding data related to an electric power system, and the means for acquiring it. An electric grid is an integration of four distinct networks of electricity generation, transmission, distribution, and consumption, each of which produce distinct metrics. While each of these networks present their own idiosyncratic challenges to be overcome, the overarching problem for the grid is that electricity supply must constantly satisfy ever expanding consumer demand in a reliable, efficient, and increasingly sophisticated manner [68]. In order to do so, electricity is transported through many buses or nodes and, in many cases, over long distances; the overload or failure of any single node or edge between nodes forces the redistribution of load to other nodes, which can compromise the operation of the entire network through a cascading or percolating effect [69]. Percolation is not unique to electric grids, and other networks that have demonstrated potential resilience to the phenomenon appear to share common topological features, such as modularity and long path length across the network, which serve to isolate disturbances, provide alternate flow routes, and delay total network exposure [70]. However, robustness to percolation is not a viable solution, as grid operators must also deliver electricity efficiently, which excessively circuitous or entangled transmission and distribution lines would preclude. In contrast, optimally efficient networks are characterized by short path lengths formed around highly linked central nodes or hubs [71]. Ideal grid architectures strike a balance between resilience and efficiency by featuring a core of interlinked hubs and a periphery of leaf nodes, which facilitate connectivity throughout the network while maintaining resilience to percolation [72]. However, the North American bulk power grid was not designed with resilience in mind, and it contains a very limited corpus of hubs, which are so highly connected that it has been characterized as a scale-free network [73]. While these hubs are the main source of the grid's connectivity, they are also a critical vulnerability if they are compromised.

The development of smart grid technologies is precipitating several significant changes in data generation, network topology, and system dynamics that impact resilience. First, automated metering infrastructure and other communications enhancements are increasing the volume, variety, and velocity of power system data. In addition, power generation resources are becoming increasingly diverse and decentralized, with the flow of electricity transitioning from unidirectional to bidirectional as energy consumers also become part-time energy providers [74]. These changes add layers of complexity to electric grid operations, which in turn drives an increase in the amount of information required to maintain operability that exceeds human capabilities to process in terms of speed and volume [75]. In naturally occurring complex adaptive systems, biodiversity is an asset that enables various components to self-organize effectively in response to disturbances [76]. As the electric grid becomes an increasingly complex system of diverse components, it is prudent to consider what measures can be taken to catalyze resilience as an emergent property among these components.

In this regard, key principles from corporate enterprise governance provide an informative isomorphic contrast to those of ecosystem resilience. For service-oriented enterprises that must cope with discontinuity and uncertainty, the ability to sense and respond to change is paramount; context and coordination replace command and control as procedural operating paradigms, with the addition of an adaptive loop that facilitates systemic learning and the ability to improve responses to successive perturbations [77]. The Sense and Respond concept was originally articulated in the context of managing large corporations with diverse teams working on a variety of missions; however, it also bears relevance to a large physical system with diverse components that must fulfill a variety of functions. In this regard, the need to acquire maximally granular data regarding the operational status of the grid's various components becomes a principal requirement, and the advent of phasor measurement units (PMU) or synchrophasors represent a significant increase in granularity over conventional supervisory control and data acquisition (SCADA) capabilities [78].

