The global thrust for macro-climate discussions (UN IPCC, etc.) are increasingly at the center of our collective social conscience. The granularity of the distributed impact is less known, perhaps due its complexity. How many know about NASA SMAP using 102GHz radar for soil moisture mapping?

We are approaching an inflection point where it may become feasible to use nano-satellites, on a pay-per-use basis, to download daily or hourly precision micro-climates (by GPS coordinates, zip codes) via apps on smartphones, tablets or portals, connected to (intelligent?) decision support systems.

In the context of using nano-satellites for micro-climates, acquired data, analytics, information and knowledge, may be of limited value without complementarity of advances in data granularity, on the ground. If these two resources can converge or if combined, the output may be more valuable.

Haphazard thoughts and a clutter of amorphous ideas about "ground" and “complementarity” may be found in this lengthy essay ‘SIGNALS’ (PDF is here https://dspace.mit.edu/handle/1721.1/111021).
Document contains a trio of essays spewing haphazard clutter of amorphous thoughts:

1. SITS
2. SIP-SAR
3. SARS♣AG

"What we know is not much. What we do not know is immense."--Last words of mathematician/scientist Pierre-Simon Laplace • March 23, 1749 – March 5, 1827
Can SITS serve as an exploration to usher a confluence of systems?

Essay by Shoumen Datta • Auto-ID Labs, Dept of Mech Eng, MIT • MDnP Lab, Dept of Anesthesiology, MGH, Harvard Medical School

BACKGROUND

Perhaps from a philosophical or pensive perspective, human health and soil health may share some common grounds. Both are complex examples of convergence shaped by genes, geology and the environment. This essay, triggered by the discussion surrounding “signals in the soil” (SITS), presents a few suggestions about tools to monitor soil health. A few thoughts may be useful in the short term and some may seem pragmatic in a few more years. Other thoughts may be absurd, incorrect or simply wrong. Still others may take shape a few decades from now. For the latter, *is earlier always better*⁴ if one wants to plant a few seeds? Early adopters may fail² but one must *dare, to propose new ideas*³ and admit errors.

ABSTRACT

In the context of a recent request for proposal (NSF 19-556 RFP⁴), signals in the soil (SITS) offers yet another opportunity to re-address, the universal context of how to benefit from data. This essay is not about soil health *per se* and is not focused on soil research⁵ data⁶.

Provided data is sufficiently valid, relatively noise-free and originated from a reliably calibrated instrument, data may be *agnostic of the source*, instrument or sensor. The acquisition of the data by the source, instrument or sensor, is *agnostic of the medium* of the raw data (waves, voltage, chemicals) as long as the signal transduction is reproducible, does not violate laws of physics and its authenticity is uncontaminated by signal processing artefacts⁷ (introduced due to digital signal processing, analog to digital converters, digital to analog conversion).
Content, validity and use of the data is agnostic of the subject of the data (for example, soil, humans, machine) as long as the data is untarnished and the subject of the data has value for our purpose or question (for example, is soil still valuable? Are humans, machines, still valuable?).

Stripped of its source, medium and subject, data is a vehicle to answer questions. To move forward, imagine data represented as a vehicle (cartoon of an automobile). Science, engineering and technology sits under the bonnet. Soil, humans and machines are in the boot. Questions hop on the passenger seat, when they reach their destination (when the question is answered), they hop off. The automobile is a convertible, an open platform. Time to time, elements under the bonnet and the boot may be serviced, replaced and upgraded. Semantics and ontology are crowdsourced fuels. The (abstract) journey continues irrespective of the incongruities expected on the unpaved road, and unknown challenges that lie ahead.

For SITS, we may transform the abstract journey into reality, by viewing the infrastructure of the proposal as a scaffold for convergence of systems science and data science with soil science. Other groups may wish to add to this systems approach and install infrastructure for, for example, food science, water science, health science. As a consequence, the systems science module grows to accommodate the dynamics, connections and networks relevant to the systems (food, energy, water, sanitation, health). Data, rising to the top, from these different domains, may begin to reveal deeper patterns beyond obvious correlations (decreasing levels of blood calcium and increasing instances of osteoporosis). Data science, in this trifecta, delivers the obvious, yet poised to address questions that were not asked. With advanced tools, we can now pursue non-obvious relationship analytics.

Non-obvious relationship analysis (NORA) may move beyond the awareness realm to empower a new breed of thinkers to ask the correct questions, and tinker with unknown unknowns, in quest of answers or breakthroughs. The shrink-wrapped version of the existing conventional wisdom is void of vision, mundanely obvious and offers tired recipes for incrementalism. The latter may induce one to wonder whether the electric light bulb could have evolved from incremental improvement of candles?
In the SITS context, the proposal may focus on the trinity of soil science, systems science and data science. Systems science refers to the vast ecosystem accompanying the domain of interest. It must be accounted for in terms of interrelationships relevant to sub-domains [for example, soil science may include microbes, chemicals, water, weather (local, global), geochem/geobiology, vegetation, etc]. However, a very tiny slice of the system may be the focus in any proposal, for research and funding.
OUTCOME

The amorphous idea of NORA is the implicit optimism latent in this triumvirate mode of thought (Figure 0). Non-obvious relationship analysis is strenuously non-linear, difficult to predict, may be uncertain in its meaning or output, if viewed with canonical perspectives (for example, if the imagination is out of focus). It may provide clues that are undetectable or remain undiagnosed or unrecognized. The latter is the greatest loss of value. NORA-like approaches may hold untold potential for exploring questions that we did not even know how to frame or ask.

Non-obvious clues are beginning to emerge from diverse domains, for example, biomimicry and cross-kingdom tools in plants and humans perhaps structured as evolutionary immune response strategies against viral invaders. However, one must hasten to add that phages and viruses may also provide solutions to our problems. There may be “million times more virus particles than there are stars in the observable Universe” but less than 5,000 virus species of the virome are currently documented.

With respect to soil science, whether the outcome of this miniscule slice of the system (SITS) will generate any results, remains to be seen. Developing tools to answer obvious questions generates data. Analytics may reveal cryptic relationships, which may hold clues for molecular interactions (plants, microbiomes, viromes) influenced by macro- and micro-environments (soil, water, pH, gases, air, temperature, chemicals, irradiance, insolation). The central question concerns the ambiguity whether observers can ask the correct questions. It is even more dubious whether we may possess the incisive foresight to decipher and/or recognize that new data and/or information has emerged.

At this time, we cannot answer these questions. Hence, we will focus on outcomes driven by tools to acquire data, to catalog attributes and characteristics of soil. As mundane and obvious as it may be, the path to the outcome may have to travel through unchartered territories. From the point of view of scientific research proposal on pedology, perhaps this knowledge is of value. We hope, when combined with a thoughtful data management plan, this approach may illuminate the meaning of this data, at the least, with relevance to SITS.
Table 1 – Lists data acquisition with respect to each entity/characteristic (upper). Mapping physico-chemical attributes in the context (environment) of the microbes detected (lower).

<table>
<thead>
<tr>
<th>Microbe 1</th>
<th>Microbe 2</th>
<th>Microbe 3</th>
<th>Microbe 4</th>
<th>Microbe 5</th>
<th>Microbe 6</th>
<th>Microbe 7</th>
<th>Microbe 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>moisture</td>
<td>salinity</td>
<td>metals</td>
<td>ions</td>
<td>density</td>
<td>particles</td>
<td>color</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Microbe 1</th>
<th>Microbe 2</th>
<th>Microbe 3</th>
<th>Microbe 4</th>
<th>Microbe 5</th>
<th>Microbe 6</th>
<th>Microbe 7</th>
<th>Microbe 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>Moisture</td>
<td>Salinity</td>
<td>Metals</td>
<td>Ions</td>
<td>Density</td>
<td>Particles</td>
<td>Color</td>
</tr>
</tbody>
</table>

Figure 1 – Sensor engineering\(^1\) combines thin film with flex circuit manufacturing using liquid crystal polymer (LCP) and other thermoplastic polymers. LCP substrate with embedded sensors, communication coil (left top) after folding over frame (left bottom).

7 ◆ Triumvirate Modes of Thought: SITS - an opportunity for imagination, invention, innovation & implementation
SITS may reveal its value in terms of the creative convergence of science and engineering to create tools and technology, to detect signals in the soil, faster and cheaper.

The collaboration with physicists and engineers for breakthroughs in remote sensing may be a pivotal part of SITS in pushing the envelope. It is encouraging to note that reflected (RF) radiowaves can be used for remote applications in humans. Remote sensing in the context of SITS include [a] concepts for moisture detection by satellites and data from 102GHz frequency, as well as [b] use of 2.45GHz (WiFi) to estimate soil moisture using drones rather than WSN to acquire data from immobile tethered sensors, which pollutes the soil, yet, often promoted by laissez-faire titles and published by august/elite journals.

Moving from gigahertz (GHz) to terahertz (THz) improves the spatial resolution of the signal (difficult to obtain with WiFi). The interest in THz have waxed and waned over the past half century. THz may be useful as a frequency cluster for detection of radio signals from proteins (protein electro-dynamics views proteins as radios, with distinct signatures, but too “noisy” to decipher the data, for example, differentiate between wild-type and mutant protein signatures, which could be of immense value in biomedical diagnostics).

Therefore, advances from the convergence of terahertz integrated electronic and hybrid electronic-photonic systems, may lend itself to new applications in sensing. Sengupta et al mentions waveguides emulating artificial dielectric medium, which, for SITS may translate to better soil moisture sensor using THz. The role of physics and engineering experts is to [a] help us understand the continuing interest in (artificial) dielectrics and [b] guide SITS in combining that knowledge with improved spatial resolution (THz or sub-millimeter wavelengths), for breakthroughs in remote sensing and future instrumentation.

Dielectrics seem to appear from several dimensions and for many uses. It may hold clues for remote sensing tools to measure pH, salts and enzymes in the soil, as previously suggested. Static dielectric constant of water drops with addition of salts, but the reverse is true, if the complexity of permittivity of water is explored with THz transmission spectroscopy (the value increases, Debye relaxation, with salt concentration).
For SITS, is it possible to exploit any correlation\textsuperscript{38} between dielectric permittivity, ionic charge, charge density (phosphates) and chemical structure? Dielectric permittivity\textsuperscript{39} is the primary diagnostic physical property\textsuperscript{40} for GPR\textsuperscript{41} (ground penetrating radar)? Creative use of ISAR\textsuperscript{42} (inverse synthetic aperture radar) in the development of WiVi (Katabi et al) emerged from radar technologies. Dielectric permittivity is the primary diagnostic physical property in ground penetrating radar. Tools from radar technology are quite pragmatic.

The use of dielectric medium as a “marker” to extrapolate data relevant to SITS (pH, salts) may gain momentum from prior research which indicates that it is possible to model\textsuperscript{43} dielectric medium with electric and magnetic reservoirs. The latter unleashes electromagnetic field quantization, which may serve as another principle for remote sensing. We need to delve deeper into dielectrics\textsuperscript{44} as well as investigate other ion-based detection systems (GC/MS\textsuperscript{45}).

The importance of “markers” cannot be overemphasized. Glucose oxidase was the first enzymatic marker\textsuperscript{46} (coupled to an amperometric electrode for monitoring oxygen in blood) followed by various forms of spectroscopy. Raman spectroscopy\textsuperscript{47} and bio-impedance spectroscopy\textsuperscript{48} are emerging as principles of choice for non-invasive remote applications.

Impedance spectrum, or dielectric spectrum, is measured in the range 0.1-100MHz. To measure variations in plasma glucose concentrations\textsuperscript{49} (critical for diabetes patients), the primary “marker” is the detection of the change in red blood cells due to the variation of plasma glucose concentration.

The variation changes the membrane potential of red blood cells (RBC) due to decrease in [Na\textsuperscript{+}] and increase in [K\textsuperscript{+}] concentrations. The changes in the membrane potential is estimated by determining the permittivity and conductivity of the cell membrane through the dielectric spectrum. This data is used to extrapolate plasma glucose concentration.

SITS is in quest of such markers. The key is to extract the rigor of basic science, and use the principles, in applications, to identify “markers” to serve as quantitative standards, or indices, for remote sensing with respect to the characteristics of soil, we wish to measure, for SITS, perhaps using THz sensing.
SITS, therefore, is seeking ground-breaking scientific principles and affordable engineering tools, which will further enable mobile data collection technologies. Half a century ago, the concept of projection reconstruction changed the field of medical imaging. By analogy, it is tempting to speculate that nanoscale vector magnetometry from the domain of magnetic field sensing, may be one such tool, to detect signals in the soil (SITS) using vector sensing of static fields to reveal properties of soil (moisture, pH, chemicals, etc.). The authors claim far-reaching consequences for nanoscale NV-NMR (nitrogen-vacancy nuclear magnetic resonance) to map spin arrangements of single proteins with increasing spatial resolution. The ability to control nanoscale quantum sensors may unleash new paradigms in precision metrology (measuring atomic-scale magnetic fields with great precision, not only up and down, but sideways, as well). One branch of this development, hopefully, may create new low-cost mobile tools to decipher signals from the soil. Applying super-resolution quantum spectroscopy for agriculture may seem obtuse, now, in the same manner that applying NMR/MRI (nuclear magnetic resonance/magnetic resonance imaging) for monitoring growth of roots may have appeared, once upon a time, a far-fetched idea.

Figure 2: Top - The “bucket brigade” for data collection. Sensors in the soil vs modern wearables. Bottom - Glucose sensor. (A) Graphene printed on bio-resorbable silk with wireless coil. (B) Bio-transfer of sensor on tooth. (C) Non-invasive wireless glucose sensing (D) Self-assembly of pathogenic bacteria bound by peptides on nano-transducer surface. SITS seeks nanoscale quantum sensors and super-resolution quantum spectroscopy tools.
The next open question in our scientific exploration of SITS may address whether we can use bio-inspired principles from dolphins and bats, in addition to electromagnetic radiation (radiowaves, radar) and quantum tools, to help us create tools for remote sensing.

Reflected sound waves coupled with timing devices are echo-signal transduction medium for sonar and ultrasound tools. More than a quarter century ago, ultrasound chips were fabricated. Capacitive micromachined ultrasonic transducers (CMUT) recently reached the summit of the hype curve with respect to medical imaging. However, in this instance (for SITS) air-borne applications may be more relevant. Use of a multi-frequency CMUT device for ultrasound capnography suggests that soil gas analysis using ultrasound may be possible. Advances in precision ultrasound chips treats gas as an interference, to be damped. What if we can reverse the experiment and measure the gas (multi-frequency ultrasound device for multiple gases to be measured) in the soil? Can we use gas as markers? Gas-as-a-marker may be suitable for what type of soil characteristics? In the final category of waves, lidar’s laser pulses may be useful for measurements of density and particles in, the soil. Lidar tools on a chip may be the future of pressure sensors.

I digress to point out that advances in (any) systems on a chip (radar, sonar, lidar) are creative tools which suffer from a common chronic problem - noise. The ubiquity of “noise” infects and introduces errors in almost all systems (waves, sensors, data). Hence, error detection, error correction codes, and error mitigation tools, are in great demand and hold immense significance to assure performance levels below the fault-tolerance threshold, because of noise associated with signals. Key milestones in error correction are due to Shannon (1948), Kalman (1960), Granger (1969), Engle (1982) and Granger and Engle (1987). A recent paper (2019) may be interesting, too. Noise in any communication introduces transmission errors. Source coding may remove redundancy from source data. Channel coding may make noisy channels appear "noiseless" by controlled addition of redundancy. One (of many) unsolved problem in error correction may concern protein dynamics and implied by Martin Karplus and suggested by others. It appears that proteins are radios. Could we use THz waves to explore protein signatures? The signal will be contaminated by noise from vibration of water molecules (think van der Waals radii).
Soil microbiology has deep roots. Circa 1904 witnessed the germination of the concept of rhizosphere\(^{76}\) and soil flora\(^{77}\) has been studied intensely since the turn of the last century. Our task is simpler in the context of SITS. We are not focused on biological function of phyto-microbiomes\(^{78}\) or viromes\(^{79}\) associated with soil, crop, or plants. We wish to detect individual microbes in soil samples (reflect on the content of Figure 1 and Figure 11).

We are searching for *bio-available* extra-cellular molecules on bacterial cell walls and/or membranes which we can use as molecular targets (analytes). We prefer to select molecular targets with sufficient degree of *species specificity* and limit the identification to species. We may exclude subspecies, biovars\(^{80}\) and serovars, as well as analysis involving genes, genomics\(^{81}\) and metagenomics. The latter takes weeks to months to obtain data. Our detection time range aspires to remain in the seconds to minutes, or hours, if necessary.

Therefore, rather than choosing a microbiome (set of bacteria, fungi or algae specific for a habitat, for example, citrus grove), we are in quest of *accessible microbial molecular targets* (Figure 3) which can serve (bind), as an analyte, with a complementary molecule on a sensor. The binding may or may not be reversible, but the binding must elicit a detectable signal (impedance spectroscopy, plasmon resonance, photonics, electrochemical, SERS). Signal transduction, analog to digital conversion, data analysis and data visualization\(^{82}\) in seconds to minutes, is the expectation. Access to data via open platforms and visualization via any mobile platform\(^{83}\) (for example, using an app on a smartphone, tablet, or laptop) is key to mass consumption of data, to reduce transaction cost\(^{84}\) based on economies of scale.

Identifying available molecular targets for microbes in the soil will enable us, initially, to select a few microbes, to be detected with new tools (see Table 1 & Figure 1). For growers, the most abundant forms of soil bacteria\(^{85}\) may be less useful compared to the microbiome specific for their plant or crop. The context of specific microbiomes/viromes are key to understanding potential molecular and cellular interactions. But, we cannot reach such goals without these tools of detection and the advances in science and engineering fundamentals.
Detection using plant lectin-based bio-sensors are one of the widely used sensing tools (in addition to other uses and biological role of lectins). Recent advances using fluorescent emission spectra are improving the quality and intensity of the signal from lectin agglutination assays. Our goal is to use a variety of tools, combined, as appropriate, to identify microbes using sensors in the soil. From the published literature, we hope to determine specific accessible molecular targets, for example, in, *Rhizobium leguminosarum*, *Bacillus subtilis*, *Pseudomonas fluorescens*, *Nitrosomonas europaea*, and Cyanobacteria (Synechococcus sp and if there are soil equivalent of Prochlorococcus). We will select other microbes if we can identify specific extracellular molecules (Figure 3). An example of the use of molecular targets for detection is illustrated in Figure 4.

