COMMENTARY [E]
" SIGNALS "

Shoumen Datta

HAPHAZARD REALITY – IOT IS A METAPHOR

https://dspace.mit.edu/handle/1721.1/111021

"SIGNALS" contains a series of essays spewing amorphous thoughts:

1. SITS
2. SIP-SAR
3. SARS💧AG
4. ART
5. PEAS

Please download "SIGNALS"
● PDF from MIT Library
https://dspace.mit.edu/handle/1721.1/111021
Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)
"IoT is a Metaphor" (see Commentary E)
Please download "IoT is a Metaphor"
● PDF from MIT Library
https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.

"SIGNALS" contains a series of essays spewing amorphous thoughts:

1. SITS -------- **You are here**
2. SIP-SAR
3. SARS♠AG
4. ART
5. PEAS

Please review "SIGNALS"

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)

"IoT is a Metaphor" (see Commentary E)

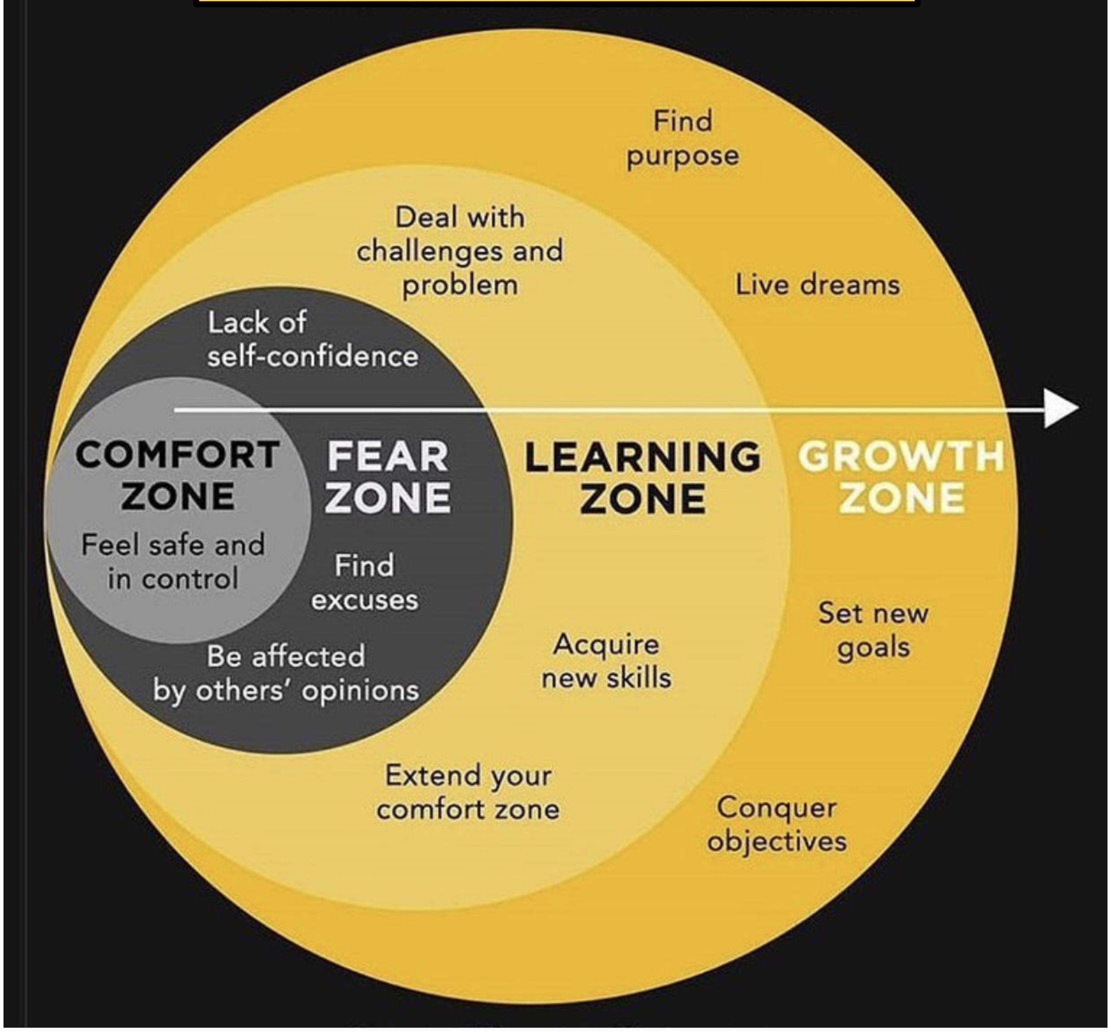
Please review "IoT is a Metaphor"