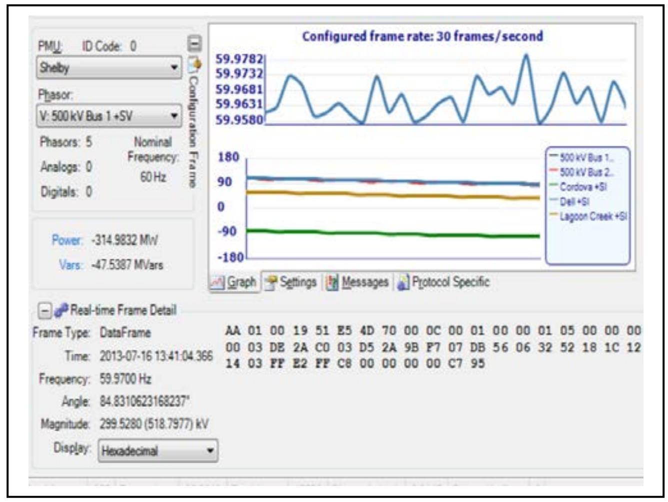

Figure 8. Sample phasor measurement unit output, Open PMU

Leveraging the PMU as a primary data acquisition platform enables the operation of power systems in a Big Data Paradigm. Measurements, such as the example pictured above in Figure 8, translate electric power signals into sinusoidal waveform and are capable of recording as many as 120 geo-referenced time-synchronized measurements per second in a streaming fashion, whereas conventional SCADA monitoring systems are event-driven, and typically record a single measurement every 2-4 seconds. In addition to acquiring a larger volume of data, PMUs are capable of measuring a variety of variables, such as voltage and frequency fluctuations, rate of change of frequency, harmonics, and phase angle difference. The veracity of these measurements is ensured with time synchronization via GPS arbiter clocks in each unit. By employing such an increased observational capacity, grid operators can sense the onset of disturbance or fault events with much greater precision. However, this requires both the ability to ingest large continuously streaming datasets [79], as well as the analytic capability to perform complex computations on the incoming data in order to discern between typical system fluctuations and dangerous anomalies. In this regard, standards for PMU performance, such as IEEE C.37.118 continue to evolve as thresholds for harmonic and interharmonic filtering are calibrated to preserve valuable signal components [80].

With regard to data ingestion and the velocity of PMU record generation, communications latency is a nontrivial issue worthy of note. In order to maximize the utility of PMUs, they must be able to communicate through a robust protocol such as TCP/IP, with significant bandwidth (i.e., at least 5 Megabytes per second). In light of the synchronized nature of PMU data, a network of geographically dispersed devices can facilitate a wide-area measurement system (WAMS), provided that each device enjoys uniform connection speed. Whereas conventional fault recording devices are generally triggered only in event-driven circumstances, the need for persistent Internet connectivity between substations and central monitoring facilities is a novel requirement for many utility operators. Similarly, the streaming nature of synchrophasor data requires utility operators to develop data retention strategies or policies in order to effectively manage such a large increase in data generation. Although the dynamic nature of power systems means that PMU data are highly volatile for functions that demand a fast response, the preservation of data over time is critical to yielding deep insights through an A2O process.

As an electric power signal contains numerous variables of interest, the analytic phase of our approach involves numerous procedural layers. A variety of competing algorithmic techniques exist for detecting anomalies, such as analyzing rates of change of voltage, frequency, and phase angle difference, detecting fluctuations in harmonic distortion, and measuring voltage unbalance, many of which have historically been executed independently of one another. As computational capability has advanced, the ability to leverage support vector machines, artificial neural networks, decision trees, and other intelligent classifiers presents the opportunity to more quickly and accurately detect events of interest. In particular, we are exploring how cognitive computing in a layered analytic process has the potential to enable the rapid detection of the onset of electrical islanding scenarios. By developing an integrated islanding detection method that is both sensitive to target events and stable against false tripping, we aim to improve the integration of renewable energy sources such as photovoltaics and wind turbines. In turn, we intend that such an advanced analytic process will also help to identify and visualize patterns related to other destabilizing events in electric power systems. We have chosen to focus our research on archipelagos due to the unique set of circumstances which these environments represent. Given their physical isolation, islands are inherently bounded problem sets. In addition, the power systems that serve islands are characterized by lower inertia and lower blackstart/quickstart ratios than larger systems such as the North American bulk power grid.

In exploring these problem sets, we employ a combination of algorithmic and heuristic analytics. Although the PMU's increase in observational granularity coupled with advanced computational analytic capabilities represent the value of algorithmic insight for enhancing power system resilience, the generation of heuristic insight will also be a significant pathway towards a more resilient and adaptive operating paradigm. In addition to its internal system dynamics, many external natural and anthropogenic environmental factors each exert influence on the operation of the grid. Therefore, including data on seismic events, weather patterns, and other natural phenomena as layers in the analytic stack will help to enrich the observational space in which we analyze power system dynamics. In addition, human-generated content is a valuable source of situational awareness regarding service disruptions, power outages, and latent vulnerabilities. By way of example, power companies in the U.S. have historically become aware of the majority of outages only after their customers called in to report that they were without power [81]. However, vigilant customers can serve as sensors not only to detect instances of power outage, but also to identify conditions that may be pre-cursors to fault or disturbance events, such as vegetation overgrowth affecting power lines, abnormal spark emission from transformers, and suspicious activity around substations. Similar to the U.S. Department of Homeland Security's "See Something, Say Something" public ad campaign, the encouragement of customer vigilance and crowdsourced infrastructure protection can be a useful tool for augmenting the limited capacity of system operators and public safety officials charged with safeguarding the grid. As depicted below in Figure 9, our approach aims to integrate these various data sources and analytic capabilities in order to achieve power systems that operate with greater resilience.