Figure 3 – Rhizobial cell surface cartoon, illustrated with polysaccharides involved in rhizobial attachment to roots. The “accessible” nature of EPS and KPS suggests its potential as molecular targets for binding to sensors. OM-outer membrane; PS-periplasmic space; PG-peptidoglycan layer; PM-cytoplasmic membrane; EPS-exopolysaccharide; CG-cyclic glucan; PL-phospholipid; MP-membrane protein; KPS-capsular polysaccharide (K-antigens); LPS-lipopolysaccharide.
Figure 4 - Attachment of *Rhizobium leguminosarum* to lectin released from pea and vetch root hairs is pH dependent\(^98\). (Top) Alkaline: Lectin released from root-hair tips. Rhizobial Rhicadhesin mediating attachment. (Bottom) Acidic: Lectin remain anchored. Rhicadhesin released from bacterial surface. Plant lectin and bacterial glucomannan involved in bacterial attachment to plant root hairs.
Lectins, also referred to as phytohemagglutinins, are proteins with at least one non-catalytic domain that can reversibly bind specific carbohydrate moieties. Figure 4 illustrates research results indicating that lectin from roots of the pea (Pisum sativum) and vetch (Vicia sativa) plants interact with a specific glucomannan polysaccharide from strain RBL5523 of Rhizobium leguminosarum.

The interest of SITS is in identifying, for example, Rhizobium leguminosarum using a sensor. There are probably several ways to accomplish this task but a published research paper seems to be a good example, except that we must perform the task in reverse. In the paper, carbohydrate moieties were functionalized noncovalently on the surface of CCG-FET and SWNT-FET devices (field-effect transistor devices comprised of chemically converted graphene or single-walled carbon nanotubes). These devices were then tested for nano-electronic detection of lectins from Canavalia ensiformis (concanavalin A from beans) and Pseudomonas aeruginosa (12kDa and 13kDa, two separate proteins).

For SITS, the nano-bio sensor surface will display the lectin from pea or vetch plant (attached via spacer, linker, innovative 3D printing). If the soil sample contains Rhizobium leguminosarum, the attachment of the bacteria to the lectin on the sensor surface will trigger a (net positive) signal. Hence, in principle, the binding of an analyte (in this case, microbes) to the immobilized substrate with a complementary molecule (may be a lectin), is the modus operandi, which is quite common, in general.

From a pragmatic perspective, “dipping a sensor” in the soil to estimate any microbe may suffer from a variety of problems, the foremost of which is “wetness” or the ability of the signal to attach to the sensor in a preferred medium. How the “medium” may affect the signal strength is a question for experts in energetics and thermodynamics. The involvement of thermodynamics in the energetics of evolution and natural selection may be unexplored but the role of thermodynamics in chemical reactions and molecular interactions are central to future discoveries.
Amplification of the signal strength, mentioned above, may probe the potential to create/introduce a cascade effect to augment the signal when microbes bind the sensor. In this context the role of nanozymes\textsuperscript{104} in sensing\textsuperscript{105} devices\textsuperscript{106} may be an important topic\textsuperscript{107} for SITS. Dissociation of the target molecule from the sensor substrate (to create multi-use sensors) may be a “cleaning” task which may be engineered by co-locating nanozymes as cleaning agents\textsuperscript{108} (perhaps mimic how enzymes are used in the laundry\textsuperscript{109} industry).

We will continue to identify molecules sensitive as a detector for sensor development, suitable for signal transduction, amenable to generate rapid results and provide specificity in identification. The ability to create a reversible reaction (sensor) for continuous monitoring (for example, using pH as a switch or trigger) which can re-calibrated in the field, may be a sign post for SITS 2.0 proposal of the future.

Sensors based on binding of an analyte follow the “lock-key” model\textsuperscript{110} which may be the most widely used analogy\textsuperscript{111} in biology. The correct analogy, the induced fit\textsuperscript{112} hypothesis, may be relatively unknown. The proposal of induced fit was a déjá vu moment if we recall that the Periodic Table\textsuperscript{113} is not\textsuperscript{114} what it used to be\textsuperscript{115} at the turn of the 20th century\textsuperscript{116}, a fact which is rapidly receding from general knowledge, judging from a recent title\textsuperscript{117} of at least one global publication, generally regarded to be adequately informed.

The reason for this preface (digression) is the unexplored sense, that sensors, which bind analytes in an irreversible manner (single use), wishes we re-visit old ideas\textsuperscript{118} to find new room for basic science research. Can we deploy nanozymes to break the lock-key and perform a direct cleaning action? If the direct approach is not feasible, can we engineer indirect use of nanozymes (or cations/anions) to trigger allosteric\textsuperscript{119} transitions? The critical role of allosteric ligands\textsuperscript{120} and the value of allostery\textsuperscript{121} remains under-appreciated. The observation of long-range allosteric effect in hemoglobin\textsuperscript{122} is now relegated only to the hidden pages of history.
DATA SCIENCE – UNDERSTANDING DATA IN THE CONTEXT OF SYSTEMS SCIENCE

*Those who cannot remember the past are condemned to repeat it* is an apt quote applicable to the history of our approach to making sense of data. In 1954, Texas Instruments touted transistors as bringing “electronic ‘brains’ approaching the human brain in scope and reliability” closer to reality. In 2015, IBM’s Modha made the idiotic claim to have “the brain in a box by 2020” and corporate lunacy followed. In 2000, President Bill Clinton declared that the Human Genome Project would lead to a world in which “our children’s children will know the term cancer only as a constellation of stars.” And now with quantum computing, a complex tool with potential but far from panacea. The hoax and hubris from snake oil salesmen resonates on the Minsky scale, “within a generation the problem of creating ‘artificial intelligence’ will substantially be solved.”

A tsunami of data handling tools and software have flooded the market, but few, if any, can make sense of the data and extract contextual information. The claim of the semantic web as a tool for “understanding” was, albeit, temporarily crushed, a decade ago. Since *Nature abhors a vacuum*, hysteria and deadly sins of AI raced to fill the void. Media frenzy and stupidity overshadowed the importance of machine learning (ML) and even experts used the cover of AI to sell books, rather than inculcating the society of mind paradigm. The paradigm has shifted in favour of news cycles as catalysts for polishing the chrome, rather than the ardor for tuning the engine.

From the dawn of computing, the role of context in understanding the meaning of data was not a part of the process, for example, the difference engine and what followed thereafter. Yet there is little need to overemphasize that information arbitrage tools can make a difference between life and death. AI, ML are important in this respect as long as we do not dwell on the terms and admit that these are manufactured labels. The hype of AI/ML is due to smug and glib marketing by folks unaware of the vast field of data driven methods. Labels to build classifiers is referred to as “supervised ML” but for over a century it is called regression. So-called "unsupervised learning" includes clustering, dimensionality reduction, principal component analysis (PCA), support vector machines (SVM), and a long list of statistical modeling techniques that do not require labels (response variables).
Data anomalies\textsuperscript{153} and lack of data interoperability\textsuperscript{154} is the third leading cause of death in the US\textsuperscript{155} (actual number may exceed quarter of a million deaths per year). Deaths due to medical errors is analogous to more than one 747 jumbo jet, with >500 passengers in each plane, crashing every day, with all lives lost in each plane crash, daily. Shocking?

Lack of understanding is not an \textit{irremediable injustice}\textsuperscript{156} perpetrated by the binary system. Proof of concept\textsuperscript{157} and examples\textsuperscript{158} from late 20\textsuperscript{th} century reveals that we may know\textsuperscript{159} how\textsuperscript{160} to make computers \textit{understand}. Fundamental principles of these tools remain in relative obscurity except perhaps one\textsuperscript{161} application which was “dumbed down” to serve as a weather\textsuperscript{162} “app” for select smartphones. The basic problems are not so difficult, at least in principle, as illustrated in Figure 5.

The solutions and so-called standards (Figure 6) are good attempts but was not designed to promote understanding and context, of data and information. The herculean task of understanding and the failed approach from the last century is outlined in Figure 7. The global wave of digital transformation\textsuperscript{163} calls for digital semantics based on ontology schema with URN (see Figure 8, introducing the digital concept of universal resource numbers).

The very limited scope of SITS may not be the platform for the global revolution necessary to usher in digital semantics. But, SITS may move beyond traditional data architectures discussed\textsuperscript{164} elsewhere and attempt to embrace data, context, and connectivity, combined. The latter implies that certain tools from the old semantic web collection shall creep into the SITS data architecture. The hope is that the expert teams involved in this key segment of SITS may find ways to bring creativity and innovation to the task, albeit, handicapped by the dead weight of old technology, to some extent.

Context and understanding are inextricably linked with value, in any instance. Making sense of data is crucial for adoption of any system. It will be a mordant irony if the brilliance of science is useless. Users must be able to ask questions, receive answers in real-time and profit from information.
Figure 5 – Problems that can be solved still fuels lack of interoperability even between intra-company databases. The syntax “callsperday” hard-coded in this software lacks any attribute which can be included in any semantic data dictionary because “callsperday” is void of key linguistic structure (grammar). In another instance, another programmer, in another country, with another type of education, with another natural language proficiency (but not English) may choose to hard-code the syntax “callseachday” or “callsper24hour” or “callseveryday” which are linguistic heresies but semantically equivalent. In the absence of (infrastructure) ontological schema and domain knowledge the semantics of the syntax is lost in translation. Data cannot merge if the syntax differs between databases. Call records in two branch offices may use data loggers sourced from different vendors, which may use different or proprietary programming languages (for example, ABAP used by SAP).
Figure 6 – One answer to the dilemma (above, Figure 5) is the use of a standard. Illustration displays multiple forms of a “standard” (HTML). Each application modified the standard to serve its niche! Why? English, with its vocabulary of a million words, can’t express every thought unambiguously, so with ~100 words we can use in HTML, there are many situations when the standard element may be unsuitable for a piece of the content, in the context of the application (for example, 6 different types of PML are highlighted in red, used by object naming service, ONS, a conceptual cousin of domain name system, DNS). It is not far fetched to imagine that the umbrella of tools and techniques referred to as “AI/ML” can address and even solve some of these problems, in select domains, in certain sub-categories and specific contexts. But, one success or one solution does not imply that “AI/ML” can now be generalized to solve all problems, even if the problems appear to be similar or related. The dictionary of context and the labyrinth of connections are far deeper than meets the eye.
Figure 7 – The task of understanding is far more difficult than perceived by the Anglo-Saxon world. For a binary system agnostic of natural language, the difficulty in understanding the meaning of the word call in the context of its use and user, may be difficult (if not impossible) to accomplish using the traditional semantic web ideas, which lack binary translation features in its tools. RDF (resource description framework) is the key to data interchange and relationships between things (triple) which leads to directed graphs (edges of relationships). Other tools, for example, OWL (ontology working language) which builds on RDF, further fine tunes the descriptive structure with the (misguided notion?) that it helps interoperability of data between dissimilar communities. The illustration above points out the semantic variability of the word “call” in only 2 languages. The descriptive structure without a binary translation holds the entire system hostage to words and syntax. Digital transformation of the descriptive structure with binary translation for context (call.1, call.2, call.3, call.4) will be beyond machine readable, it may be machine understandable.
Figure 8 – Digital Semantics – URN – Universal Resource Numbers for ontological frameworks? Each context has an unique id [numerical (knowledge) representation]. The list can grow (number spaces similar to EPC\textsuperscript{174} or UPC\textsuperscript{175} barcode) as communities/countries crowdsource and contribute to the “context list” the descriptive structures relevant to their language and cultural context. Mapping between a numeric system and categorization (taxonomy\textsuperscript{176}) with relevant to specific ontologies, that is, by associating a number with the context of use, the context can now be translated to binary. “Mireille bought a mirror” indicates the meaning of mirror (URI) in the context of an “object” (URN) but the statement, “it mirrors my life” conveys the meaning that Mireille is perhaps choosing to be “philosophical” about her life\textsuperscript{177}. The failure to promote URN as a global standard and enabler of the digital semantic web may be linked to its complexity. The latter may be one reason why the “marketing” department has failed to create a “value proposition” for the diffusion of URN. Irrespective of how great a technology may be in the context of advancing the knowledge society\textsuperscript{178}, its adoption, especially in business and industry, is solely a matter of economics. Elegant research on corn\textsuperscript{179}, dynamo\textsuperscript{180} and electricity\textsuperscript{181} has repeatedly demonstrated that the economics of technology is catalyzed by proof of profitability.
DATA SCIENCE – DATA FROM THE EDGE CONNECTS WITH SYSTEM OF SYSTEMS

The most important outcome from, and use of, any system, is its data. SITS may provide crucial information about the characteristics of soil and its microbial flora. This information may help decision support systems to feed 10+ billion people on the Earth, at the dawn of the 22nd Century. Data may be inert if divorced from its relationships. The effect of data and information on interconnected resources may be viewed through the lens of decision nodes. A node is where other paths may branch, meet, start or finish. Each node is a resource which connects to other nodes (may have weighted relationships with other nodes). The relationships between nodes, and sum of the attributes, are collectively referred to as a graph. Each edge of a graph (the line connecting the circles, the nodes), is a relationship. The internet is an example of a directed graph. In computing, a graph database is a database that uses graph structures for semantic queries with nodes, edges and properties to represent and store data which users can access, analyze and interpret, to derive value.

Figure 9 – Example of a graph (top, L), a system dynamics model (top, R), cartoon of an artificial neural network (b, Left) and sketch of a biological network of neurons (b, Right).
With decreasing cost of computation, memory and storage, graph databases are rapidly emerging, catalyzed by GPU servers (graphics processing units). A central processing unit (CPU) consists of four to eight CPU cores, while the GPU consists of hundreds of smaller cores, which is used for the graphics card, necessary for 3D gaming applications. CPU and GPU, together, operate to crunch through the data in applications. This parallel architecture is what gives the GPU its high compute performance. Graph databases demand high end computation because they hold edges, connectivity and relationships, in the form of graphs. The most interesting queries on graph data structures tend to have computational (requiring traversal), not analytical (closed-form), answers.

Because relationships are central to SITS, graph databases may be a key component of the SITS information architecture. Experts may point out that graph databases are still developing, which has implications for the ecosystem of connectors to write data to them, clients to read data from them and security/ops tools which are necessary for maintenance. In addition, certain forms of graphs (DAG, Directed Acyclic Graph) can be represented and traversed in relational databases, hence, the question, do we need graph databases?

Figure 10 – Graph DB: a future forward strategy? Is graph database right for SITS?
The primary interaction between SITS and SUGs (SITS user groups) may focus on the use of SITS as a repertoire for users to query relevant information. Compatibility with different query languages, for example, SQL, SPARQL, NoSQL DSLs (domain-specific languages), is mandatory. Graph databases are likely to handle multiple query languages. Perhaps, this may be an opportunity for innovation in semantic query languages in the context of SITS.

In the context of the data, context determines the users focus on the data and the focus, hence, the context, could vary. In other words, users may view the same data from different contexts (in software jargon, need multiple cursors). Depending on the application, the user is going to define the context (externally). Therefore, by separating data and context (application), the information architecture may provide users the flexibility to view the data from all possible perspectives (rather than fixed schema or set of relationships, typical in a relational database). Graph databases (semantic databases) enables creation of context-free data systems. For example, the value < soil pH 7.8 > may be important to one grower in the context of planting seeds, for another grower the context may be nitrogen fixation and still another user may use it to determine if upstream leaching of salts may be influencing the pH of downstream soil. The relationship between pH and planting seeds, pH and soil nitrogen fixation, pH and water management, is an enormous task, expected to be captured in graphs.

This task cannot be “complete” on day one. Graph databases may crowdsource this information and grow the repertoire from information curated/contributed from vast number of users. How can users “drag and drop” this information in the database? Open interfaces (API) to import sensor data (local, global) in the graph database, agnostic of file format (CSV, JSON, XML), is the expectation from smart data hubs (semantic data warehouse).

For context-agnostic data ingestion, the characteristics of ingestor logic gains prominence. Smart hubs facilitate direct ingestion of data (think, time series data from sensors) from external databases. One segment for creativity and innovation is how to embed metadata during this ingestion process, when the incoming data is converted to the form which is useful for the recipient database. How can we be creative in seeding metadata and improve this step continuously to better support [a] search [b] navigation [c] data identity keys [d] provenance information and [e] data governance?
To amplify the value of the SITS information architecture, can we use SITS graph database to serve as a (local, state, national, global) repertoire for different types of sensor data? The value of the future forward data and information architecture of SITS may not be limited by the sensors in Table 1. For example, if a laboratory offers sensor data unrelated to Table 1, how do we incorporate the sensor types in the SITS graph database? It will be necessary to create open interfaces (APIs) as “feeder tubes” to ingest data (from anywhere) relevant to the sensors that we added to the database.

One obvious approach may involve creation of an automated feature selection tool. It may be downloaded by labs interested to share sensor data. Feature engineering and selection, will enhance compatibility of data uploaded to the SITS repository. Labs may export their data using open APIs from the SITS portal. If feature selection criteria are similar, then feeding data directly to SITS database may be feasible.

However, it is desirable for feature engineering to be independent of data persistence, so that raw data can be replayed through new feature engineering ideas. In this respect the design of IoT (internet of things) in applications is a detail worth pursuing because in many IoT applications raw data has a narrow window to be persisted, since the attached storage, at the source of sensing, is typically limited (less storage may mean less cost). Therefore, if the data at the source is lost (historical data deleted from storage to make room for new data), then “replaying” the raw data (if there is a new feature engineering idea) is only possible if the transmitted raw data from the source was stored in the parent (graph) database (for example, using an open source time series database).

Importing raw sensor data and populating structured data fields may be sufficient in relational databases but in SITS graph database, setting up the graphs, the relationships, are equally critical. Replaying raw data in the context of a new feature(s) may create new graphs and other relationships may emerge, which could lead to questions not yet asked. Hence, these attributes (independence of data persistence, context-free data) allows data to remain accessible to multiple levels of queries. It enables questions to be asked by other users using “new eyes” to see old problems, which could lead to new approaches to trigger new solutions.
Knowledge graphs, at this time, are a human-driven endeavor to install semantic architecture (metadata, data, may have to be curated). Furthermore, building ontological schema is required to perform Google-like searches. Management of metadata in case of overlapping ontological schema (from different sources) may be ripe for automation to avoid human-centric curation as the rate limiting step, especially in the management of mapping from one ontology to another. The erroneous assumption is that the sensor data, in question, has associated metadata, semantics or any ontological schema.

Perhaps the fallacy, in this context, is the organizational ambition to create a single conceptual model for the expression of information between organizations. It means that everyone should use the same model, if they wish to communicate information, for example, RDF. It requires that the data in question is in RDF and the external content contains enough of the metadata model used by the source system to successfully execute the query.

But, SITS and its ecosystem (distributed databases that may contain sensor data and other information relevant to relationships between data, models and their context, from global organizations), are likely to be highly heterogenous, non-RDF based databases. In this context, knowledge graphs are better suited to handle heterogenous data. Multiple structured and unstructured data silos, connecting their data sets in a meaningful way and contextual manner (things and related data of things) makes this possible through knowledge graphs.