● Download PDF from MIT Library

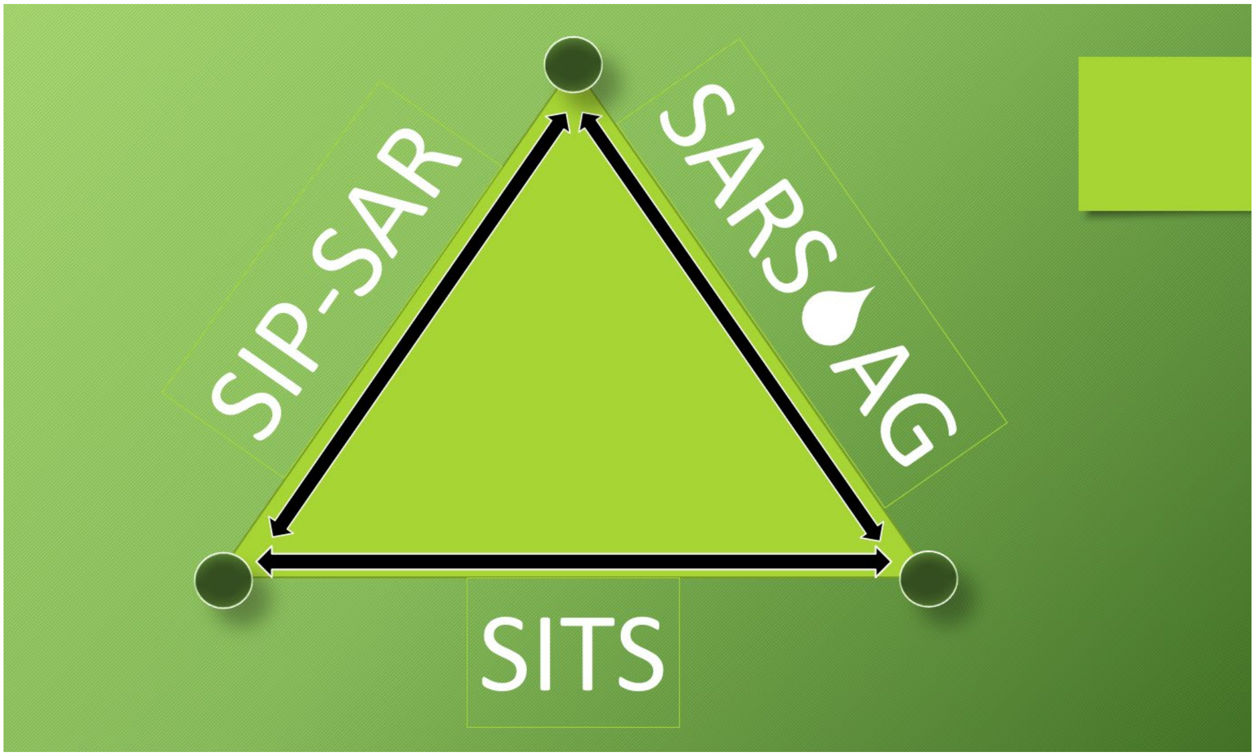
https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.

THE DISCOMFORT ZONE



SITS – SIGNALS N THE SOIL



SIGNALS IN THE SOIL

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?