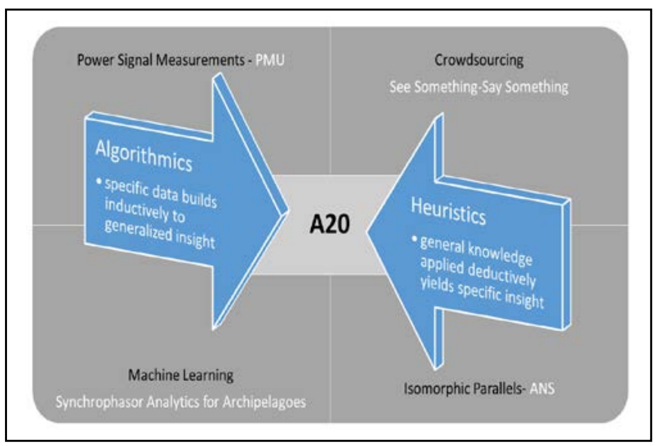

Figure 9. Analytics on Analytics (A2O) Concept

As technology and connectivity advance us closer to the realization of a smart grid, a coherent and logical integration of data acquisition, analytics, and visualization will be critical. The ability to assimilate system state data from PMUs and other devices deployed across the grid with a variety of environmental data and geo-referenced content from human observers and remote sensors will be vital for operating power systems that are both reliable and sustainable. In addition, gaining a better understanding of how complex dynamics impact the grid's operation will directly inform how automated protocols can be established to develop computer-assisted dynamic fine tuning measures.

In this regard, nature-inspired engineering and insight from the life sciences may be instructive. In particular, the autonomic nervous system (ANS) is one potentially useful source of biomimetic and isomorphic insight for how a self-healing smart grid could operate. Also known as the involuntary nervous system, the ANS maintains conditions for a steady state or homeostasis within the body and plays a role in many of its critical processes, including the defense reaction or "fight or flight" response, thermoregulation, coping with metabolic challenges [82], and regulation of organ functions related to the biological clock [83]. The ANS is composed of two primary subsystems that influence bodily functions; the sympathetic nervous system, which activates function, and the parasympathetic nervous system, which limits function [84].

An important aspect of the ANS that translates well to the electric grid is its feedback-regulated nature; as conditions within the body change, the ANS acts to mitigate potentially harmful effects. For example, when we stand up after sitting for an extended period, the potentially dangerous drop in blood pressure that could result from blood pooling in the legs is counteracted as the ANS initiates a series of baroreceptor reflexes that increase heart rate and blood flow while mediating constriction of the blood vessels to restore steady blood pressure throughout the body. However, the ANS is also capable of anticipatory responses that serve to manage the danger presented by system perturbations or destabilizing events before they occur. The anticipatory release of insulin prior to a meal is one simple example. And yet, maintaining a homeostatic steady state may not always be the best strategy for ensuring the long-term survival of a system. In response to stress imposed by extreme circumstances, ANS processes within the body may need to shift to an altered state or allostasis[1] characterized by adaptive levels of system output like increased heart rate and blood flow through the muscles in order to facilitate escape from or confrontation with a physical threat [85]. Similarly, cyber-physical infrastructures such as the grid must also operate during periods of unusually high demand, or acute instability such as in the midst or aftermath of natural disaster.

[1] Allostasis refers to a system's ability to maintain functionality through change. In contrast to a homeostatic state in which a system maintains a relatively stable balance, an allostatic state is characterized by temporarily unbalanced ratios in system input vs. output, or other adaptations to internal processes that enable the system to remain in operation during external perturbations, after which the system returns to homeostasis.