Rather than spreadsheets and folders, data in knowledge graphs are connected with respect to context and relations, which are critical for almost all users and industries (finance, manufacturing, banking, utilities, healthcare, retail, logistics, and agriculture). To extract value (use and re-use) from the immense volume of data in distributed data silos, users turn to algorithms (statistical, machine learning) to improve their search. Knowledge graph models, of knowledge domains, are created by subject-matter experts who weave the foundation of how the domains are connected. Therefore, the strength and value of the knowledge graph model is as good as, or limited by, the knowledge and bias, of the expert. We need tools to increase the breadth of knowledge and reduce the spread of bias. By creating structures, standards and common interfaces, the knowledge graph model may be improved by crowdsourcing the “knowledge” rather than reliance on a few experts.
Tools to create “low-bias, universal models” are few and far between. The dynamic breadth and incisive insight of knowledge graph models will be quintessential to generate not only the obvious outcomes but also the non-obvious relationships. Because knowledge graphs overlay existing databases or data sets (structured or unstructured), agility increases when new data is added or new data sources are linked. Standard outcomes and performance of NORA improves with increasing relationships in context of data. The ability of knowledge graphs to establish relationships is determined by (hence, restricted to) the knowledge of the knowledge domain expert(s) establishing rules, including vocabulary, ontology, taxonomy, related to the data. Thus, these rules cannot be governed by specific groups or remain limited within certain geographies (introduces cultural bias). The crowd-sourced modus operandi is key for knowledge graphs to learn, un-learn, and grow to embrace the rules stemming from different vocabularies (global users and their social bias) as well as ontologies from different bodies and plethora of tools (statistical techniques, machine learning, algorithms). This may not happen on day 1 but must be a salient feature for any knowledge platform of the future.

Figure 11 – Knowledge graphs come alive when we query “Alexa” what is the weather? (Left box). What we propose to achieve is the SITS version for ag. We may ask “Hypatia - which types of bacteria are in the soil sample from Alexandria?” (Right box). Note: re-visit Figure 1
TEMPORARY CONCLUSION

The suggestions here are amorphous but enabling mass adoption\textsuperscript{200} of these tools remains the Holy Grail. Adoption of these tools are a pre-requisite if we view SITS as the future anchor and resource for “ag Google” which may serve as an “one-stop shopping mall” for scientists/researchers to upload their data and users/growers to access that data to answer questions. The emphasis on open platforms and open APIs in the information architecture is key to engage with diverse sources of information, ingest big data, if relevant, and invite non-compete alliances to enable industry to contribute and benefit (profit) from this digital initiative through data analytics and intelligent information arbitrage. SITS may be ag’s answer to the “geological Google” (Deep-time Digital Earth Initiative\textsuperscript{201}) which is one reason for suggesting open source\textsuperscript{202} GPU\textsuperscript{203} database in the design of SITS.

Figure 12 – Viewing the ancient library card catalog with “\textit{new eyes}” in the Semantic Data Catalog. SDC contains a link (URL, URI, URN) to a resource and information/connections to other SDCs, to navigate across the information space. It does not actually include the resources, only their addresses and metadata, to identify a particular representation of that resource, in a traditional database. It\textsuperscript{204} appears to be, in principle, similar to the transition when actual data on the RFID tag was replaced with an unique 64-bit EPC, which used PML and ONS to indicate, where to find (link to an URL) the information relevant to the unique ID (EPC), that was read by the RFID reader from the RFID tag. IoT systems\textsuperscript{205} are increasingly using of semantic graphs to keep up with the edge\textsuperscript{206} of connectivity.
Significant breakthroughs may be necessary. Part of the scientific content in this haphazard essay is almost pure speculation (quantum spectroscopy). The data science section is a regurgitation of common knowledge. Creating the edge connectivity (SITS graphs) may be a mammoth undertaking, which may mature over years, or decades. An example of a related task for oncology may be just the tip of the iceberg. It will be foolish to assume that the primary repertoire for the future “ag Google” may be an easy undertaking.

In attempting to provide my two cents on this *threepenny opera* I attempted a task much too great for my abilities, the extent of which I didn’t perceive when I started. We choose to pursue these ideas “not because they are easy, but because they are hard.”

Figure 13 – Signals in the soil databases will not only provide access to the data but also display GIS maps to show source of the signal (and environment, for example, tomato growers in Florida or rice paddy on terrain in Yunnan Province, China). To render realistic views of images at a high frame rate, GPUs process massive amount of geometries and pixels in parallel, at high speed. The clock-rate increase for processing units has plateaued but the number of transistors on a chip is increasing. GPU computation speeds, (gigaflops per sec, GFLOP/second), are increasing. Graph compares GPU vs CPU trends.
REFERENCES

8. https://canvas.harvard.edu/files/3486763/download?download_frd=1&verifier=qRAG5jUIbng83WzEBnQOp9GgeXfw0hCkeMYiJ
11. https://www.nap.edu/download/25059
14. http://science.sciencemag.org/content/360/6393/1126
15. https://www.nature.com/articles/s41586-018-0238-4
16. https://www.nature.com/articles/d41586-019-00599-8
25. https://www.youtube.com/watch?v=BCIohEJYxdY
27. http://robotics.sciencemag.org/content/4/27/eaaw1875
29. https://www.nature.com/articles/s41928-018-0173-2
34. http://jgp.rupress.org/content/jgp/44/1/19.full.pdf
36. https://pdfs.semanticscholar.org/i363/6dedc27227fa5a52603ccc16a5127f48085.pdf
37. https://pdfs.semanticscholar.org/839b/dbc3ce47df32b04b9722aa492d51b3921268.pdf
38. https://core.ac.uk/download/pdf/61188787.pdf
Triumvirate Modes of Thought: SITS - an opportunity for imagination, invention, innovation & implementation
Triumvirate Modes of Thought: SITS - an opportunity for imagination, invention, innovation & implementation

173 https://dspace.mit.edu/handle/1721.1/41902
174 https://www.gs1.org/standards/epc-rfid
175 https://www.gs1.org/standards/barcodes/ean-upc
177 https://www.tandfonline.com/doi/abs/10.1080/2287108X.2012.11006070
178 https://www.economics.ox.ac.uk/materials/working_papers/paper084.pdf
179 https://pdfs.semanticscholar.org/02be/4bd7e9f91a20fe522edd041ac6708889dc7.pdf
183 http://snap.stanford.edu/class/cs224w-readings/broder00bowtie.pdf
184 https://neo4j.com/developer/graph-database/
185 http://dx.doi.org/10.4135/9781526410993
186 https://www.youtube.com/watch?v=6Z_XVBECJzw
188 https://hackernoon.com/which-gpu-database-is-right-for-me-6cedf6a17505
197 https://ieeexplore.ieee.org/document/8081906
202 https://erijs.mit.edu/
205 https://www.nature.com/articles/s41598-018-36973-1
207 https://earlyprintedbooks.com/ames_typographical-antiquities_1749_a1r/
208 https://eng.uber.com/aresdb/
210 https://er.jsc.nasa.gov/seh/ricetalk.htm
211 http://www1.cs.columbia.edu/~ji/F02/ir02/p44-perlman.pdf
214 http://eyerijs.mit.edu/
215 https://www.nature.com/articles/s41598-018-36973-1
222 https://weng.uber.com
223 http://www1.cs.columbia.edu/~ji/F02/ir02/p44-perlman.pdf
224 https://www.bloomberg.com/graphics/2018-us-land-use/
Algorhyme

I think that I shall never see
A graph more lovely than a tree
A tree whose crucial property
Is loop-free connectivity.
A tree which must be sure to span
So packets can reach every LAN.
First the Root must be selected
By ID it is elected.
Least cost paths from Root are traced.
In the tree these paths are placed
A mesh is made by folks like me
Then bridges find a spanning tree.
While the U.S. benefits from an overall agricultural trade surplus, Americans imported 15 percent of their food and beverage products in 2016. More than 30 percent of the fresh fruits and vegetables Americans consume come from other countries, predominantly Mexico and Canada. The amount of U.S. land used to produce citrus fruits alone is larger than Rhode Island.
SIP □ SAR flows into SARS💧 AG

Shoumen Palit Austin Datta
Table of Contents

- SIP □ SAR
- SARS ♦ AG
SIP SAR and SARS AG are similar and related ideas, but neither contains any new thinking, just different perspectives. If read in series, it perhaps boils down to systems, data interoperability, platforms, extraction of value from data to generate actionable information and/or knowledge which may update experience (bank).
For those who wish to think, connect ideas using thought experiments and may discover happiness in allowing their imagination to wander.

"Not all those who wander are lost“ from the poem "All that is gold does not glitter”

J. R. R. Tolkien
SIPSAR

Shoumen Palit Austin Datta
Email - shoumen@mit.edu
PREFACE

In the US, a quarter million patients die each year due to preventable medical errors. One leading cause is the lack of medical device interoperability and the lack of data integration platforms. The latter is quintessential to treat the patient, as a whole. The widespread practice of viewing stand-alone data from multiple medical devices, attached to the patient, manufactured by different medical device equipment companies, increases the risk of fatal errors, due to the reliance on medical professionals, who must integrate all the data, in order to actuate treatment.

The human heart, as an example of a machine, may be deconstructed into parts and components. Hence, a bill of materials (BOM) may be prepared and various parts/components may lend itself to monitoring using signals and sensors. Reconstructing the signals on a common “heart health platform” may reveal the state of heart health. The knowledge, dependencies and case logic, embedded in the platform, may remain cryptic to the point-of-care end-user. But, the combined outcome, based on analytics of the data, from signals and system of sensors, from various medical devices, may aid in the precision and accuracy of the decision support system (DSS). Humans in the loop may use the outcome to design preventive measures, recommend maintenance (for example, valvuloplasty) or prescribe medication (for example, statins, a class of lipid-lowering drug which is a HMG-CoA reductase inhibitor).

The thrust of this position paper is to recommend that we pursue an analogous approach for all machines and devices, to maintain the health of machines. We may create an open source approach to synergistic integration of data, using “plug-n-play” modular platforms, running analytical engines to extract information, from system of sensors. Promoting interoperability between platforms may unleash the potential for diagnostics using real-time data. Taken together, it may mitigate mechanical risks due to usage, reduce uncertainty of operational downtime, prevent energy waste, optimize output efficiency, catalyze savings from economies of scale and use of big data.

The application of this modus operandi is scalable for manufacturing, through integration with legacy manufacturing execution systems (MES), product lifecycle management (PLM) systems and advanced planning and optimization (APO) routines, commonly found in classical enterprise resource planning (ERP) systems. Shop floors with a few or a few hundred machines, are equally capable of harvesting the value from synergistic integration. We can extend this approach to [1] a collection of diverse machines, for example, farm equipment, [2] different devices, for example, home health monitoring, wearable body sensors for wellness and healthcare, [3] transportation network of vehicles, combined with traffic data, [4] optimize load balancing in energy distribution using smart grids.

Our suggestions are neither unique nor novel. The strength of this proposal is grounded in an open-source, modular approach, which can support scalability, interoperability and real-time communication for transparency, with suitable data protection (cybersecurity, privacy, identity and authorization). Proprietary platforms often retard productivity by reducing economies of scale, may exclude outlier events which could influence performance and increases TCO or total cost of ownership (for example, capital expenses, cost to operate) for global operations, if multiple types of platforms, which may lack interoperability, give rise to local silos or cannot “talk” between regions. Open-source optimization for customer-centric value is not synonymous with corporate profit-centricity. Hence, transforming this vision into reality, albeit in parts, requires a PPP (public-private partnership) framework with pre-competitive collaborators from industry and users (customers), along with academia and government agencies, to fund this initiative, and enable large scale implementations. Monetization of open source platforms, and the service supply chain of data acquisition tools (sensors, communications), are quite common. To accelerate global diffusion, a pay-per-use, micro-revenue business model, may lubricate market entry, and amplify, long term, ethical profit.

1 https://doi.org/10.1136/bmj.j2139
2 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052027/
FRAMEWORK

The conceptual structure, of the synergistic integration platform infrastructure, may be loosely illustrated by the old telephone switchboard (Fig 1). The “plug-n-play” input resembles the potential to aggregate data streams from sensors, actuators, RFID tags and other tools. In the logic layer background, the nature or type of data may trigger data-dependent selection of one or more tools, for example, analytical engines, “containers” with algorithms for data-driven micro-services and other virtual machines, which may be on-site or sourced, ad hoc, from clouds, fog, or mist computing repertoires. Automated programming and feature engineering may be included with ML.

Fig 1: The old telephone switchboard may be conceptually analogous to future ‘plug-n-play’ integration platforms.

The analytical output may feed decision support systems, and/or actuators, and/or executable software programs, to act/modify physical systems (for example, cyberphysical systemic control of valves based on rate/flow). The visualization of the data on a mobile device (smartphone) may resemble the cartoon in Fig 2. The illustration is adapted from an agent-based inventory monitoring example which used RFID data. For our current discussion, each component from the BOM which can be sensed, may be attached with a sensor (vibration, temperature, pressure). The sensor data will be transmitted to an Agent system where one dedicated software agent will be responsible for acquiring the sensor data and feeding it to an ‘agency’ which will converge inputs to trigger alerts or execute action.

Fig 2: Monitoring mean time between failure (MTBF) from sensor data on a mobile smartphone. With decreasing residual lifecycle (upper panel) the probability of failure (broken part may precipitate unplanned downtime) increases (lower panel). Just in time spare parts supply chain coordination with maintenance, using MTBF data, prevents potential machine failure. Probability of failure drops to “zero” (lower panel) when the part at risk is replaced. This example of an agent-based system uses MTBF metrics and ML to guide preventive maintenance.

3 https://ieeexplore.ieee.org/document/7506647/
6 http://dspace.mit.edu/handle/1721.1/111021
IMMENSE SCOPE

Data from sensors integrated into sensing-compatible components of machines (for example, hoses, seals, belts) can transmit data about operating conditions (temperature, pressure) as well as micro- and macro-system wear, as a measure of system health. In principle the ‘sense & response’ paradigm may not be restricted to devices, machines, and energy. The scope of sense & response can embrace much more than manufacturing (e.g. Wabash), farm operations (e.g. Fair Oaks Farms) and robotics (e.g. Automotive Robotics). The scope may extend to system of systems including machines, buildings, roads and emergency response systems. In other words, this synergistic integration platform (SIP) can distribute the sense & response (SAR) paradigm to gauge the real-time “health” of the university campus, community, city, county, state and a bird’s eye view of the networked society (cartoon below).

SIP SAR is driven by the mantra of connectivity, the primary barrier preventing the implementation of the scope. Lack of communication between systems, loss of data due to absence of interoperability, and the chasm between manufacturing processes and products, are all remediable injustices, yet they are still holding back global productivity. However, tracking money and financial transactions are making better use of connectivity.

One purpose of this proposal is to identify gaps which may be addressed through research to improve connectivity and, in turn, productivity. In parallel, with a greater priority, this proposal expects to coalesce PPP to deliver very large scale implementations of SAR (sense & response paradigm), which has been popularized by the metaphor of internet of things (IoT). IoT is a digital by design approach, which encapsulates the connectivity of the networked physical world where the duality of the physical, and the digital, converges into a fluidic continuum.

Fig 3: IoT is a metaphor without borders. It represents a continuum made up of discrete objects, decisions, processes and outcomes, each of which, in turn, impacts and influences, man and machines, life and living.

7 https://dl.acm.org/citation.cfm?doid=3131672.3131702
CONNECTIVITY AND CHEMISTRY, IN COMBINATION, PRESENTS THE NEXT EMERGING CHALLENGE TO THE STATUS QUO OF MANUFACTURING. THE QUANTUM LEAP FOR MATERIAL SCIENCE IS A RAGS TO RICHES STORY WHICH HAS THE POTENTIAL TO CHANGE ALMOST EVERYTHING IN CLASSICAL MANUFACTURING AND Usher IN CYBER-MANUFACTURING AS WELL AS THE FUTURE DIGITAL FOUNDRY.

Additive manufacturing of metals, composites, biomaterials (joints, prosthetics), will re-form the supply chain using distributed additive manufacturing on-demand and replenishment (DAMODAR). The inclusion of “replenishment” signifies the profound supply chain disruptions, and business discussions, that distributed on-demand direct digital manufacturing (cyberphysical) systems, is poised to usher. The “distributed” local physical presence, will benefit from the confluence of ideas from global digital connectivity (metaphor of IoT-by-design). Digital design tools (cyber component) may harvest “best of breed” outcomes from experts and benefit from location-agnostic convergence of deep knowledge. The best parts can be printed in Mongolia and Mogadishu (as well as, on a military aircraft carrier). Thus, in combination, it may influence the context, quality, metrology, and performance, of local physical manufacturing (3D printed additive manufacturing products (DAM) as a service).

3D printed objects with integrated sensors presents the next opportunity for precision in detection, diagnostics and advanced analytics. Imagine a 3D printed hip joint using metal-ceramic composites or other metal alloys that may be a partially hollow mesh integrated with sensors (gyroscopic sensor, glucose sensor, CTx-1 collagen sensor for bone loss) and room for cancellous bone (spongy bone) containing hematopoietic precursors (forms blood cells). Therefore, direct digital manufacturing brings us back to where we started. Cyber-manufacturing can work as a catalyst to democratize healthcare. Using distributed additive manufacturing, the advanced hip joint may be printed on-site for fortunate folks in Anchorage as well as the less fortunate people in Addis Ababa.

Fig 4: Mechanical properties of additive manufactured alloys compared to conventionally processed counterparts.

12 https://patents.google.com/patent/US4575330
13 https://www.darpa.mil/attachments/DARAPA60_publication-no-ads.pdf
14 https://pubs.acs.org/doi/10.1021/acsami.8b06903
16 http://www.mdpi.com/2224-2708/7/1/10
TRANSCENDENTAL CONNECTIVITY

Chadwick (the discoverer of the neutron) was a student of Rutherford (discoverer of the proton) who was the student of Thomson (the discoverer of the electron).

SIPSAR is Appendix II in Chapter 5 of the book ‘Haphazard Reality - IoT is a Metaphor’ available from the MIT Library.

http://dspace.mit.edu/handle/1721.1/111021
INTRODUCING SARS – A DESIGN METAPHOR CATALYZED BY UBIQUITOUS CONNECTIVITY

Unless adopted and consumed, products and services are sterile and impotent. R&D exercises, and experiments, may serve socio-economic needs and offer return on investment (grants, funding), if it is guided by users, as a part of its user-centric design. However, often people do not know what they need. Hence, user-centricity is a suggestion, not a doctrine.

Generally, usability should not be an after-thought. The outcome and user adoption, are essential elements of design. SARS for ag (agriculture) prefers an user directed design, may commence with the “sense” segment initiated by data acquired from sensors, often at the behest of the end user. The idea is agnostic of vertical and serves as a foundation.

The SARS paradigm is an example of synergistic systems integration, an open platform approach to deliver outcomes, decision support or suggest solutions, mostly quantitative, using defined concepts (sense, analyze, response, systems). It is applicable to almost any domain.

SARS is not a paradigm shift. It is a concept which has come of age due to decreasing cost of computation, storage, telecommunications and systems. The principle of SARS existed from the dawn of human thought, procedural sequencing and the organization of logic.