In light of its role in both feedback-regulated and anticipatory responses that facilitate homeostatic and allostatic system states, the ANS offers conceptual parallels for enhancing the resilience of critical infrastructure systems like the electric grid. Just as the ANS acts within the body to mitigate against potentially harmful conditions without an individual's conscious awareness, machine automation or computer assisted dynamic fine tuning can act within the grid to mitigate against potentially harmful conditions before they are even recognized by human system operators. Yet, just as the ANS relies on input from the five senses to drive its operation, human-engineered systems like the grid require robust, precise, and diverse data in order to operate effectively. Technological advances such as synchronized phasor measurements, cognitive computing, and machine learning capabilities are particularly beneficial for operating large infrastructural systems such as the electric grid, and are therefore the primary components of our synchrophasor analytics system for archipelagos.

VII. Conclusion

As machine capability continues to accelerate giving rise to big and bigger data, the power and promise of robust analytics will grow along with potential areas of vulnerability introduced by increased system connectivity and interdependence. At the same time, our ability to make sense out of evolving circumstances quickly, and adapt social and physical structures accordingly will be important determinants in the shape of things to come. Such competing dynamics call for a suitably balanced approach to data management and systems operation. While complex cyber-physical infrastructures such as the electric grid are akin to the Abelian Sandpile in its susceptibility to cascading effects, tools such as PMU-enabled response mechanisms are the bamboo that can enable such systems to remain resilient amidst destabilizing events. In the context of electric grids in particular, the increasing penetration of distributed sources of renewable energy necessitates an improvement in the ability to detect and counteract the destabilizing effects of unintentional electrical islanding and similar phenomenon. In response, our synchrophasor analytics system for archipelagos aims to achieve a precise islanding detection capability by integrating the granularity of PMU data acquisition with the robust analytic capabilities of cognitive computing and machine learning.

Our participation in the D4D Challenges as well as our investigation of infrastructural resilience demonstrate that much opportunity exists to improve our understanding of how systems operate through the application of analytics in a Big Data Paradigm. We believe that open and inclusive approaches such as the **Sensemaking Methodology** have the potential to enhance numerous dimensions of resilience, including those of cyber-physical systems, societies, and individuals. Systematic Decision Engineering requires a robust and iterative process of data collection, layered analytics, and insight visualization that in turn has the ability to identify critical blind spots and mitigate harmful vulnerabilities. We hope that such a methodology can facilitate positive developments, such as the smart integration of green technologies into sustainable Blue Economies [86], and improvement in our roles as both environmental stewards and engines of social progress. Each of these areas represent exciting and relatively unexplored realms of research that we have designated as targets for future work. Specifically, we plan to further investigate how technological capabilities such as remote sensing and cognitive computing can be effectively integrated with human Sensemaking techniques to achieve increasingly useful insights and practical Decision Engineering solutions.

Acknowledgment

The authors would like to thank our mentors and collaborating partners, as well as the advisory board of the Sensemaking Fellowship.

References

[1] R. Spousta, S. van den Elzen, S. Chan, and J. K. Buenen, "From Cheese to Fondue: A Sensemaking Methodology for Data Acquisition, Analytics, and Visualization," in *IARIA 2015 Data Analytics* Nice, France, 2015, pp. 38-43.

[2] N. R. Council, *Frontiers in Massive Data Analysis*. Washington, DC: The National Academies Press, 2013.

[3] R. Spousta and S. Chan, "Milk or Wine: Are Critical Infrastructure Protection Architectures Improving with Age? ," *Journal of Challenges,* vol. 2, pp. 43-57, 2015.

[4] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: An explanation of the 1/f noise," *Physical Review Letters,* vol. 59, p. 381, 1987.

[5] W. Bryant, "Cyberspace Resiliency: Springing Back with the Bamboo," in *Evolution of Cyber Technologies and Operations to 2035*, M. Blowers, Ed., ed Cham: Springer International Publishing, 2015, pp. 1-17.

[6] M. E. Porter and V. E. Millar, "How information gives you competitive advantage," *Harvard Business Review,* vol. 63, July-August 1985.

[7] E. Fritsch and B. Rondeau, "Gemology: The Developing Science of Gems," *Elements,* vol. 5, pp. 147-152, 2009.

[8] N. Biehn. (2013, May) The Missing V's in Big Data: Viability and Value. *Wired*. Available: http://www.wired.com/2013/05/the-missing-vs-in-big-data-viability-and-value/. Accessed May 27, 2016

[9] C. Alcaraz and J. Lopez, "Wide-Area Situational Awareness for Critical Infrastructure Protection," *Computer,* vol. 46, pp. 30-37, 2013.

[10] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks,* vol. 54, pp. 2787-2805, 2010.

[11] D. Evans, "The internet of everything: How more relevant and valuable connections will change the world," *Cisco IBSG,* pp. 1-9, 2012.

[12] K. M. Chandy, "Sense and respond systems," in *Int. CMG Conference*, 2005, pp. 59-66.

[13] T. Yu, L. Bai, J. Guo, and Z. Yang, "Construct a Bipartite Signed Network in YouTube," *International Journal of Multimedia Data Engineering and Management (IJMDEM),* vol. 6, pp. 56-77, 2015.

[14] T. Ray, "YouTube's 2 Billion Videos, 197M Hours Make it an 'Immense' Force, Says Bernstein," in *Barron's*, ed, 2016.

[15] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, "Dublin core metadata for resource discovery," *Internet Engineering Task Force RFC,* vol. 2413, p. 132, 1998.

[16] D. ElBoghdady, "Market quavers after fake AP tweet says Obama was hurt in White House explosions," in *Washington Post*, ed, 2013.

[17] T. D. Heidel, J. G. Kassakian, and R. Schmalensee, "Forward Pass: Policy Challenges and Technical Opportunities on the U.S. Electric Grid," *Power and Energy Magazine, IEEE,* vol. 10, pp. 30-37, 2012.

[18] Y. Pradeep, S. A. Khaparde, and R. K. Joshi, "High Level Event Ontology for Multiarea Power System," *Smart Grid, IEEE Transactions on,* vol. 3, pp. 193-202, 2012.

[19] "IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems," *IEEE Std 1547-2003,* pp. 1-28, 2003.

[20] D. M. Legler, H. J. Freeland, R. Lumpkin, G. Ball, M. J. McPhaden, S. North, *et al.*, "The current status of the real-time in situ Global Ocean Observing System for operational oceanography," *Journal of Operational Oceanography,* vol. 8, pp. s189-s200, 2015.

[21] M. Alavi and D. E. Leidner, "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly,* vol. 25, pp. 107-136, 2001.

[22] C. Hajjem, S. Harnad, and Y. Gingras, "Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact," *arXiv preprint cs/0606079,* 2006.

[23] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences,* vol. 275, pp. 314-347, 2014.

[24] S. F. Oliveira, K. Furlinger, and D. Kranzlmuller, "Trends in Computation, Communication and Storage and the Consequences for Data-intensive Science," in *High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS), 2012 IEEE 14th International Conference on*, 2012, pp. 572-579.

[25] D. Ferrucci and A. Lally, "UIMA: an architectural approach to unstructured information processing in the corporate research environment," *Natural Language Engineering,* vol. 10, pp. 327-348, 2004.

[26] R. Kurzweil and M. L. Schneider, *The age of intelligent machines* vol. 579: MIT press Cambridge, 1990.

[27] L. van Zoonen, "Gendering the Internet: Claims, Controversies and Cultures," *European Journal of Communication,* vol. 17, pp. 5-23, March 1, 2002.

[28] D. Gurer, "Pioneering women in computer science," *Communications of the ACM,* vol. 38, pp. 45-54, 2002.

[29] H. H. Goldstine, *The computer from Pascal to von Neumann*: Princeton University Press, 1980.

[30] M. Glassman and M. J. Kang, "Pragmatism, connectionism and the internet: A mind's perfect storm," *Computers in Human Behavior,* vol. 26, pp. 1412-1418, 2010.

[31] C. E. Shannon and W. Weaver, *The mathematical theory of communication*: University of Illinois press, reprinted 2015.