Feeling with your fingers, the water temperature of the sea water, is a form of sensing. Based on the “sense” of the temperature (analysis), the response may be a decision to swim in the sea/ocean. The data, analysis (based on stored data, past experiences) and response, was transmitted by the sensory system (skin\(^1\)), neural system (signal transduction) and the musculo-skeletal system (use of arms and legs), using a platform to aggregate data, coordinate signals and actuate motor neuron activity, that is, the brain. This bio-inspired paradigm is SARS and we have replaced the human functions (for example, feeling with fingers) with engineering tools (temperature sensor). Tools help us to sense vast number of parameters and analyze volumes of data but, depending on the complexity, the tool-based outcome feeds decision support systems with humans in the loop, generally, to execute necessary steps or trigger processes.

Examples of SARS may include the “low fuel” gauge in an automobile, the home blood glucose test which may prompt you to visit your physician or the smoke alarm in your home. SARS in the context of internet of things (IoT\(^2\)) may be intelligent machine maintenance\(^3\) which may be viewed as a part of Digital Twins\(^4\), another emerging trend. In every case, connectivity, between different components, is key for data fusion and data analytics, to determine if there is any actionable information in the data, before the value of the data perishes.

---

1 10.1016/j.protcy.2014.09.015  
2 https://dspace.mit.edu/handle/1721.1/111021  
4 https://arxiv.org/abs/1610.06467
Figure 0: The concept of SARS is embedded in the metaphor of Digital Twins. Shell’s Floating Production Storage Offloading unit (Bonga FPSO, top, has a storage capacity of 2 million barrels of crude oil). It is a digital twin project (left cartoon, top panel) executed by Akselos, a MIT spin-out, a high-tech start-up with advanced mathematics at its core. Akselos helped Shell create “digital twins” or computational models of physical structures (bottom panel). Digital Twins mirror the exact characteristics and operating performance data (of assets). It offers decision support for operators (abnormal stress, strain, temperature), allowing information to assist maintenance decision systems, and insights on asset life cycle management and supply chain.
SARS FOR AG (AGRICULTURE)
This essay addresses SARS for agricultural end-users (water, soil, air, crops, animals). Due to a vast array of potential scenarios, this proposal often refers to a proof of concept (PoC), which may be achieved using a sub-set of data (for example, sensor data from water/soil/air). The limited scope of the PoC is with respect to the deliverable, that projects often demand. It is temporary and only due to the financial resources which may be restrictive in a grant funded case. In this approach, the scalability of the PoC surfaces as an intrinsic litmus test for the PoC.

At the architecture level (for example, data lake\(^5\)), the scaffold of the PoC will be highly modular, to facilitate “plug-n-play” operation with structured and unstructured data streams, from a variety of sources (including, so called, big data, and, the pursuit of open\(^6\) standards, for example, for publish/subscribe data distribution services). The PoC, may limit itself to use only a few sources of data: (a) water quality [water sensors] and (b) soil components [soil sensors], before and after, using specific protocols (for example, waste water treatment). The “skeleton” of the SARS “tool” for agriculture must be an open platform. Other groups and consortia can add to this data platform, leading to organic growth of the tool for specific or general purposes.

SARS PLATFORM – TOOL DESIGN AND PROOF OF CONCEPT TASK
The platform (and its PoC sub-set) is expected to connect different components, sub-systems and domains\(^7\) to add value to the outcome (pragmatic solution), intended to deliver tangible benefits. For example, the sensor must successfully transduce the signal to (preferably) a mobile device to upload, and/or analyze, the data. The outcome (value of data, meaning of data) may be visualized on a device (smartphone), in a manner that a non-expert may extract actionable information from data, and know, how to, benefit from suggestions or instructions.

The description of the outcome starts with the end user. The outcome, which the end user anticipates, is the “deliverable” scenario. This is the task the SARS TOOL must accomplish.

The design of the SARS TOOL starts from the end, that is, the outcome in the context of the user. By sequential deconstruction of the “outcome” we arrive at sub-systems, components and sub-components (analogous to BOM or bill of materials, in manufacturing). These “parts” when assembled, synthesized and harmonized, creates a functional system or system of systems. The functional and synergistic integration between parts, flow of data and analytics, capable of extracting information (if there is any), will generate the outcome, expected by the end user. The user interface (GUI or app on mobile device) is the conduit to initiate the cascade and return the outcome to the GUI or visualized as another app on the device. The user needs to know how to ask the correct question, which in turn, will trigger the cascade and how to utilize or benefit from the outcome.

\(^6\) http://opendds.org/
\(^7\) https://pubs.rsc.org/en/content/articlelanding/2018/an/c8an00065d
SARS PLATFORM – SCENARIO, SUB-COMPONENTS, SYSTEMS

General SARS is a “google of ag” approach where humans (“user push”) input query or data, for example, the following may be entered on a mobile app (dialog box pop-up):

[a] “blood glucose = 157mg/dl” “hello system - what does it mean?”
[b] “mercury in water = 85 micrograms per liter” “hello system - suggest water treatment”

In phase II, we move forward by going back, almost a hundred years. SARS infrastructure will not only handles ‘user push’ but will complements with proactive system response or ‘system push’ (“Hello Rebecca – I have new soil spectroscopic sensor data, click on this link”). Bioinspired “cybernetics of ag” is bidirectional, discussed by Weiner in the context of sensors at the edge of the eye, which vie for the attention of the analytical core in the fovea centralis.

The SARS TOOL PoC may focus on [a] water/soil sensors and [b] water treatment. In that context the platform rests on two key sources of data (data issues are discussed later).

In Figure 1, the SARS “platform” (table top in the illustration, below) is a multi-layered application design environment, which connects to the user interface (smartphone, Fig 1). The application logic, decision support tools, machine learning algorithms, feature engineering, cybersecurity and others, including APIs to facilitate data interoperability, and re-distribution, essential to the function of SARS, connects to this design environment.

Figure 1: Over simplified illustration of the SARS PLATFORM metaphor and its connectivity to data (database or datastore may be real or virtual, that is, sourced from cloud, fog or mist). The platform “aggregation” role makes it imperative to offer an open platform architecture where data sources may continuously feed in or drop out (dynamic composition / decomposition). The visualization tool for the platform may be a mobile device, eg, smartphone. The dialog with the platform is executed via an app on the smartphone, which is the user-controlled interface. The red “strings” illustrate link to data resources (eg sensor data, wastewater treatment data). The architecture provides for multi-directional communication (user push, system push, push-pull).

---

8 https://www.nature.com/articles/s41467-018-06773-2.pdf
10 https://libraries.mit.edu/archives/research/collections/collections-mc/mc22.html
11 https://oaspub.epa.gov/tdb/pages/general/home.do
The view of the SARS PLATFORM with two “red strings” or two data resources (Figure 2) illustrates the hypothetical PoC linked to limited data [a] water/soil sensors and [b] wastewater treatment protocol. The “user push” query in natural language may be in this form “what is the waste water treatment protocol when the mercury sensor reads mercury in water is equal to 85 micrograms per liter?” The dependencies and inter-relationships between sensor data and treatment protocols are defined in the application logic layer. The data and query triggered by the user generates a response, in another dialog box, which pops-up on the smartphone.

Figure 2: Proof of concept for a SARS platform tool may be created and delivered with only two sources of data, illustrated by two red strings (illustration source14).

The software architecture of the open platform, its connectivity to data lakes and access to cloud storage, may be designed to embody the broad-spectrum vision of the SARS platform, which can serve a plethora of purposes, and is not limited to digital connectivity of atoms to bits. This platform may also serve as a “news and views” clearinghouse for information storage and “analog” information arbitrage by humans, for example, the successful CPS-VO15 model.

Open architectures may enhance or catalyze “plug and play” operation when integrating other data sources or connecting to “big data” resources (for example, weather, crop, fertilizers) from different clouds, to feed the SARS platform. The PoC, at hand, may functionalize the tool with an app that can receive (input) and send (output) data and suggestion related to [a] water/soil sensor data and [b] wastewater treatment protocols.

The potential for dynamic composability unleashes unlimited flexibility to add different data resources and receive feedback for complex questions. SARS platform needs a visionary open architecture as a part of its infrastructure which can act as a scalable scaffold. The nature of data, dependencies between data, and application logic, that is suitable to connect / mine / combine the data, to generate useful information, are software functions which can be added or subtracted to the application design environment. Modularity and fluidity of the software system will determine the ability, or the inability, of the SARS platform tool to evolve. Hence, the use of agent-based architectures16 may be central to the design of the SARS platform.

16 https://dl.acm.org/citation.cfm?id=122367
User adoption of the SARS platform tool may be directly proportional to the functional strength of the tool, that is, the **value of the output**. The strength of the tool to deliver value-added output, in the form of recommendations, decision support and predictive suggestions, may be proportional to [a] depth of the noise-free curated data and [b] foresight embedded in the logic layer, which must correlate and synthesize data dependencies **between** domains. The latter is a difficult task with increasing number of variables (explosion of state space).

SARS will evolve as a data-driven logic-based platform. Data analytics in the application design environment is a pivotal role and a key performance indicator (KPI) for value and efficacy of the SARS platform. It will benefit from old\(^{17}\) ideas and new developments\(^{18}\) in data and analytics. The latter are widely practiced\(^{19}\) in several fields\(^{20}\) based on statistical (ML) machine learning techniques (neural nets, deep learning\(^{21}\)) and attempts to automate\(^{22}\) data science.

---

\(^{17}\) [https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html](https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html)


\(^{19}\) [http://dx.doi.org/10.1016/j.energy.2012.10.052](http://dx.doi.org/10.1016/j.energy.2012.10.052)

\(^{20}\) 10.1021/acs.energyfuels.7b01415


SARS PLATFORM – PHYSICAL COMPONENTS AND RELEVANCE TO CYBERPHYSICAL SYSTEMS

The physical components of the SARS platform are an integral part of the paradigm. Therefore, by extension, SARS may be viewed as a subset\(^23\) of cyberphysical systems (CPS\(^24\)). The physical “things” in the context of SARS, must be connected to deliver value, for example, sensors for sensor data. Therefore, SARS is also a part of the internet of things (IoT) metaphor. The SARS platform may benefit from the wisdom of experts, which has generated thoughtful works related to CPS\(^25\) and IoT\(^26\) architectures. Other generic\(^27\) frameworks are also common.

Sensors and “things” are located in the physical layer (bottom of Figure 4). Data from the physical layer is uploaded to the cyber layer where connectivity, between systems, is key. The top layer is the (IoT) connectivity between systems of systems (internet of systems).

The SARS platform proposal may immensely benefit from the CPS framework, especially the cybersecurity guidelines. The risk due to food safety and security may be reduced if SARS for agriculture may follow NIST recommendations\(^28\) for security and threat assessment.

---

Figure 4: Architectural framework\(^25\) for cyberphysical systems proposed by NIST. Most platform paradigms (for example, digital twins, SARS) may be a subset\(^29\) of CPS (cyberphysical systems).

---

\(^{23}\) [https://doi.org/10.1007/s10845-018-1433-8](https://doi.org/10.1007/s10845-018-1433-8)

\(^{24}\) [https://cps-vo.org/](https://cps-vo.org/)


\(^{28}\) [https://www.nccoecnist.gov/](https://www.nccoecnist.gov/)

\(^{29}\) [https://ptolemy.berkeley.edu/projects/cps/](https://ptolemy.berkeley.edu/projects/cps/)
SARS PLATFORM DEPEND ON DATA

A cursory survey\textsuperscript{30} of types of sensors (bundled by loose similarity\textsuperscript{31}) is overwhelming, simply due to volume\textsuperscript{32} and variability. The inordinate complexity of sensor types in agriculture (weather, water, soil, pathogens, chemicals) presents a curse of dimensionality\textsuperscript{33} with respect to signals. The latter presents an almost insurmountable barrier to make collective sense of data, from different sensors and network of sensors. One approach to streamline the quagmire is a call for protocols\textsuperscript{34} and standards\textsuperscript{35} which offers economic\textsuperscript{36} incentive. Agnostic of verticals, standards may accelerate\textsuperscript{37} adoption if industry values such non-competitive agreements. This effort may not possess the resources to organize the sensor industry\textsuperscript{38} to begin any discussion on standards. Even the FDA had to resort to legislation\textsuperscript{39} just to define what is a medical device.

These issues are detrimental to making collective sense of data, unless all data sources were transmitting data with common characteristics and handled by a single analytical engine. Taken together, that would have to be a sole source, proprietary product, closed to open innovation and an anathema for distributed systems (ag is a globally distributed system).

Making collective sense of data is crucial, even without sensor standards for specific analytes (eg manufacturers pre-agree to specifications for sensors). Therefore, we must think about attempts to “standardize” the data rather than the sensor. The signal (data) must be characterized, then, grouped, collected and analyzed, as a data set with specific characteristics. In other words, feature selection\textsuperscript{40} and tools\textsuperscript{41} for feature engineering\textsuperscript{42} are emerging as vital tools to identify key, or selected, characteristics. Sensor-agnostic data must feed these features. Without shared features, data cannot be joined (merged, combined). This is a strength and may be a weakness, too. What if, the data with shared features was joined (merged) but the process eliminated outliers or data with uncommon features? Outlier events and unique data may offer clues to non-obvious relationship analysis and/or determining unknown unknowns.

The power of this proposal, partly, rests on such foresight, infrastructure and dynamic analytical strategy, if embedded, at the foundation of the SARS platform, with respect to its ability to aggregate, curate and extract value from widely diverse sources of unstructured data.

\textsuperscript{30} http://bit.ly/PLASMONICS
\textsuperscript{31} http://bit.ly/GAS-SENSORS-01
\textsuperscript{32} http://www-analytik.chemie.uni-regensburg.de/wolfbeis/P2-www.pdf
\textsuperscript{33} https://doi.org/10.1155/2018/7467418
\textsuperscript{34} http://www.ieee802.org/
\textsuperscript{35} https://www.gs1.org/standards/epc-rfid
\textsuperscript{36} http://www.econ.yale.edu/growth_pdf/cdp984.pdf
\textsuperscript{37} https://www.futuremedicine.com/doi/10.2217/pgs-2018-0028
\textsuperscript{38} http://www.sens2b-sensors.com/directory
\textsuperscript{39} www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM401996.pdf
\textsuperscript{40} http://dx.doi.org/10.1016/j.compeleceng.2013.11.024
\textsuperscript{41} https://people.eecs.berkeley.edu/~dawnsong/papers/icdm-2016.pdf
The management of structured data, based on defined syntax and semantics, in data dictionaries, is helpful for maintaining homogeneous institutional data. Its value in information arbitrage is declining. Therefore, we must seek dynamic strategies and agile architectures\(^{43}\) with built-in change control\(^{44}\) to mine unstructured data, which is emerging as the norm, rather than the exception. The value of the PoC in the context of IoT\(^{45}\) for farming and agriculture 2.0 concept\(^{46}\) must address data granularity and how each piece of data may be made to reveal its information, if there is any information in the data and if the information is actionable.

The management of unstructured data, therefore, is central to the proposed platform. Structured data is a subset of unstructured data. Management of structured data can only deal with structured data. Strategies to address platform requirements (and PoC) using structured data management, may be a nail on the coffin of data driven decision support systems (DSS).

The gradual diffusion of the ideas related to precision agriculture\(^{47}\) is directly related to the ability of the farming community to rapidly adopt tools and technologies. With the deep penetration of social media (messaging, twitter, etc), it is well nigh impossible to expect that communication between systems and people will exclude social media formats (natural language). The exchange of data and information may not be limited to apps on smartphones or tablets, provided as a part of the service pack by service providers (companies, agencies, organizations). To mine and use the social media data, in this context, unstructured data handling architecture must be at the core of the SARS infrastructure. The definitive answer is to fully embrace unstructured data and build capacity for change control\(^{48}\) to deal with big data, because most forms of data, under the “big data” category, may be strenuously unstructured.

However, just because we have acquired data, does not guarantee it will contain any information, irrespective of the data acquisition tool, and no matter how sophisticated the analytics may be or the perceived power of algorithms in an engine. Data devoid of information may be natural or an artifact, if the value of the data perishes before information retrieval.

In any effort, to manage\(^{49}\) data and to make sense of data in real-time (before the value of the data perishes), it is vital to use a diverse portfolio of mathematical and statistical tools, solvers and analytical engines, to curate and extract actionable information, by unpacking and unleashing value\(^{50}\) from data, and then using the information, in near real-time, to actuate response (sense, analyze, response system \(\Rightarrow\) SARS). The latter is a segue to a new paradigm, which extends “sense, analyze, response” to include sense, analyze, response, actuate (SARA).

---

\(^{43}\) https://ieeexplore.ieee.org/document/7868325?part=1
\(^{44}\) https://rmas.fad.harvard.edu/pages/change-control
\(^{45}\) https://ieeexplore.ieee.org/document/8126169
\(^{48}\) http://vcaf.berkeley.edu/sites/default/files/change_control_process_au.pdf
\(^{49}\) https://journals.aom.org/doi/10.5465/amj.2014.4002
STRUCTURE OF THE INFRASTRUCTURE – ARCHITECTURE, DATA HANDLING, OPEN TOOLS

The SARS tool, may be a convergence of deep\textsuperscript{51} systems\textsuperscript{52} thinking\textsuperscript{53} leading to system dynamics\textsuperscript{54} combined with lightweight but agile IoT \textbf{connectivity} services, IFTTT\textsuperscript{55}. In the first step, the PoC may be a web-based simulation\textsuperscript{56} to display its potential value\textsuperscript{57} for users and/or socialize the SARS tool via extension\textsuperscript{58} and education\textsuperscript{59} programs for dissemination\textsuperscript{60}. The PoC tool, when functionalized with real data, temporarily limited to [a] water/soil sensor data and [b] wastewater treatment, may provide limited set of “answers” from an “user push” approach.

Irrespective of whether it is a simulation sandbox or real data driven “google of ag” tool, and no matter how small or large the data set and/or data volume may be, we must begin the task of laying down the \textit{foundation} of this platform, which may be extended and expanded by future teams. We begin by addressing the “big picture” and user-centricity of the global SARS tool. The short term PoC deliverable is a local, nano-version “sandbox” of the future.

Open system architectures and modularity, are crucial, to enable the system to grow on demand, amplify in any dimension, and scale to handle large number of variables, diverse data structures, and \textit{unstructured} data. Structure/schema rigor must co-exist with \textit{ad hoc} event driven architectures, eg, using containers for portability on multi-cloud platforms \textit{as well as} serverless functions (nano-services). For example, architect to create a serverless function, which triggers a task, which, in turn, spawns a container and runs a longer-term process, like a complex time series data from a sensor network (which may require hours to complete). At present, we will loosely focus on 2 issues: [1] feature selection, and [2] software tools for data.

\textbf{Feature Selection}

To establish a common plane of reference and a calibrated baseline, the performance of the sensor and the sensor data (value derived from the measurement of the specific analyte) are equally important. Agnostic of the use case, without robust performance evaluation, any measurement data will lack credibility and the (sensor) measurement (data, output) may not be trusted to drive decision\textsuperscript{61} support (for example, operations or supply chain\textsuperscript{62} decision support).