[32] S. McCartney, *ENIAC: The Triumphs and Tragedies of the World's First Computer*: Walker \& Company, 1999.
[33] A. M. Turing, "Turing's treatise on Enigma," *Unpublished Manuscript,* 1939.
[34] S. Lohr, "For big-data scientists, janitor work is key hurdle to insights," *New York Times,* vol. 17, 2014.
[35] M. Viser, "House panel details failures in run-up to Marathon attack," in *Boston Globe*, ed, 2014.
[36] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems,* vol. 90, pp. 111-127, 1997.
[37] G. S. Sureshrao and H. P. Ambulgekar, "MapReduce-based warehouse systems: A survey," in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on*, 2014, pp. 1-8.
[38] N. Ford, "Information retrieval and creativity: towards support for the original thinker," *Journal of Documentation,* vol. 55, pp. 528-542, 1999.
[39] J. Stevenson, "Jihad and Piracy in Somalia," *Survival,* vol. 52, pp. 27-38, 2010.
[40] A. Shortland, "Treasure mapped: using satellite imagery to track the developmental effects of Somali Piracy," *London: Chatham House*, 2012.
[41] S. Yikona, *Pirate Trails: Tracking the Illicit Financial Flows from Pirate Activities Off the Horn of Africa*: World Bank Publications, 2013.
[42] B. E. Byland and J. M. Pohl, *In the realm of 8 Deer: The archaeology of the Mixtec codices*: University of Oklahoma Press, 1994.
[43] J. Z. Gao, L. Prakash, and R. Jagatesan, "Understanding 2D-BarCode Technology and Applications in M-Commerce - Design and Implementation of A 2D Barcode Processing Solution," in *Computer Software and Applications Conference, 2007*, pp. 49-56.
[44] W. Neil Adger, N. W. Arnell, and E. L. Tompkins, "Successful adaptation to climate change across scales," *Global Environmental Change,* vol. 15, pp. 77-86, 2005.
[45] E. Bumiller. (2010, April 26) We Have Met the Enemy and He is Powerpoint. *New York Times*. Available: http://www.nytimes.com/2010/04/27/world/27powerpoint.html?_r=2. Accessed May 27, 2016.
[46] S. H. Strogatz, "Exploring complex networks," *Nature,* vol. 410, pp. 268-276, 2001.
[47] S. G. Eick, J. L. Steffen, and E. E. Sumner, Jr., "Seesoft-a tool for visualizing line oriented software statistics," *Software Engineering, IEEE Transactions on,* vol. 18, pp. 957-968, 1992.
[48] A. Ozment and S. E. Schechter, "Milk or wine: does software security improve with age?," in *Proceedings of the 15th conference on USENIX Security Symposium-Volume 15*, 2006, p. 7.
[49] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. M. T. Do, *et al.*, "From big smartphone data to worldwide research: The Mobile Data Challenge," *Pervasive and Mobile Computing,* vol. 9, pp. 752-771, 2013.
[50] V. D. Blondel, M. Esch, C. Chan, F. Clérot, P. Deville, E. Huens, *et al.*, "Data for development: the d4d challenge on mobile phone data," *arXiv preprint arXiv:1210.0137*, 2012.
[51] J. Poole. (2013, May 6) Winning Research from the Data 4 Development Challenge. *United Nations Global Pulse*. Available: http://www.unglobalpulse.org/D4D-Winning-Research. Accessed May 27, 2016.
[52] N. Elhefnawy, "Societal Complexity and Diminishing Returns in Security," *International Security,* vol. 29, pp. 152-174, 2004.
[53] V. Hughes, "Stress: The roots of resilience," *Nature,* vol. 490, pp. 165-167, October 10, 2012 2012.
[54] S. S. Luthar and L. B. Zelazo, "Research on resilience: An integrative review," in *Resilience and vulnerability: Adaptation in the context of childhood adversities*, ed New York, NY, US: Cambridge University Press, 2003, pp. 510-549.
[55] C. S. Holling, "Resilience and Stability of Ecological Systems," *Annual Review of Ecology and Systematics,* vol. 4, pp. 1-23, 1973.
[56] L. Briguglio, G. Cordina, N. Farrugia, and S. Vella, "Economic Vulnerability and Resilience: Concepts and Measurements," *Oxford Development Studies,* vol. 37, pp. 229-247, 2009.
[57] S. E. Chang and M. Shinozuka, "Measuring Improvements in the Disaster Resilience of Communities," *Earthquake Spectra,* vol. 20, pp. 739-755, 2004.
[58] Y. Y. Haimes, "On the Definition of Resilience in Systems," *Risk Analysis,* vol. 29, pp. 498-501, 2009.
[59] B. Walker, "Resilience, Adaptability and Transformability in Social--ecological Systems," *Ecology & Society (formerly Conservation Ecology),* vol. 9, p. 5, 2004.
[60] K. Tierney and M. Bruneau, "Conceptualizing and measuring resilience: A key to disaster loss reduction," *TR news*, 2007.
[61] T. McDaniels, S. Chang, D. Cole, J. Mikawoz, and H. Longstaff, "Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation," *Global Environmental Change,* vol. 18, pp. 310-318, 2008.
[62] C. S. Holling, *Engineering Resilience versus Ecological Resilience*: The National Academies Press, 1996.
[63] B. E. Beisner, D. T. Haydon, and K. Cuddington, "Alternative stable states in ecology," *Frontiers in Ecology and the Environment,* vol. 1, pp. 376-382, 2003/09/01 2003.
[64] C. Folke, "Resilience: The emergence of a perspective for social–ecological systems analyses," *Global Environmental Change,* vol. 16, pp. 253-267, 2006.
[65] N. N. Taleb, *The black swan: The impact of the highly improbable fragility*: Random House, 2010.
[66] T. D. O'Rourke, "Critical infrastructure, interdependencies, and resilience," *Bridge - Washington National Academy of Engineering,* vol. 37, p. 22, 2007.
[67] A. Boin and A. McConnell, "Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience," *Journal of Contingencies and Crisis Management,* vol. 15, pp. 50-59, 2007.
[68] J. O. Petinrin and M. Shaaban, "Smart power grid: Technologies and applications," in *Power and Energy (PECon), 2012 IEEE International Conference on*, 2012, pp. 892-897.
[69] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review,* vol. 45, pp. 167-256, 2003.
[70] J. Ash, "Optimizing complex networks for resilience against cascading failure," *Physica A,* vol. 380, p. 673, 2007.
[71] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, "Network Structures from Selection Principles," *Physical Review Letters,* vol. 92, p. 198701, 2004.