\textsuperscript{51} \url{https://www.thinkmind.org/download.php?articleid=bustech_2015_1_30_90048}
\textsuperscript{52} \url{http://jasss.soc.surrey.ac.uk/17/4/2.html}
\textsuperscript{53} \url{http://www.sfu.ca/~vdabbagh/Forrester68.pdf}
\textsuperscript{54} \url{https://en.wikipedia.org/wiki/System_dynamics}
\textsuperscript{55} \url{https://doi.org/10.1080/01639269.2014.964593}
\textsuperscript{56} \url{http://web.mit.edu/jsterman/www/SDG/MFS/simplebeer.html}
\textsuperscript{57} \url{https://ctl.mit.edu/sites/ctl.mit.edu/files/attachments/Josue%20-%20Beer%20Game%20Run%20%20Ex%20Ed%20January%202017%20sub.pdf}
\textsuperscript{58} \url{https://beergameapp.firebaseapp.com/}
\textsuperscript{59} \url{http://web.mit.edu/jsterman/www/SDG/beergame.html}
\textsuperscript{60} \url{http://www.beergame.org/}
\textsuperscript{61} \url{https://www.theseus.fi/bitstream/handle/10024/132787/Thesis_Bohne_DM%20and%20IoT.pdf?sequence=1}
\textsuperscript{62} \url{https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-smart-sensors.pdf}
For the PoC, we will focus on two types of data [a] water/soil sensor data repository and [b] waste water treatment repository. We will now address the various characteristics of data in these repositories and explore how to select key features.

For performance characteristics of sensors, the features (dimensions) germane to SARS are:

- Sensitivity (output per input)
- Selectivity (accuracy, discrimination in complex matrix)
- Limit of detection (minimum number of detectable targets)
- Response time (time to 95% pulse input per IUPAC)
- Hysteresis/reusability

Features for wastewater treatment data repository may include the following characteristics:

- Conversion rate (output per input)
- Selectivity (related to inactivation)
- Loading rate (max number targets inactivated / volume of water)
- Kinetics of inactivation (kd, μ_max and Ks)
- Hysteresis/reusability

The characteristics of the 2 types of data, are likely to be a small sub-set. The number of dimensions (from which we will extract contextually relevant features) may be reasonable. There may be more dimensions (more important features) related to the actual measurement data. In future versions of the SARS tool, when the number of dimensions increase with the number of different types of sensor data, an initial reduction of dimension may be performed (filter). The value of data is contextual. Context of data is paramount during feature selection. Relevancy and redundancy of the variables/dimensions, which makes it suitable or unsuitable, to create combinations of features, may be sorted using a portfolio of algorithms. The ability to create combinations, enhances the potential to meaningfully merge data.

![Figure 5: Feature selection (left panel) helps to generate data format (right panel). From ref 64.](http://www.bioline.org.br/pdf?se11003)

![Figure 5: Feature selection (left panel) helps to generate data format (right panel). From ref 64.](http://www.iri.upc.edu/files/scidoc/2071-Sensor-Placement-for-Classifer-Based-Leak-Localization-in-Water-Distribution-Networks-using-Hybrid-Feature-Selection.pdf)

![Figure 5: Feature selection (left panel) helps to generate data format (right panel). From ref 64.](https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19870013024.pdf)
Feature selection is central to data handling (a short list of selected features will be required for the PoC). Using these features, simulation under different conditions is expected to generate data, organized in a particular data format (Fig 5, right). If data from other sources, repositories and projects, are to be merged, to create the collective view, then external data must be able to work with the data format of the tool. Semantics may play a substantial role in data merger between classes of sensor networks.

Because we may not standardize sensor manufacturing, we use feature selection to extract the data. It does not matter who made the sensor. As long as the data output from the sensors can fit the relevant features (e.g., the SARS tool), the tool can use diverse sources of data, generated by a plethora of different sensors. In each instance, we are seeking selected types of data, which can feed the features that are selected by the SARS tool, or any other future tool.

However, data types and data schemas can vary widely even between data from similar types of sensors. For example, waveforms and time stamps. In the 3rd dimension of the data format illustration (Figure 5, right), the time samples may differ in the time elapsed, between measurements. If data sampling is set to every 10 seconds, 1 min or 10 min, the time series representation of data may prevent merger or data fusion, unless the units are adjusted.

Software Tools for Data

Data level granularity, even for what if scenario planning must proceed with clarity about the question we are asking, or user-centric queries (uses data to answer). The quality, specificity and value of any answer, depend on data which is acquired by design, understanding dependencies between data and the influence of external or outlier events. The scenario outlined below is an example, to provide clues, with respect to the architecture necessary for future SARS tool. The imagination about the “big picture” must be in focus to help design SARS.

---

**Scenario**

A plot of land is instrumented with multiple different sensors (water, temperature, air pressure, soil chemistry). Similar instrumentation is replicated on other plots of land. Sensors sourced from different manufacturers (multiple manufacturers make soil, temperature and mercury sensors). Metadata about plots of land (crop type, acreage, soil composition) is available. Rainfall/drought data is available. SARS tool to build infrastructure to store select (disparate) data and answer ad hoc user directed queries using a mobile app: Can heavy rainfall explain elevated levels of mercury in the water?

---

66 [https://www.w3.org/2005/Incubator/ssn/wiki/Agriculture_Meteorology_Sensor_Network](https://www.w3.org/2005/Incubator/ssn/wiki/Agriculture_Meteorology_Sensor_Network)
67 [https://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/](https://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/)
69 [https://www.mdpi.com/1424-8220/9/10/7771](https://www.mdpi.com/1424-8220/9/10/7771)
The following is an incomplete list of concerns, software suggestions, and data issues, to be synthesized, when designing the foundation for SARS platform. The concerns are as follows:

[a] Time assurance of sensor data (variable). Limited bandwidth or transmission cost are reasons to compress\textsuperscript{70} data\textsuperscript{71} or skip (repeat/flatline data) values or aggregate at the edge.
[b] Geotagging – explicit (observation offers lat/lon) vs implicit (insert coordinates)
[c] On-chip memory to store data or “one shot” upload? fault tolerance? speed?
[d] Data upload interfaces (802.X protocols, LPWAN, LTE), gateways, cloud storage

Architecture Notes:

[a] Time Series DB: Storage of ingested sensor data in TSDB is increasingly\textsuperscript{72} popular (InfluxDB\textsuperscript{73} or Prometheus, open\textsuperscript{74} source). Optimization for time-ordered data reduces storage footprint (sensors often flatline, transmit the same value for some period). Non-TSDB stores each data point but TSDB only stores the change. Anomaly detection must become a routine.

[b] Specialized Databases: Applications must deal with variety of data, including data embedded in text (environmental inspection notes). Open\textsuperscript{75} Elasticsearch\textsuperscript{76} with support for search primitives (topic modeling, synonym resolution, fuzzy search) and associated open services\textsuperscript{77} are recommended. RDBMS may not answer "How often after heavy rainfall are mercury concentrations elevated?" or geospatial\textsuperscript{78} queries\textsuperscript{79} “Are there mercury levels above 100ppb within a 10km radius of “this” plot of land?” NoSQL\textsuperscript{80} databases, (eg, MongoDB\textsuperscript{81} or Elasticsearch) offer flexible and expressive query languages, do not require defined schemas.

[c] RDBMS: Schemas must be defined before adding data. SQL needs to know what you are storing in advance. Schema guarantee offers advantage (compared to flexibility of NoSQL) when the query is complicated ("sensor readings from corn or soybean plots within 5 hours after a lightning strike that is within 3 kilometers of the center of the plot") and RDBMS SQL\textsuperscript{82} offers flexibility\textsuperscript{83} for unknown query patterns. PostgreSQL\textsuperscript{84} is a credible\textsuperscript{85} and robust option.

\textsuperscript{70} https://www.itworldcanada.com/article/canadian-telematics-firm-geotab-doubles-down-on-data-services-as-it-finds-u-s-growth/408647
\textsuperscript{71} https://www.nrel.gov/docs/fy18osti/70223.pdf
\textsuperscript{72} https://db-engines.com/en/ranking_categories
\textsuperscript{73} https://www.influxdata.com/time-series-database/
\textsuperscript{74} http://opentsdb.net/
\textsuperscript{75} https://qbox.io/blog/what-is-elasticsearch
\textsuperscript{76} https://www.elastic.co/
\textsuperscript{77} https://aws.amazon.com/elasticsearch-service/
\textsuperscript{78} http://toblerity.org/shapely/manual.html
\textsuperscript{79} https://www.omnisci.com/
\textsuperscript{80} http://nosql-database.org/
\textsuperscript{81} https://www.mongodb.com/nosql-explained
\textsuperscript{83} https://pdfs.semanticscholar.org/3635/12f3dc6659d05eba4b6f07339f8542bd2737.pdf
\textsuperscript{84} https://www.postgresql.org/
\textsuperscript{85} https://www.csail.mit.edu/person/michael-stonebraker
Data and Other Issues:

To enhance performance and automate queries, data curation may be necessary. At a simpler level, pre-computing combinable representations of the data in the system may help: [i] converting units of measure, [ii] aligning time series data to same resolution (standardizing on n-hour means, percentiles of all observations), [iii] enriching ingested data with additional metadata describing its content (eg, Einstein AutoML system) [iv] aligning geospatial data to one or more known grids (eg, Uber's H3).

At the architecture level, separating the data persistence (store) and retrieval (query) functions of databases are recommended. Common data stores include open source distributed file systems - HDFS, GlusterFS and Amazon S3 (built for sensors and IoT). These can be paired with open analytics/query layers for which the underlying storage is swappable (Apache Spark, Apache Drill, Impala and Presto). This “divide and conquer” approach is based on the observation that far fewer professionals understand statistics and database technology. But analytical code is portable across environments. Due to surge in data science, more human resources may become available in this area (for example, statistical machine learning tools).

Figure 6: Asking correct questions may better define our goals and extract value from data.

---

86 https://github.com/salesforce/TransmogrifAI
87 https://engineering.salesforce.com/open-sourcing-transmogrifai-4e5d0e098da2
88 https://www.salesforce.com/video/1770953/
89 https://github.com/uber/h3
91 https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
92 https://docs.gluster.org/en/latest/
93 https://aws.amazon.com/s3/
94 https://spark.apache.org/
95 https://drill.apache.org/
96 https://impala.apache.org/
97 https://prestodb.io/
98 https://research.fb.com/category/machine-learning/
TRANSFORMATION OF SARS TO SARA - PARADIGM SHIFTS

Basic research\(^9\) in advanced\(^10\) material\(^11\) science may be the catalyst for forthcoming paradigm shift from SARS to SARA by using materials that not only sense but actuate\(^12\) and communicate. SARS may evolve to include and necessitate, for select applications, bidirectional communication, where the response, post-analysis, results in actuation, which may involve modifying the sensor for selectivity or sensitivity (SARA $\Rightarrow$ sense, analyze, response, actuate).

Using radio frequency (RF) signals to modulate molecules, such as DNA\(^13\), may be used in emerging sensing tools which uses aptamers. The structure-function\(^14\) relationship of DNA can be tuned by remote RF signals. With the emergence of the IoT\(^15\) concept formalized by Sanjay Sarma\(^16\) (the marketing\(^17\) term IoT was coined by Kevin Ashton, the concept of IoT has a rich\(^18\) history\(^19\)), the radio frequency identification (RFID) tool, the RFID tag, evolved as a target for disenfranchised individuals. Functional modification of RFID tags frequently used parasitic backscatter\(^20\) attack, a tool used by cryptanalysts (power analysis attacks). The EPC standard\(^21\) utilizes the remote mechanism to “kill” RFID tags using modulated backscatter. Sensing, actuation, transmission and communication has now converged with material science.

Convergence of material science, radio frequency (RF), mobile communications, nanobiotech, chemistry of biomolecules, biochemistry of infectious disease, and device systems engineering, may enable us to actuate\(^22\) functions and modify sensors using smartphones.

Figure 7: Smart materials can be monitored and actuated with a smart phone (ref 112).

9. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.044302
10. http://science.sciencemag.org/content/354/6317/1257
11. http://advances.sciencemag.org/content/4/8/eaat4634
12. http://science.sciencemag.org/content/347/6228/1261689
13. https://www.nature.com/articles/415152a
18. http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=34057
22. https://docs.wixstatic.com/ugd/6a9318_408ad8c2617847708e64cc3ba50d903a.pdf
The SARS platform tool and the future paradigm shift from SARS to SARA, depends on our ability to express our incisive foresight by asking and framing the **correct questions**. Success depends on defining our goals, validation of analytics, and the pragmatic drive to promote end-user adoption. The outcome will be shaped by the questions we ask and the evidence-based responses. Development of a “sense, analyze, response” system (SARS) as a platform and data tool for application in agriculture (water, soil, etc) is an important task at the nexus of food, energy, water, and sanitation (FEWS). The move from SARS to SARA may be a milestone, but SARA is not limited to any vertical. For this discussion, agriculture is an use-case (SARS-AG).

Figure 8: Paradigm of SARA illustrated in cybersecurity (phishing). Agents\textsuperscript{113} may sense content, find evidence of potential phishing fraud upon analysis, responds by actuating termination.

To substantiate ideas latent in SARS/SARA, as a first step, we advocate a PoC using limited volume of data related to water/soil sensors and wastewater treatment. The outcome of the PoC may be a web-based simulated model which may be disseminated to potential end-users through extension services and serve as an educational resource. By feeding the PoC with actual sensor data, the end-user (growers, farmers, researchers) can evaluate the PoC through a mobile app posting user-centric queries. The temporary functional limitation of the PoC does not imply functional limitation of the open infrastructure of the SARS platform. The modular architecture will be flexible, scalable and adaptable.

The thinking to guide this vision, must include scale, from conception. The compass to continuously map need-based innovation, must be fluid, not static. Future-ready architecture must overlay all computation and complexity with simple forms of knowledge representation, such as “drag & drop” icons\textsuperscript{114} and “traffic light” guidance (red, yellow, green) as the outcome. The complexity of this simplicity is a test of our collective erudition. It may stretch beyond the horizon of our collective imagination. Are we, then, going to be limited by our imagination?

\textsuperscript{113} https://www.technologyreview.com/s/608036/we-need-to-talk-about-the-power-of-ai-to-manipulate-humans/
\textsuperscript{114} http://bit.ly/TOPOLOGY-OPTIMIZATION
IMAGINATION – WHAT IF YOU ALLOWED IT TO WANDER

The discussion here, and elsewhere, about sensors, centers on the classical binding of an analyte. The pragmatism about reversible binding, in order to re-use the sensor, is a distant laggard. Even if re-usability was a possibility, concomitant calibration of the sensor, in the field, may be challenging, especially for low-cost sensors. Decreasing efficacy (sensitivity, selectivity) may tarnish data acquired from re-usable sensors, unless robust re-calibration was verifiable.

What if we pursued sensors that do not bind to an analyte? One alternative medium may be waves, in particular, radio waves, all around us. Can we use RF waves as sensors?

Figure 9: Dina Katabi\textsuperscript{115} and Fadel Adib\textsuperscript{116} and others have demonstrated how reflection of RF waves\textsuperscript{117} may be used to detect objects behind opaque barriers. Using the ability of frequency modulated carrier waves (FMCW, a radar technique) to separate reflections from different objects, it is possible to detect\textsuperscript{118} with accuracy, respiration and heartbeat from individuals. Applications\textsuperscript{119} of this principle (reflection of RF waves) is only limited by our imagination. Panel (right) shows how WiFi passive radar (#2) may assist real-time decision support\textsuperscript{120} for workers.

Detecting chemical and biological agents in water and soil presents grave challenges. Foremost, attenuation of RF signals and perhaps in second place, the damping or disturbance due to other “similar” molecules which may add significant noise to the reflected wave/signal. Molecular detection using IR/NMR/GCMS may inform this exploration, in terms of the features which may be extracted from signals (reflected waves?). Feature-related data may train new\textsuperscript{121} algorithms, and use machine learning\textsuperscript{122} techniques to detect “needles” from many haystacks!

\textsuperscript{115} http://people.csail.mit.edu/dina/
\textsuperscript{116} http://www.mit.edu/~fadel/
\textsuperscript{117} https://people.csail.mit.edu/fadel/papers/wivi-paper.pdf
\textsuperscript{118} http://www.mit.edu/~fadel/papers/vitalradio-paper.pdf
\textsuperscript{119} http://openaccess.thecvf.com/content_cvpr_2018/CameraReady/2406.pdf
\textsuperscript{120} https://pure.coventry.ac.uk/ws/portalfiles/portal/7646003/tancomb.pdf
\textsuperscript{121} http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.888&rep=rep1&type=pdf
\textsuperscript{122} http://bit.ly/BOOKS-ML
As a lay person, it is impossible to suggest tools from the vast field of waves/spectrum (electromagnetic, matter) and related (radar, sonar) techniques, but one may dare. It is quite encouraging to note that we can identify between Coke and Pepsi using RF\textsuperscript{123} signals. Using UWB (ultrawideband) radios, we differentiate between liquids using permittivity\textsuperscript{124} of liquids. RF waves travel slower in soil with higher permittivity. By measuring\textsuperscript{125} soil permittivity, we can extrapolate soil moisture and soil conductivity, to aid data-driven smart farming (IoT).

What about molecules? Consider non-destructive testing (NDT) techniques used for sophisticated aerospace components\textsuperscript{126} as well as simple construction material\textsuperscript{127} (Fig 10L). Now imagine the combination of NDT with ISAR (inverse synthetic aperture radar, Fig 10R).

The sensor “housing” may find inspiration from the NDT illustration (Fig 10L) if we imagine that the “substance” (soil, water) sits in the ‘sample holder’ (Fig 10L, illustrates the material under testing, as a flat cartoon, located on the ‘sample holder’ platform).

The sensor “detection” mechanism may find inspiration from the use of ISAR (Fig 10R). In a conventional approach, an antenna array may be required to locate an object by steering its beam spatially (see (a) in Fig 10R). The authors inverted this concept and used the motion of the object (top part, (b), Fig 10R) to emulate an antenna array. Hence, the creative application of inverse synthetic aperture (ISAR). Wi-Vi\textsuperscript{116} leverages this principle to beamform the received signal in time (rather than in space) and locate the moving object.

For water and soil sensors, is it even rational to form a hypothesis, that this “moving object” in the context of Wi-Vi, may be analogous to the “moving analyte” in soil, water, or air? Can we even extrapolate the concept of “motion” (Fig 10R) at the molecular (micro, nano) level, necessary for detection of analytes? The assumption is that the intrinsic motion of molecules (random, Brownian) can be used (substituted?) in the context of the “motion” explored in the Wi-Vi paper (reference 116). The kinetics of molecular motion must be accounted in the design of the detector, but are these motions (motion of object vs motion of molecules) comparable?

Figure 10: Rational basis for convergence of ideas? Can we combine NDT\textsuperscript{127} and ISAR\textsuperscript{117} types of tools & techniques to create sensors, based on differential reflection of waves (RF, EM, sound)?

\textsuperscript{123} https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4878&context=sis_research
\textsuperscript{124} https://synrg.csl.illinois.edu/papers/liquid_mobisys2018.pdf
\textsuperscript{126} https://www.mdpi.com/1424-8220/18/2/609
\textsuperscript{127} https://www.sciencedirect.com/science/article/pii/S1876610217321689
UNPLUGGED, UNBOUND, UNTETHERED

The value and flexibility of “wireless” sensor networks (WSN) partially (often rapidly) evaporates, when the hardware (sensor) is tethered or is made immobile (Figure 11). Isn’t it almost oxymoronic to picture immobile RF sensors?

Fig 11: The conventional practice of immobilizing sensors may restrict their dynamic potential. It adds enormous burden of pollution when sensors are non-functional and not biodegradable.
With advances in RF sensing, the use of drones as a gateway\textsuperscript{128} for mobile sensing will increasingly rise in prominence, in the portfolio of data acquisition tools. Mobile sensing is certainly not a panacea, but it may help reduce pollution due to non-biodegradable sensor hardware. But, these advantages are at the mercy of energy systems and advances in portable energy must be feasible for large scale deployments.

Fig 12: The many advantages of UAVs as mobile sensor platforms and data acquisition tools may remain unrealized due to on-board energy limitations. Food, water, soil, sanitation and the economy, globally, may be traced to a single scientific/engineering bottleneck – energy.

Convergence of concepts, manifested through confluence of computation, will make itself useful, if we can harvest and invest in data and data analytics, to make sense of data. The infectious inclination to claim advances by power-point (Figure 13), which are in abundance (including this essay), is in sharp contrast to the actual/real availability of the real-time tools for simple decision support (Figure 14). It is easy to illustrate what “should/could be” (this essay) and “simulate” what looks like (Figure 14) but it is a herculean task to transform the vision into reality.

Transformation cannot stop at decision support. Detecting harmful levels of pathogens or metals in a sample, calls for treatment. To suggest treatment without implementing protocols, does not remedy the problem. We “polished the chrome” but did not “tune the engine” which helps with economic growth. In this essay, we discuss an interim step, decision support, as a haphazard compilation of amorphous ideas. The real work is on the road ahead.

\textsuperscript{128} https://www.mdpi.com/1424-8220/18/2/624
Fig 13: Data-driven intelligent information arbitrage has advanced by power-point\(^\text{129}\) but the real task of driving the data, and merging it to make sense, is the task of feature engineering\(^\text{130}\) which few can accomplish (and rarely seem to surface in august monographs, reference 129). The illustration (on the left) may take into consideration “discovery” processes\(^\text{131}\), include principles of dynamic model-based enterprise\(^\text{132}\) concepts and assume most data to be fuzzy, incomplete, unstructured, and punctuated. This data, devoid of schemas or structure, must be incorporated, in system dynamics models using agent systems\(^\text{133}\) which can update values (state, parameter), as and when, new data arrives, in streams or trickles.

Fig 14: Sensors\(^\text{134}\) to detect pathogens and transmit data to smartphones\(^\text{135}\) are initial steps to remedial action. All “ideas by power-point” must translate to mobile systems (right) but we still need to “move the needle” by implementing solutions. Knowledge of contaminants must not generate complacency. The task to equip users with solutions and remedies is not an academic role. Hence, the importance of public-private partnerships between academia and industry.

\(^{131}\) [https://dl.acm.org/citation.cfm?doid=2836075.2822529](https://dl.acm.org/citation.cfm?doid=2836075.2822529)
\(^{133}\) [https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-017-2726-9](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-017-2726-9)
\(^{134}\) [https://doi.org/10.1016/j.ebiom.2018.11.031](https://doi.org/10.1016/j.ebiom.2018.11.031)
\(^{135}\) [https://doi.org/10.1016/j.ebiom.2018.09.001](https://doi.org/10.1016/j.ebiom.2018.09.001)
The mobility of display (Fig 14) is not necessarily mobility of measurement or the sensor. Data from immobile sensors (Fig 11) on smartphones is mobile access. For mobility by design, remote sensing using reflected RF (Fig 10), is one solution. SMURF\textsuperscript{136} soil moisture sensor uses 2.4GHz WiFi band (Fig 16). NASA’s SMAP\textsuperscript{137} radar and radiometer observe the Earth’s surface at 1.2GHz and extrapolates soil moisture. If we probe the versatility of RF and its unlimited\textsuperscript{138} potential, we may find an embarrassment of riches. If we move beyond tools for soil moisture, we can shift our exploration from microwaves to sub-millimeter frequencies, terahertz. Higher frequencies may improve spatial resolution. Can remote sensing measure salts in the soil?

Fig 15: Sign posts of our spectrum: MHz (megahertz, $10^6$ Hz), GHz (microwaves are in gigahertz, $10^9$ Hz) and THz (sub-millimeter waves are in terahertz, $10^{12}$ Hz).

Fig 16: IEEE 802.11 standard protocol provides several distinct radio frequency ranges for use in WiFi communications: 900 MHz, 2.4 GHz, 3.6 GHz, 4.9 GHz, 5 GHz, 5.9 GHz, 60 GHz bands. Each range is divided into a multitude of channels\textsuperscript{139} (illustrated here for the 2.4GHz band).

\textsuperscript{137} https://smap.jpl.nasa.gov/faq/
\textsuperscript{138} https://doi.org/10.1038/s41928-018-0186-x
\textsuperscript{139} https://upload.wikimedia.org/wikipedia/commons/8/8c/2.4_GHz_Wi-Fi_channels_%28802.11b%2Cg_WLAN%29.svg
Terahertz (THz) sensing and advances in terahertz integrated hybrid electronic-photonic systems\textsuperscript{140} may lead to new dimensions in remote sensing, including, potentially, reflected RF at THz frequencies. It is merely speculation at this point, but the prominent question is whether algorithms can be trained by machine learning techniques, to differentiate between different soil compositions based on data from reflected RF waves. It is encouraging that this approach works for microwaves (WiFi, Fig 9) to differentiate between heart rate and respiratory rate in humans from reflected RF signals. Will it work reliably for terahertz, sub-millimeter waves?

The sensor housing, and acquisition of the signal, may be made mobile-by-design using UAV mounted sensors with tiny TSDB (time series database) embedded in SDR (software\textsuperscript{141} defined radio) transceivers for collecting reflected RF data (ultrawideband transceivers). Algorithms and ML techniques in data analytics may differentiate the signals and extract actionable data and/or information. Due to damping of THz frequencies, error correction routines will be required to improve signal to noise ratio (for example, Kalman filter, Shannon noisy channel). Deploying a swarm of UAVs will require collision avoidance, path optimization, energy minimization and intrusion detection (cybersecurity). Application of swarm intelligence and the principles of collaborative mobile robotics are expected to be integral to this process.

Continuing the speculation to ask even more from remote sensing, the discussion about sensors in the soil may progress from soil moisture to chemical ingredients in the soil (nitrogen, phosphorous) and then to the biological activity in the soil due to nitrogen-fixing bacteria. One idea is to explore whether we can take advantage of the paramagnetic vs diamagnetic nature of the key elements, oxygen and nitrogen. Measuring biological nitrogen fixation using remote sensors may serve as an index or indicator of root health, plant physiology and food production.

![Biological Nitrogen Fixation](https://www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419)

Fig 17: Nitrogen fixation (Beijerinck, 1901), is carried out by a special group of prokaryotes.\textsuperscript{142}

---

\textsuperscript{140} https://www.nature.com/articles/s41928-018-0173-2
\textsuperscript{141} https://dspace.mit.edu/handle/1721.1/9134
\textsuperscript{142} https://www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419
The central metalloenzyme for nitrogen fixation is a dimeric protein, nitrogenase, with a moiety of Molybdenum (Mo) and Iron (Fe), one for each of its monomers, but with minor exceptions where Vanadium (V) is the metal of choice.

The suggestion here is to exploit the differential between the rapid inactivation of nitrogenase by atmospheric concentrations of oxygen (even an oxygen concentration as low as 57nM within a soybean nodule can reduce nitrogenase activity) versus the demand for oxygen for ATP synthesis which is essential to energize the nitrogenase. These competing needs are met by modulating oxygen concentration via an oxygen diffusion barrier in the nodule, by use of leghemoglobin to buffer oxygen concentration and by a class of bacteroid cytochrome c oxidases with a high affinity for oxygen (in a classic feedback loop, low oxygen signals the transcriptional activation of nif and fix genes in nitrogen fixing bacteria, eg, Rhizobium sp).

The recent focus on leghemoglobin may increase the inclination to develop tools to determine its concentration using the principles of spectroscopy (absorption spectra of human hemoglobin) which were worked out almost half century ago and patented a quarter century ago. Modifying non-invasive hemoglobin measurement tools may be useful but some of the tools available may not be reliable or suitable for leghemoglobin measurements.

Since the differential in this scenario is an understanding of the oxygen concentration (or its gradient) and its modulator, leghemoglobin, it is imperative that we measure oxygen or oxygenated compounds. In addition to the potential use of terahertz signaling, paramagnetic property of oxygen makes it possible to use robust tools, such as, electron spin resonance which is increasingly popular as electron paramagnetic resonance (EPR) spectroscopy and used for a variety of purposes. I don’t know if EPR spectroscopy can function in a form factor to be an oxygen sensor and transported via UAVs. Understanding which questions to ask about the type of data necessary, remains the most important issue in this complex scenario.
CONCLUDING COMMENTS

In one segment of this meandering essay, I discussed SIP-SAR, a synergistic integration platform for machine intelligence, which can better support decision systems, the response part of the SARS paradigm (sense, analyze, response systems).

By now, any astute reader will have made the connection between SARS and PEAS\textsuperscript{161} which is an easy to remember mnemonic describing Agent action (Performance metric, Environment, Actuators, Sensors). SARS is a partial inverse of PEAS where we start with the data to drive toward a data-driven decision system rather than a goal driven system where we optimize to reach the performance metric. In the real world, SARS and PEAS are not conceptual compartments, they are dynamic push-pull elements, which may be used in any combination.

The reference to digital twin is supposed to be a generalizing approach, because a “twin” can be made of anything, a machine, a sensor, a greenhouse, an orchard, a capnometer. In this respect, SIP-SARS is not about machine intelligence but about the data of objects. The objects may be real, with a virtual representation, hence, the digital twin concept. All the data elements from Figure 18 may be synthesized in a digital twin for urban water management or the illustration may be modified to make it relevant to water for irrigation or farm use. The system data may be accessible in real-time on smartphones for users and city managers.

The familiarity of internet of things (IoT) makes it easier to understand the meaning of *internet of objects* and the ideas related to “digital objects” which are at least 25 years old. If data about objects are in this repository\(^\text{162}\) then the *internet of objects* may be also referred to as the *internet of data*.

In the second part of this essay, SARS-AG, the digital thread continues as data thread. The suggestion that SARS may evolve to SARA is based on data-driven actuation of function. These paradigms, and their shifts, may be traced back to data, and the connectivity between data sources and sinks.

In the accompanying presentation, the SIP-SARS-SARA thinking seeks to merge with the conceptual OODA\(^\text{163}\) framework and tangles with the six (0-5) levels of autonomy\(^\text{164}\) highlighted in automotive automation, and occasionally used as a management tool\(^\text{165}\) in decision making. The common denominator in all these acronyms and frameworks is data.

Data acquisition tools (hardware, telecommunications, software, services, sensors) and data analytics (machine learning, AI, statistical techniques) enables us to extract information from data (if there is information in the data). This information may be of value to the end user, if its delivered, at the point of contact or point of use, perhaps in near real-time, before its value perishes. The role of mobility in this process is central, with respect to distributed data acquisition and data driven decision support at the edge (the latter is also a distributed system).

Therefore, any application, of real value, must engage with diverse industry groups, to synthesize the end-to-end components of the outcome, the user/customer/farmer demands.

The industry-academia partnership suggested in Figure 14 is not only between like-minded companies in the same business segment but must involve the value chain partners. In the context of machine tools (SIP-SAR) and applications related to food (SARS-AG), we must engage with behemoths and SMEs, including telecommunications, hardware, software, sensor manufacturers, services, and tool developers. The portfolio of solutions for FEWS (food, energy, water, soil) needs data, convergence of ideas, and implementations which are of value to users.

The last section of this essay are suggestions which may be partially or even completely incorrect because of my lack of relevant depth of knowledge. My thinking involves a quantum leap where one leg is based on the knowledge that in neonatal intensive care units (NICU’s), pre-mature babies are sensitive to oxygen concentration. NICU’s continuously monitor oxygen using the paramagnetic vs diamagnetic differential (between oxygen and nitrogen in air). I have jumped from that knowledge to connect with other dots. It remains to be seen whether parts of that principle can be applied to remotely sense the nitrogen fixation status of soil microbes.

---

\(^{162}\) [https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf](https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf)

\(^{163}\) [https://apps.dtic.mil/dtic/tr/fulltext/u2/a590672.pdf](https://apps.dtic.mil/dtic/tr/fulltext/u2/a590672.pdf)

\(^{164}\) [https://journals.sagepub.com/doi/abs/10.1177/1555343417726476](https://journals.sagepub.com/doi/abs/10.1177/1555343417726476)

\(^{165}\) [https://www.maths.tcd.ie/~nora/FT351-3/DSS.pdf](https://www.maths.tcd.ie/~nora/FT351-3/DSS.pdf)
In making this leap, I have relied on my feeble understanding of basic school chemistry. Classical molecular orbital theory (but not VSEPR) indicates oxygen will fill the orbitals following Pauli’s Exclusion Principle (sigma2s, sigma2s*, sigma2p, sigma2p*). Two electrons in 2p* will partially fill this orbital and possess parallel spins. Since the rest of the electrons are paired, the remaining two electrons in 2p* orbital gives the diatomic molecule a net total spin (it does not matter if they are +1/2 or -1/2 spins, they will both be the same). Since there is a net spin, oxygen is paramagnetic. I have assumed that these laws of physics governing the spin of oxygen and relevant properties of nitrogen, cannot change if they are in the NICU or plant root nodule.

The desire to usher mobile tools for continuous remote sensing is the basis for the purely speculative suggestion about combining ultrasound sensing on a chip166 with terahertz sensing167 with hybrid photonic-electronic systems which can take advantage of RF reflection168 to create new applications. The use of WiFi169 and NASA SMAP170 for soil sensing may suggest that a potential convergence, at least in principle, is in progress. The latter may be useful in remote sensing for diverse applications (human blood hemoglobin, plant leg hemoglobin, soil composition, security, metabolites, electromyography). If adapted for use with cows, internally and externally, remote sensing may offer robust return on investment, in the US (Figure 19).

Fig 19: US - the land of cows171 - may need remote sensing tools for external and in vivo use.

---

166 https://www.nature.com/articles/s41467-018-08038-4.pdf
167 https://www.nature.com/articles/s41928-018-0173-2
168 http://people.csail.mit.edu/fadel/wivi/
171 https://www.bloomberg.com/graphics/2018-us-land-use
Remote sensing principles are agnostic of applications. Specific tools may evolve as scaffolds for digital by design metaphor (IoT) by advancing the principles and practice of connectivity. When added to mobile carriers (e.g., swarms of micro-drones), these may evolve to become tools for mobile sensing, useful for sense, analysis, and response systems (SARS) in wide area applications (e.g., animal farming, agriculture, water, energy, sanitation, and public health).

If one wonders why we should invest in science, the simple answer is that of necessity. We do not know what effort will result in what outcome. We cannot leave any stone unturned. By the end of this century, the Earth may have to feed about 11 billion people. We need food.

---

**The Importance of Being Smart (is “smart” sufficient?)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1804</td>
<td>1 billion</td>
</tr>
<tr>
<td>1850</td>
<td>1.2 billion</td>
</tr>
<tr>
<td>1900</td>
<td>1.6 billion</td>
</tr>
<tr>
<td>1917</td>
<td>2 billion</td>
</tr>
<tr>
<td>1950</td>
<td>2.5 billion</td>
</tr>
<tr>
<td>1955</td>
<td>2.8 billion</td>
</tr>
<tr>
<td>1960</td>
<td>3 billion</td>
</tr>
<tr>
<td>1970</td>
<td>3.7 billion</td>
</tr>
<tr>
<td>1975</td>
<td>4 billion</td>
</tr>
<tr>
<td>1980</td>
<td>4.5 billion</td>
</tr>
<tr>
<td>1985</td>
<td>4.8 billion</td>
</tr>
<tr>
<td>1990</td>
<td>5 billion</td>
</tr>
<tr>
<td>1995</td>
<td>5.7 billion</td>
</tr>
<tr>
<td>1999</td>
<td>6 billion</td>
</tr>
<tr>
<td>2000</td>
<td>6.1 billion</td>
</tr>
<tr>
<td>2005</td>
<td>6.4 billion</td>
</tr>
<tr>
<td>2010</td>
<td>6.8 billion</td>
</tr>
<tr>
<td>2011</td>
<td>7 billion</td>
</tr>
<tr>
<td>2030</td>
<td>8.5 billion</td>
</tr>
<tr>
<td>2050</td>
<td>9.7 billion</td>
</tr>
<tr>
<td>2100</td>
<td>11.2 billion</td>
</tr>
</tbody>
</table>

- GPS Guidance / Semi-Autonomous
- Section Control (Sprayers)
- Row Control (Planters/Seeders)
- Yield Monitoring
- Remote Sensing
- In-field Sensing
- LiDAR
- Variable Rate Applications
- Telematics
- Robotics / UAV
- Data Analytics / Actuation
- Information Management Systems

---

Fig 20: Necessity is the mother of invention. The demand for food and economic development makes it imperative that we search deep within ourselves, and science, to better serve society.
SARS\textsuperscript{AG} - ANIMAL\textsuperscript{AG} CASE STUDY

DYNAMIC FEED FORMULATION, REAL-TIME NUTRITION, DIGESTIVE ANALYTICS & PRODUCTIVITY

INFLUENCE OF FEED ON FOOD - BIOCHEMISTRY AND METABOLOMICS OF THE BOLUS

This case starts with the adage that \textit{we are what we eat}\textsuperscript{172} and the vast plethora of efforts\textsuperscript{173} to bring home\textsuperscript{174} this message. In this section, we attempt to capture a few ideas and suggestions embedded in SARS\textsuperscript{AG} to propose an application for the animal ag industry, relative to the influence of feed on food production.

We start with the bolus, that is, the chewed food, at the moment of swallowing the feed\textsuperscript{175} mixture, the animal is eating (think about ruminants, cows). The biochemistry\textsuperscript{176} of processing this bolus and the kinetics of that process, defines the efficacy of the feed mix with respect to its nutritional content, digestion, excretion and contribution to food production from the animal, for example, quality and quantity of milk production.

Where do we start? Biochemistry of nutrition is very well investigated\textsuperscript{177} and needs little introduction or justification of its importance. We may address this case, as follows:

|-------------------------|-----------------------------|

In the biochemical lifecycle of a bolus, can we list [1] the key molecules/pathways which may serve, \textit{in sequence}, as molecular indicators (sugars, fats, acids) of progress (\textit{time series}), from ingestion of the feed to digested products, \textit{in vivo}, ready for absorption in the lumen.