[72] M. Brede, "Networks that optimize a trade-off between efficiency and dynamical resilience," *Physics letters. A*, vol. 373, p. 3910, 2009.

[73] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the Barabási–Albert network model," *Physica A: Statistical Mechanics and its Applications*, vol. 355, pp. 667-677, 2005.

[74] E. Zio and G. Sansavini, "Vulnerability of Smart Grids With Variable Generation and Consumption: A System of Systems Perspective," *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, vol. 43, pp. 477-487, 2013.

[75] H. Yih-Fang, S. Werner, H. Jing, N. Kashyap, and V. Gupta, "State Estimation in Electric Power Grids: Meeting New Challenges Presented by the Requirements of the Future Grid," *Signal Processing Magazine, IEEE*, vol. 29, pp. 33-43, 2012.

[76] J.-M. Lehn, "Toward Self-Organization and Complex Matter," *Science*, vol. 295, pp. 2400-2403, March 29, 2002 2002.

[77] S. H. Haeckel, "Adaptive enterprise design: The sense ‐ and ‐ respond model," *Planning review*, vol. 23, p. 6, 1995.

[78] A. G. Phadke, "Synchronized phasor measurements-a historical overview," in *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, 2002, pp. 476-479 vol.1.

[79] "IEEE Standard for Synchrophasor Measurements for Power Systems," *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1-61, 2011.

[80] K. E. Martin, "Synchrophasor Measurements Under the IEEE Standard C37.118.1-2011 With Amendment C37.118.1a," *Power Delivery, IEEE Transactions on*, vol. 30, pp. 1514-1522, 2015.

[81] D. W. Caves, J. A. Herriges, and R. J. Windle, "Customer Demand for Service Reliability in the Electric Power Industry: A Synthesis of the Outage Cost Literature," *Bulletin of Economic Research*, vol. 42, pp. 79-121, 1990.

[82] C. B. Saper, "The Central Autonomic Nervous System: Conscious Visceral Perception and Autonomic Pattern Generation," *Annual Review of Neuroscience*, vol. 25, pp. 433-469, 2002.

[83] R. Buijs, C. van Eden, V. Goncharuk, and A. Kalsbeek, "The biological clock tunes the organs of the body: timing by hormones and the autonomic nervous system," *Journal of Endocrinology*, vol. 177, pp. 17-26, April 1, 2003.

[84] G. G. Berntson, M. Sarter, and J. T. Cacioppo, "Autonomic nervous system," *Encyclopedia of cognitive science*, 2003.

[85] B. S. McEwen and J. C. Wingfield, "The concept of allostasis in biology and biomedicine," *Hormones and Behavior*, vol. 43, pp. 2-15, 2003.

[86] G. Pauli, "The blue economy," *Our planet*, pp. 24-27, 2010.