In [2] we explore which molecules may serve as key performance indicators (KPI) in the context of engineering specific sensors. In other words, create sensor(s) to sense the rate of change (in real time?) of the indicator(s), with respect to the identified \textit{sequence} of steps.

In part [3] we focus on closed loop\textsuperscript{178} sensing, sensor engineering and sensor form and factor. The potential for use of soft robots\textsuperscript{179} and closed-loop sensing is key to wireless signal acquisition from \textit{in vivo} measurements. Calibration and \textit{in vivo re-calibration} of sensors are central to data reliability. Calibration and mechanism of sensing (binding, signal, reusability, saturation, sensitivity, mobility, degradation, excretion) are key areas for innovative R&D.

\textsuperscript{172} https://link.springer.com/content/pdf/10.3758%2Fs13423-015-0908-2.pdf
\textsuperscript{173} https://www.nap.edu/read/1365/chapter/4
\textsuperscript{174} https://www.cdc.gov/scienceambassador/documents/we-are-what-we-eat.pdf
\textsuperscript{175} https://cdn2.hubspot.net/hubfs/745395/2017%20Global%20Feed%20Survey%20(WEB)%20EDITED%20EM.pdf
\textsuperscript{176} http://bit.ly/BOLUS-BIOCHEM
\textsuperscript{177} https://www.springer.com/us/book/9781475713510
\textsuperscript{178} https://www.ncbi.nlm.nih.gov/pubmed/30503617
\textsuperscript{179} https://www.nsf.gov/awardsearch/showAward?AWD_ID=0938047

29 • Essay - Sense of Convergence - PEAS for Agricultural Information Systems. Shoumen Datta (shoumen@mit.edu)
In part [4], data capture, data analytics and data visualization on smartphones, are core elements. However, the most important issue in this vein is analytics. The meaning of the data is the single most critical outcome which generates value for the end-user.

Parameters in [2] are the bedrock of digestive dynamics and metabolomics. This is the most exhaustive segment which requires knowledge, wisdom and on-farm working experience to identify targets, without getting lost in the myriad of molecules. The ability to ask correct questions (in the context of this specific scenario) may reveal molecular signatures and/or clues to in vivo manifestations, due to viromes and microbiomes (phenotypes).

Predictive analytics [4] must deliver value for the user. If this data is aggregated and coupled with information from other farms (animals and species, anywhere), it may begin to reveal patterns or genetic footprints (genotypes), with respect to the food/feed ingested and in vivo signals. Information sharing in the “big data” context could be useful, too.

Using data and evidence, optimizing or re-formulating genotype-phenotype balanced bolus may be transformative for the future of the feed industry. The convergence of real-time data with the principles of nutrigenomics may influence the future of food productivity. Data and analytics [4] may aid future metabolic engineering of high methane animals. Different feed formulations may be less expensive than CRISPR-ization to reduce methane emissions.180

CREATING A “PLUG AND PLAY” TOOL FOR THE ANIMAL AG INDUSTRY

Visualization of in vivo nutritional kinetics data and analytics on smartphones is only information. The preferred outcome for the end-user is to use this information in near real-time to address the problem. For example, re-formulate the feed mix to lower the concentration of some of the volatile (short chain) fatty acids. When this re-formulated feed mix generates the required/expected result, then, this effort may be labeled as a success. The end user can count the outcome as a positive return on the investment (ROI) and may pay for such services.

A tool is necessary to deliver parts of this service, based on the data and information. To create this model (where input data will output instructions to change the feed mix), we need to understand the kinetics, in the lifecycle of a bolus, using metabolites as markers (sense) and metabolomic sign-posts. For each metabolite, we may identify (one or more) substrate/product and at least one attribute, which can be measured (sense) to determine the status (rate of formation) of the substrate or product.

Envision the “model” to be a lego-like modular software version of a fusion between a Rubik's Cube, a multi-dimensional matrix, and the cube-on-cube concept proposed by Marvin Minsky to demonstrate agent-based connectivity. Modularity of the software enables the user to access/buy only the necessary or affordable modules (Nissan Datsun vs Rolls-Royce Cullinan).

---

180 https://www.technologyreview.com/s/612458/exclusive-chinese-scientists-are-creating-crispr-babies/
181 http://www.acad.bg/ebook/ml/Society%20of%20Mind.pdf
The modular "blocks" in this model may represent ingredients of the bolus. The pathway from ingestion to digestion, that is, sequence of molecular intermediaries, will be ‘weighted’ according to [a] biochemical significance of the ingredient and [b] ease with which it can be measured or serve as a KPI for the process (think “weights” as in artificial neural networks).

Fig 21: The feed (bolus) is made up of several complex materials which must be digested in a series of biochemical steps to extract nutritional value for the animal. Time series data from measurement of intermediates, or activity of enzymes, are indicative of the progress of these reactions. In the segment [1] we explored these metabolic pathways and identified which molecules (intermediates, enzymes, co-factors) were suitable/amenable for sensing, and acquisition of the sensor data, to map the dynamics of the progress of events in vivo.

Clicking on the “blocks” signifying ingredients, may reveal the next visual layer of the drill-down (using the software) as a chain (left panel, Figure 21) on a graphic user interface (smartphone, tablet, laptop). The intermediates will “light” up if the molecule can be sensed. These “feed” blocks and certain combinations (depends on pre-formulated mix) may indicate pre-set outcomes, for example, feed mix A generates X cubic meter of methane and Y gallons of milk. How closely the factory-suggested outcome parallels the observed results for your cows?

The combination and proportion of ingredients may be guided by data from digestive analytics. This “feed mix” may no longer be the same for all cattle. Using this system, the user can mix and match feed ingredients and optimize for milk production or reduction of methane or tackle both, at any level of priority, the user chooses.

The user is not required to “know” the kinetics of digestive process or the metabolic interrelationships that determine the nutritional absorption of molecules from the lumen and the specific/desired outcome, for example, the quality and quantity of milk produced. The user will choose "blocks" or use “auto-select” feature to “drag and drop” the blocks (signifying ingredients). The user can simulate the outcome based on her/his choices and ask the system what will happen to methane and milk production given these set of combinations.
Based on the data embedded in the knowledge and logic segment of the analytical engine, the combination chosen by the user will generate simulated results. The user can use this “tool” to map every animal or groups of animals. User can “tweak” parameters/ingredients to optimize desired outcomes. The system can crowd source independent data modules and aggregate on the open source platform of this tool, to increase number of ingredient options or any other function or knowledge enhancement, related to nutrition and/or nutrigenomics.

The presentation "tool" may resemble a mobile application which may draw on some of the ideas in Lego MindStorm (where players may cooperate to mix and match codes, bits and functions, to create their desired outcome). The data, analytics and logic, is embedded in the icons, buttons and levels, with respect to the independent variables (eg ingredients) and dependent variables (eg milk production). Relationships between the scientific and nutritional parameters versus the social and economic operators, can be modeled using agent-based models which are dynamic compared to the Jay Forrester school of system dynamics. The logic matrices, knowledge convergence and information arbitrage, perhaps operating on remote open source cloud platforms, will be imported/exported (event-triggered software defined networking) but “invisible” to the user, yet, subject to user exits to control access, authorization, privacy, and cybersecurity.

SUMMARY

This case is focused on feed and its influence on productivity. The kinetics of digestion and the combined analytics of feed vs food production (milk) is an oversimplification. The real world scenario must account for microbiomes and viromes, which can alter metabolic rates and outcomes. Implementation of this mobile tool is viewed as a software app. Users choose “lego blocks” to create or choose from a “menu” of feed, to drive their preferred outcomes, with respect to a temporary optimized state, and hopefully, maximize production of food.

---

182 https://www.nature.com/subjects/nutrigenomics
183 https://www.lego.com/en-us/mindstorms
186 https://peerj.com/articles/5012.pdf
187 https://pdfs.semanticscholar.org/286e/6e25a244d715cf697cc0a1c0e8f81ec88fbc.pdf
188 https://executive.mit.edu/faculty/profile/11-jay-forrester
189 https://jci.org/articles/view/94601/pdf
Causation – Correlation
Shoumen Palit Austin Datta, Massachusetts Institute of Technology and Massachusetts General Hospital, Harvard Medical School, Cambridge, Massachusetts 02139

If simulated\(^1\), system dynamics\(^2\) and agent based\(^4\) models\(^5\) may allow us to explore\(^6\) a range of impact\(^7\) of different variables\(^8\) on specific\(^9\) outcomes\(^10\). By altering parameters and metrics which can shape the outcome from these simulations\(^11\), we may identify to what degree the variables may influence the outcome, independently and/or collectively. The co-dependencies\(^12\) and inter-relationships\(^13\) between variables are central, and may become obvious, if models attempt to isolate and test the impact of variables, which cannot be isolated, in reality, due to functional inter-relationships (for example, what if sales and supply\(^14\) were separated). The latter is also applicable to human behavior\(^15\) related scenarios.

To transform the output of these tools\(^16\) to deliver meaning, resilience\(^17\) and relevance\(^18\) in the real world, the simulated data (use of agents\(^19\)) from pareto-optimal outcomes, may be targets which development economics must aspire to attain. For example, to reduce morbidity from diarrhoea by 10% we must increase availability of water by 20% and supply of cooking gas by 5%. On one hand, these numbers (outcome) may guide local municipalities or NGOs or development teams, to catalyze this change through capacity building and/or assess the risk\(^20\) of economic\(^21\) loss if the situation is unattended. On the other hand, data-driven\(^22\) outcomes driving decision support\(^23\) tools may guide policy\(^24\) and loan managers, in global financial\(^25\) institutions, to enable\(^26\) financial structures necessary to deliver global\(^27\) public goods. Convergence of concepts in physical and digital connectivity\(^28\), classical operations research\(^29\) with educational use\(^30\) of simulation tools may promote systems thinking\(^31\), which is essential for science\(^32\) to better serve\(^33\) society\(^34\). Understanding causation\(^35\) and true correlation is necessary to avoid hype\(^36\). Our ability to use these tools are limited, yet necessary for medicine, farming and optimizing milk production.

---

\(^1\) https://ncaef.me/loopy/
\(^2\) https://www.systemdynamics.org/assets/conferences/2004/SDS_2004/PAPERS/381BORSH.pdf
\(^3\) https://www.tandfonline.com/doi/abs/10.1080/01441647.2012.745632
\(^5\) http://www.thwink.org/sustain/glossary/SystemDynamics.htm
\(^6\) https://www.techrepublic.com/article/systems-thinking-with-a-chromebook/
\(^7\) https://tangible.media.mit.edu/project/airportsim/
\(^8\) http://vensim.com/vensim-software/
\(^9\) https://doi.org/10.1016/j.phpro.2012.03.263
\(^10\) https://afpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=310977
\(^11\) https://www.systemdynamics.org/tools
\(^12\) https://link.springer.com/content/pdf/10.1007%2F978-0-387-74157-4_45.pdf
\(^13\) https://thesystemthinker.com/from-spreadsheets-to-system-dynamics-models/
\(^14\) https://tangible.media.mit.edu/project/tangible-business-process-analyzer/
\(^15\) http://dx.doi.org/10.1016/j.jenvman.2017.04.036
\(^16\) http://sysdyn.simantics.org/
\(^17\) http://web.mit.edu/signup/repository/Rice_SCResp_Article_SCRM.pdf
\(^18\) https://rras.fad.harvard.edu/pages/change-control
\(^19\) https://dspace.mit.edu/handle/1721.1/41914
\(^21\) https://doi.org/10.1007/s13753-017-0154-5
\(^22\) http://science.sciencemag.org/content/359/6373/325
\(^23\) https://www.systemdynamics.org/tools
\(^24\) https://issues.org/esty-2/
\(^25\) https://www.brettonwoodsproject.org/2005/08/art-320747/
\(^28\) https://dx.doi.org/10.1109/JIOT.2017.2755620
\(^29\) http://www.bookmetrix.com/detail/book/7153d969-09a3-4211-a6cf-34588ef367e9#reviews
\(^30\) http://supplychain.mit.edu/supply-chain-games/beed-game/
\(^31\) http://www.sfu.ca/~vdabbagh/Forrester68.pdf
\(^32\) https://theconversation.com/marie-curie-and-her-x-ray-vehicles-contribution-to-world-war-i-battlefield-medicine-83941
\(^33\) http://science.sciencemag.org/content/295/5557/929
\(^35\) http://links.jstor.org/sici?sici=0012-9682%28196909%29297%3A3%3C3A%3CICRBEM%3E2.0.CO%3B2-L
\(^36\) https://www.amazon.com/Freakonomics-Economist-Explores-Hidden-Everything/dp/0060731338
Principles and Practice of Platforms and Paradigms
The importance of being smart (is "smart" sufficient?)

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1804</td>
<td>1 billion</td>
</tr>
<tr>
<td>1850</td>
<td>1.2 billion</td>
</tr>
<tr>
<td>1900</td>
<td>1.6 billion</td>
</tr>
<tr>
<td>1927</td>
<td>2 billion</td>
</tr>
<tr>
<td>1950</td>
<td>2.5 billion</td>
</tr>
<tr>
<td>1955</td>
<td>2.8 billion</td>
</tr>
<tr>
<td>1960</td>
<td>3 billion</td>
</tr>
<tr>
<td>1965</td>
<td>3.3 billion</td>
</tr>
<tr>
<td>1970</td>
<td>3.7 billion</td>
</tr>
<tr>
<td>1975</td>
<td>4 billion</td>
</tr>
<tr>
<td>1980</td>
<td>4.5 billion</td>
</tr>
<tr>
<td>1985</td>
<td>4.8 billion</td>
</tr>
<tr>
<td>1987</td>
<td>5 billion</td>
</tr>
<tr>
<td>1990</td>
<td>5.3 billion</td>
</tr>
<tr>
<td>1995</td>
<td>5.7 billion</td>
</tr>
<tr>
<td>1999</td>
<td>6 billion</td>
</tr>
<tr>
<td>2000</td>
<td>6.1 billion</td>
</tr>
<tr>
<td>2005</td>
<td>6.4 billion</td>
</tr>
<tr>
<td>2010</td>
<td>6.8 billion</td>
</tr>
<tr>
<td>2011</td>
<td>7 billion</td>
</tr>
<tr>
<td>2030</td>
<td>8.5 billion</td>
</tr>
<tr>
<td>2050</td>
<td>9.7 billion</td>
</tr>
<tr>
<td>2100</td>
<td>11.2 billion</td>
</tr>
</tbody>
</table>

- GPS Guidance / Semi-Autonomous
- Section Control (Sprayers)
- Row Control (Planters/Seeders)
- Yield Monitoring
- Remote Sensing
- In-field Sensing
- LiDAR
- Variable Rate Applications
- Telematics
- Robotics / UAV
- Data Analytics / Actuation
- Information Management Systems

Cause and Effect
SENSE
SENSE
BINDING OF ANALYTE

ANALYZE

TRANSMIT

DECISION SUPPORT

CONVERGE

SYNERGIZE
Don’t bind. Reflect
Conventional Wisdom
Scientists at GE Global Research discovered that the nanostructures on the wing scales of Morpho butterflies have excellent sensing capabilities. They could allow them to build sensors that can detect heat and also as many as 1,000 different chemicals. Image: GE Global Research
SVM distinguishes gazelles, ostrich, trees and ground in Namibia, Africa.
Volvo has fitted some of its cars with sensors and software that can tell cyclists apart from other objects.
RESPONSE
Sense, Analyze, Response, Systems

SARS
Paradigm Shifts
Smart materials can be monitored and actuated with a smart phone.
Amazon Warehouse – Smart materials can be monitored and actuated with a smart phone.
Shifting Paradigms

SARS  Sense, Analyze, Response, Actuate  SARA

AQUA SENSE

H₂O
Shifting Paradigms

Sensor Nodes + SDK/ API + Gateway + Cloud Connection + Web Frontend/Apps

Sensor Nodes + SDK/ API + Gateway + Cloud Connection + Web Frontend/Apps
Smart nanomaterials meet smart phones

Eric S. McLamore\textsuperscript{1} and Carmen Gomes\textsuperscript{2}
Agricultural and Biological Engineering, University of Florida
Mechanical Engineering, Iowa State University
NSF Project No. 1511953 (Nanobiosensing program)
www.mclamorelab.com

SARS 2 SARA

https://docs.wixstatic.com/ugd/6a9318_408ad8c2617847708e64cc3ba50d903a.pdf
What fuels these shifts?
If genius has any common denominator, I would propose breadth of interest and the ability to construct fruitful analogies between fields.

My talent is making connections. That's why I'm an essayist.

Connectivity

September 10, 1941 – May 20, 2002

in Darwin's Middle Road
Complementarity

DNA Replication

Image adapted from: National Human Genome Research Institute.
Reflecting on the Edison Awards: Why It’s Important to Dream Big

On Thursday I attended a gala in New York where I had the tremendous honor of receiving the Edison Achievement Award, which celebrates leaders who have made significant contributions to innovation and whose efforts in the space are positively impacting the world. It was humbling to be mentioned in the same breath as Thomas Edison – the epitome of ingenuity – and past award winners like Elon Musk and Steve Jobs. I was so proud to accept this award both personally and on behalf of my Cisco family – we truly made it happen together.
We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices, wisely.
Digital-by-Design Metaphor

IoT

A Connected Sense of the Future
Point solutions

there aren’t any ...
Systems Integration

Waspmote Agriculture 2.0 Board by Libelium
Systems Integration

- Sensor Nodes
- SDK / API
- Gateway
- Cloud Connection
- Web Frontend / Apps
Synergistic Integration

Synergize appropriate threads to weave solution fabrics

Metaphor for decision platforms
Multi-component “solutions” which are connected and communicates using platforms for interoperability between sub-systems, protocols and interfaces

Synergistic Integration
Engineering at the Edge
Figure 1 | Schematic drawings and corresponding images of the GP-hybrid electrochemical devices and thermoresponsive drug delivery microneedles.
Figure 4 | Demonstration of the wearable diabetes monitoring and therapy system in vivo.
Platform System

- Sensing Technologies
- Software Applications
- Communications Systems
- Telematics, Positioning Technologies
- Hardware and Software Systems
- Data analytics Solutions

SMART FARMING
Platform Interoperability
WSN Sub-Systems

- Temperature Sensor
- Humidity Sensor
- Salinity Sensor

Diagram showing the connection of sensors to a base station/aggregator through RF and LAN connections.
Meshlium from Libelium is a multi-protocol router which works as the WSN gateway (radio interfaces for WiFi 2.4GHz, WiFi 5GHz, 3G/GPRS, Bluetooth (802.15.1), Xbee (ZigBee) and LoRaWAN (LPWAN), as well as GPS integration.)
Systems Approach

Gateway radio interfaces for WiFi (2.4GHz and 5GHz), GPRS, Bluetooth (802.15.1), ZigBee (Xbee), LoRaWAN (LPWAN), plus GPS
SYSTEMS SIMILARITIES
NOAA Hydrometeorology Testbed (HMT) Program (soil moisture networks)

Orientation of soil probes and surface meteorological observations used at a typical soil moisture station.

https://journals.ametsoc.org/doi/abs/10.1175/2010JTECHA1465.1
https://doi.org/10.1016/j.compag.2004.03.002
Things which aren’t Smart
Generic 'Big Data' Approach

Sensors -> Module 1 -> DB

Sensors -> Module 2 -> DB

Sensors -> Module 3 -> DB

INTERNET

Data Analysis

Browser

Smartphone
Mobile Radio Wave Sensors
Drone-based Data Collection

Immobile implanted sensors are a liability (waste) if non-functional. Drones mobilize sensors. Arm extends to sense, retracts. Drone moves to next location.
Gateway radio interfaces for WiFi (2.4GHz and 5GHz), GPRS, Bluetooth (802.15.1), ZigBee (Xbee), LoRaWAN (LPWAN), plus GPS.
Mobile ‘Big Data’ Approach

{events, aggregate}$_n$

ML, Deep Learning Tools

Cloud Mist Fog

VALUE OF INFORMATION TO END USERS & SYSTEM

Smartphone

- GPS
- Rate Gyro (internal)
- 3-axis Accel. (internal)
- Barometric Pressure (internal)
- Compass (internal)
- 1km Low-Speed Radio (802.15.4)
- High-Speed Radio (802.11) (internal)
- Front Facing Camera
- Thermal Sensor
- Sonar Ranging
- Down Facing Camera (under)
REFRESH ...
SENSE RF REFLECTION

TRANSMIT

ANALYZE

DECISION SUPPORT

Mobile RF Data Collection

Reflected Radio Waves
Mobile Sensors ➔ RF Data

Sensor Hardware
Pollution Prevention
**Mobile Sensors ➔ RF Data**

- **Peak Data Rate (Gbps)**
  - 20 Gbps
  - 10 Gbps
  - 1 Gbps
  - 10 Mbps
  - 1 Mbps
- **Reliability**
  - 99.999%
  - 99.99%
- **Latency (ms)**
  - 1 ms
  - 10 ms
  - 100 ms

**Legend**
- 4G
- 5G

**Network**
- Sustainable Data Rate (mbps)
- Many network slices for many industries
- One Network for many industries

**Jitter (packet delay variation)**
- Guaranteed
- Best Effort

**Positioning Accuracy (outdoor)**
- Reduced cost of connectivity
- Device density

**G - What a Difference at 60mph = 88 fps**
- Latency 100ms (3G) = 8.8’ = 106” = 268cm
- Latency 10ms (4G) = 0.88’ = 10.6” = 26.8cm
- Latency 1ms (5G) = .088’ = 1.06” = 2.68cm
Mobile Sensors
Terahertz Systems

https://doi.org/10.1038/s41928-018-0173-2
Mobile Sensors

Key Dependency ➔ Energy
Dajiang T16 Plant Protection UAV

T16 plant protection UAV adopts a modular design, carries high load, wide spray width, powerful hardware collaboration, AI engine technology and 3D job planning capabilities. T16 is a hexa-rotor mounted on foldable arms. It can spray up to 4.8 liters/minute of plant protection material (pesticides), fly waterproof (flies in rain), has obstacle avoidance and features RTK for precise spraying/monitoring. Radar system mounted below the drone enables specificity (sprays only when flying over treetops, and not in between trees, to reduce the wastage of plant protection).

https://dronedj.com/2018/12/14/dji-agriculture-drone-t16-china/
可根据地形起伏实现仿地飞行

T16可自动探测障碍物方向与距离

绕开障碍物后，T16可自主恢复作业任务

并生成三维飞行航线
Almost all objects (things, atoms) are controlled/actuated by data (bits)
Complex Event Processing Engine: Can ESPER Discover Data Patterns?

Streaming data
DataTurbine

High-speed high-volume real-time data streams

Event Stream Intelligence for real-time EDA

Event Stream connectors & adapters
Statements
POJOs
Output adapters

Esper engines
Event Query & Causality Pattern Language
Core container

Historical data

Esper: Lightweight ESP/CEP container

www.dataturbine.org; http://esper.codehaus.org
GRID (L) and (R) INTER GRID ROUTING

Start

Divide the area into grids

Deploy Soil and Environmental sensors, grid heads and sink node
Predictive Analytics

1. Start
2. Divide the area into grids
3. Deploy Soil and Environmental sensors, grid heads and sink node
4. If Grid Head?
   - Send Beacon Message to sensors in its grid
   - Receive beacon response
   - Receive and separate data
   - Aggregate data
5. If Normal Node?
   - Wait for beacon from grid head
   - Send beacon response
   - Sense data
   - Send data to grid
6. If Sink Node?
   - Separate the Data
   - Apply mining algorithm to extract knowledge
   - Feed the required parameter based on knowledge extracted
   - Apply predictive algorithm to predict disease
7. Stop

Atoms to Bits

My mobile phone in Boston can move a taxi in Beijing
Stretching the OODA Loop? Exploring an amalgam between SARS, SARA, OODA, low-level semi-autonomous activities and principles embedded in using a mobile phone to move a taxi.

SARSOODA

Sense

Analyse

Respond

Observe

Orient

Act

Decide
Intelligent Information Arbitrage, automation, services, consumers

Data is the glue ...

Satellites, UAVs, Automobiles, Agriculture, Finance

Comparing 4G and 5G

- Latency: 10 ms (4G) vs. <1 ms (5G)
- Data Traffic: 7.2 Exabytes/Month (4G) vs. 50 Exabytes/Month (2021) (5G)
- Peak Data Rates: 1 Gb/s (4G) vs. 20 Gb/s (5G)
- Available Spectrum: 3 GHz (4G) vs. 30 GHz (5G)
- Connection Density: 100 Thousand Connections/Km² (4G) vs. 1 Million Connections/Km² (5G)
Waste water treatment is actuated, if sensor data exceeds range...
Almost all objects (things, atoms) are controlled/actuated by data (bits)
Data as a Service

- SERVICE
- COM
- EDGE
- TIME
- CPS
- IoT
- ANALYTICS
- COGNITION
Water Contaminated by Mercury, Arsenic, Lead

Professor Eric McLamore, University of Florida • www.ncbi.nlm.nih.gov/pubmed/29629449
Water Contaminated by BACTERIA


Professor Eric McLamore, University of Florida ● www.ncbi.nlm.nih.gov/pubmed/29541704
Prior to running the support vector machine (SVM) algorithm, PCA (principal component analysis) was applied through singular value decomposition (SVD) to reduce 152 features to 2 principal components. PCA was used to reduce the dimension of 152 features in the raw EIS data to a two-dimensional principal components matrix. Depending on number of components to extract, full or randomized truncated SVD was used. To ensure general applicability across other application-specific biosensors, code screens were prepared for four types of SVM: kernels (linear, sigmoidal, radial basis function, polynomial) to identify which approach best segregates the training data.

Tuning of Gaussian radial base function (RBF) hyper-parameters (C and gamma) for chemo-sensory proteins (CSP) acetone interactions. Recombinant insect chemosensory proteins (CSP) derived from *Glossina morsitans* (Gmm, tsetse fly) were heterologously expressed and purified from *E. coli* hosts. Representative support vector machine (SVM) classification results for one training and testing set show the effects of parameters C and g in the output of the RBF kernels. Red and blue circles represent the baseline samples in training and testing sets; green and purple plus symbols represent the positive signals in training and testing sets. The background blue and red region indicated the classifier decision surface, where all data fall into the red region are predicted as positive.
Water Data Analytics: Value to user?

Water Data of Value 2 User

New reporting platform for black and smelly rivers

Black and smelly rivers can be reported directly to two government ministries through the Blue Map app, with 100% acceptance and response from the government, helping us work together to bring back blue skies and clear waters.

Data at the Edge
Decision Support at the Edge
Crowd-sourcing Data from the Edge
Sense-Analyze-Response-Systems
MOBILE PLATFORM
Smartphone-based pathogen diagnosis in urinary sepsis patients

DOI: https://doi.org/10.1016/j.ebiom.2018.09.001
DOI: https://doi.org/10.1016/j.ebiom.2018.11.031
Platform
Platform

Data
Mobile Platform
- Two Rail Cars -
Proof of Concept
- Platform -

Data
Synergistic Integration Platform
Proof of Concept / Mobile Platform
Data and analytics on mobile applications

Visualization of data + mobile app development
Integration between the data science cycle and the app development cycle

Systems Approach

agnostic of applications
Closed Loop CPS Model for Precision Agriculture
From 126 papers, 139 dairy sensor systems (2002-2012) were compared based on 4 levels above: (I) technique, (II) data interpretation, (III) integration of information, and (IV) decision making.

NONE included integration of sensor data with other farm data for decision support or automated decision making (Rutten et al). We need for sustainable dairy CPS. Mobile sensing, data analytics, controls and networking advances are ripe enough to develop CPS for precision animal agriculture.
Volatile (Short Chain) Fatty Acids: Methane Estimator

VFA Profile

Degradation rates of CHO pools vs. milk yield

Degradation rates

<table>
<thead>
<tr>
<th>Rate, %/h</th>
<th>Fast Pool Rate</th>
<th>Slow Pool Rate</th>
<th>Starch Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.5</td>
<td>21.3</td>
<td>21.9</td>
<td></td>
</tr>
<tr>
<td>4.55</td>
<td>4.79</td>
<td>4.95</td>
<td></td>
</tr>
<tr>
<td>10.4</td>
<td>11.3</td>
<td>12.4</td>
<td></td>
</tr>
</tbody>
</table>

In vitro digestion model (IFM)
Measuring Dynamic Digestion

Volatile (Short Chain) Fatty Acids: Methane Estimator

ALLTECH INC
Internet of Connected Cows – Sustainable Practices for Precision Animal Agriculture

Level 1. Wireless Body Sensor Network (WBSN)

Level 2. Mobile Ad Hoc Network (MANET)

Level 3. Repeaters and Base Station

Cloud Platform

Real-time Monitoring/Data Analysis

Closed-loop Feedback

GUT CRAWLER

SOFT ROBOT

SENSORS

Data Upload via on-site gateway WSN

PRECISION NUTRITION

Professor Richard Voyles, Purdue University
WHY WE NEED SYNERGISTIC SYSTEMS INTEGRATION?

Nutritional SARA ⇒ Real-Time Digestive Analytics
Nutritional ⇒ sense, analyze, respond, actuate

In vivo rates of digestion will enable precision mixed ration formulation and better dairy health to increase milk production.
Intelligent Information Arbitrage
Intelligent Information Arbitrage

Performance

Environment

Actuators

Sensors

RESPONSE
Data based learning

Capture, storage, analysis of data

Data collection IoT

environment

percepts

actions

agent

sensors

actuators
HARVEST and INVEST in UNSTRUCTURED DATA and (MERGE DATA IN CONTEXT FOR INTELLIGENT) DATA ANALYTICS
Intelligent Information Arbitrage

V2I - Vehicle-to-Infrastructure.
Alerts drivers to traffic lights, traffic congestion, road conditions, etc.
Due 2022.

V2D - Vehicle-to-Device.
Cars communicate with cyclists’ V2D device and vice versa.
Due 2018.

V2H - Vehicle-to-Home.
In emergencies vehicles will give power back to homes.
Due 2019.

V2G - Vehicle-to-Grid.
Electric cars will return electricity to the grid.
Due 2020.

V2V - Vehicle-to-Vehicle.
Alerts one vehicle to the presence of another. Cars “talk” using DSRC technology.
Due 2017.

V2P - Vehicle-to-Pedestrian.
Car communication with pedestrian with approaching alerts and vice versa.
Due 2018.

Enabled by 5G, high- or full-automation and enhanced V2X connectivity will start to mature around the mid-2020s. Vehicles will be completely aware not just of other cars, but of cyclists, pedestrians, traffic lights, and other infrastructure.

The Society of Automotive Engineers defines six levels of autonomy, ranging from Level 0 with full driver control to Level 5, where the AI driver equals human capabilities. www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf
Intelligent Information Arbitrage

Merge Unstructured Data – Extract Sense
Intelligent Information Arbitrage

Six Levels of Autonomy

https://blog.playment.io/autonomous-driving-levels-0-5-explained/
Intelligent Information Arbitrage: combine SARS, OODA and levels of autonomous decision making
A unique code is woven into the fabric material of the backpack given to each first-year student at MIT. Unlike a QR code, this fabric-based coding system is subtle to the eye but immediately recognizable by an app called AFFOA LOOKS. The owner can link her backpack to their mobile device and program it to display a song, a cause, or anything the owner wishes to share.

https://www.nature.com/articles/s41586-018-0390-x
https://looksapp.affoa.org/
http://go.affoa.org/
Data is the glue

Satellites, UAVs, Automobiles, Agriculture, Finance and everything else, needs, anomaly-free data

http://cucis.ece.northwestern.edu/projects/DMS/publications/AnomalyDetection.pdf
Intelligent Information Arbitrage from manufacturers to end users

Data Analytics is the gluon

use of data to generate information of value

Think this https://en.wikipedia.org/wiki/Gluon not that https://github.com/gluon-api/gluon-api/
Intelligent Information Arbitrage

Data Analytics may need AI

Why are these governments “discussing” AI strategies?
Intelligent Information Arbitrage
with help from AI & Neural Nets

Artificial neural networks (ANN) imitate humans - they can learn to translate languages, predict sales, and categorize text and images. One ANN imitates any kind of text. I’ve given them paint colors, band names, and even guinea pig names and in each case their results are mixed. The problem is that it doesn’t know what any of these words mean - it’s just picking letter combinations. This is what happened when I gave recipe for cookies. This is what human cookies sound to neural networks. Can we trust ANN to make decisions for us?

Intelligent Information Arbitrage
Elusive Quest for Common Denominators

How can we offer value for very diverse domains?

- Feature Engineering in Data Analytics
- Automated Feature Selection
- Context of Data Curation
- Anomaly Detection
- Semantic Tools
- ML, ANN
Intelligent Information Arbitrage deliver value for diverse domains

There are domain-specific application preferences.

- Cancer detection, Real-time systems:
  - Early response; Avoid false negatives!
- Robotic defense systems:
  - Delayed response; Avoid false positives!
- Emergency braking in self-driving cars:
  - Neither too early nor too late; Avoid false negatives!

Atrial Premature Contraction anomaly in human ECG

Intelligent Information Arbitrage - deliver value for diverse domains

Processes:
- Acquisition
- Storage
- Processing
- Communication
- Integration
- Analysis
- Mining
- Presentation
- Interpretation
- Retrieval

Entities:
- Data:
  - Uninterpreted
  - Meaningless Symbols
- Information:
  - Descriptive
  - Interpreted Data
- Knowledge:
  - Predictive
  - Validated Information
- Intelligence:
  - Prescriptive
  - Actionable Knowledge

Domains:
- Agriculture
- Healthcare
- Tools
  - AI/ML

Science
Informatics
Tools
AI/ML
The scalpel is only as good as the surgeon who uses it

FEATURE AUTOMATION
FEATURE ENGINEERING
FEATURE SELECTION
FEATURE AGENTS

What we haven’t addressed, yet

That's just the way the cookie crumbles :-(

I know it is a biased point of view.
Book on Digital Transformation in MIT Library
“Haphazard Reality - IoT is a Metaphor”
https://dspace.mit.edu/handle/1721.1/111021

UnPredixable World

https://www.nataliastakhanova.com/research
I know it is a biased point of view. Apologies.

Shoumen Palit Austin Datta
RE: Foot traffic, bone density, Whole Foods and your online pharmacy

wilke@amazon.com

To: Shoumen Palit Austin Datta

Monday, January 21, 2019 4:44 PM

Certainly lots of possibilities here. It was my step-father. I’ve been noodling on the impact of CS on healthcare for, oh, 36 years 😊 Hope all is well.

JAW

From: Shoumen Palit Austin Datta <shoumen@mit.edu>
Sent: Sunday, January 20, 2019 10:13 AM
To: Wilke, Jeff <wilke@amazon.com>
Subject: Foot traffic, bone density, Whole Foods and your online pharmacy

Hi Jeff -

If my memory serves me correctly, I think you told me that your father-in-law was a doctor?

The convergence of your online pharmacy (Taha K-H) and foot traffic at Whole Foods is the future of wellness IoT, which I think you are planning. Think DEXA scanners for bone density of shoppers and recommend adding milk (wellness IoT) with Calcium and Vitamin D to their grocery shopping list. Also, mammograms for $1.99 anytime and remote sensing for vitals for Amazon GO ??

I wonder if you may be interested?

Regards and thanks,

Shoumen

---

Amazon’s head of world-wide consumer talks about integrating Whole Foods—and what it is like to disagree with CEO Jeff Bezos

Jeff Wilke, Amazon CEO Worldwide Consumer, talks to The Wall Street Journal about how much sleep he gets each night, the one trait he looks for in a new hire, and what it’s like being the second-most-powerful person at Amazon named Jeff. Photo: Wixan Ang for The Wall Street Journal
From: Wilke, Jeff
Date: Thu, Jun 2, 2011 at 7:24 PM
Subject: RE: Gibberish for the Soul < ? / ? >
To: Shoumen Datta <shoumendatta@gmail.com>

Interesting, as always. Certainly much of what you anticipate will happen. Some key questions include: which technologies will become the dominant design and are they already invented? How quickly will consumers’ on-hand devices change? (skin vs. six-sided mobile device for example) Interesting plug for micro-philanthropy. Always good to think through this as large businesses are important and visible “citizens.”

I don’t have much criticism: just questions about which technologies on which to bet.

Best,
JAW
Shoumen - I finally had some time to complete a thorough reading of your recent paper, "Adapting Decisions, Optimizing Facts and Predicting Figures." It was certainly thought-provoking. As you know, we have been thinking about some of these ideas, but it is nice to see them woven together more completely than I have before. I think the trick for industry will be to fearlessly use these ideas instead of rejecting them out of ignorance, cynicism, or short-sightedness.

JAW

17 June 2004

Jeffrey A. Wilke

Jeffrey A. Wilke
CEO
AMAZON

Jeff Wilke (JAW) has served as CEO Worldwide Consumer since April 2016. From February 2012 to April 2016, Jeff served as Senior Vice President, Consumer Business, from January 2007 until February 2012, he served as Senior Vice President, North American Retail, and from January 2002 until December 2006, he was Senior Vice President, Worldwide Operations. Jeff joined Amazon.com as Vice President and General Manager, Operations in September 1999. He left AlliedSignal (Honeywell) where he was Vice-President and General Manager, Pharmaceutical Fine Chemicals. Jeff spent the preceding six years in a variety of operations and general management assignments in the chemical, polymer, and electronics industries. Jeff did his graduate work (MBA and MS in Chemical Engineering) at MIT's Leaders for Global Operations (formerly Leaders for Manufacturing) program where he focused on Total Quality and Process Improvement techniques. He began his working career writing software and leading software development at Andersen Consulting (now Accenture). Jeff also holds a BSE degree in Chemical Engineering (summa cum laude) from Princeton University, NJ.