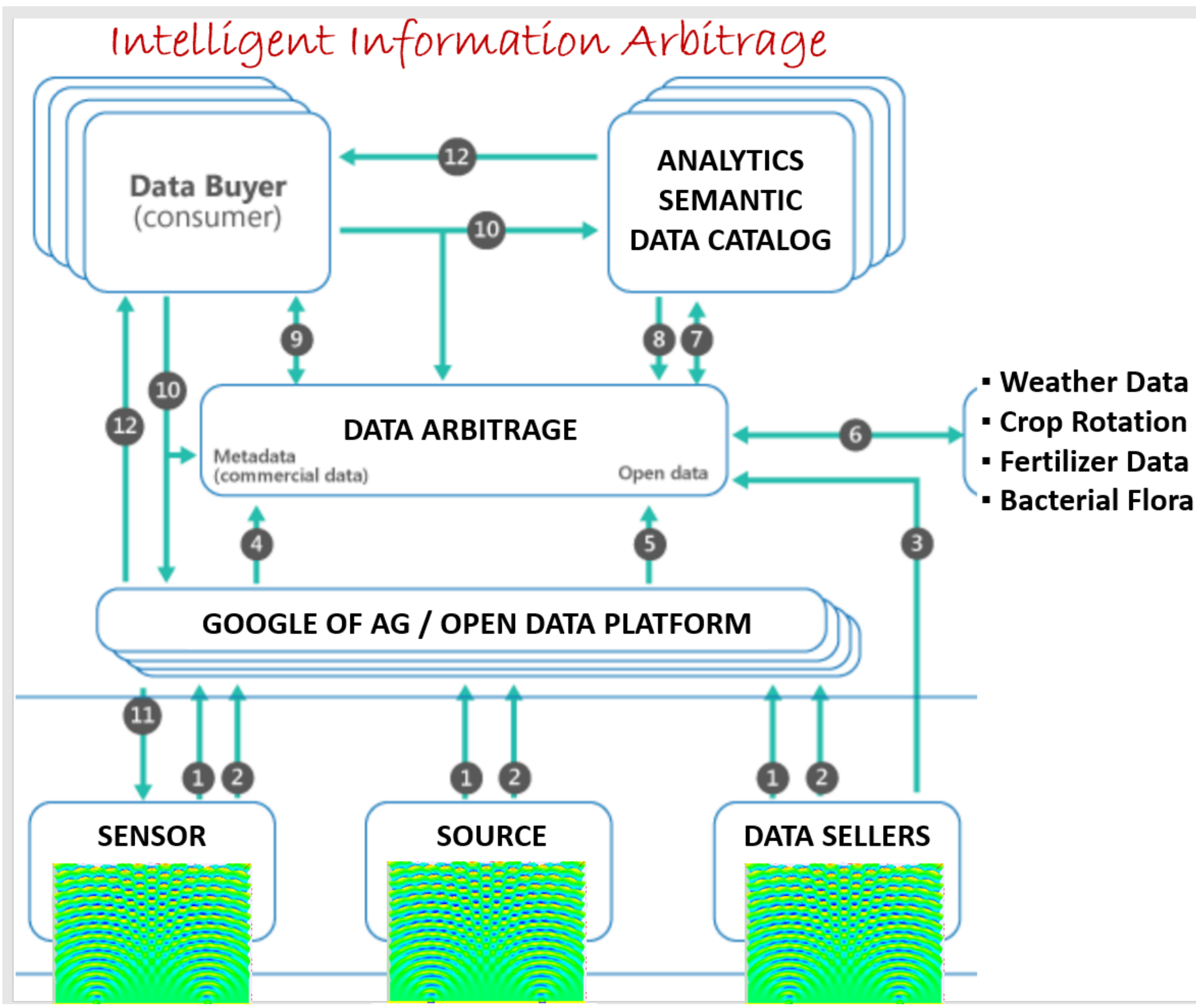
We are approaching an inflection point where it may become feasible to use nano-satellites, on a pay-per-use basis, to download daily or hourly precision micro-climates (by GPS coordinates, zip codes) via apps on smartphones, tablets or portals, connected to (intelligent?) decision support systems.

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 "PEAS" may be found in this lengthy collection of essays **'SIGNALS'** (PDF is here https://dspace.mit.edu/handle/1721.1/111021).

"What we know is not much.
What we do not know is immense."

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 ▪ MDPnP 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 ideas may be practical in the short term and some may seem pragmatic in a few more years. Other thoughts may be found in "PEAS" at the end of this PDF. Still others may take shape a few decades from now. For the latter, *is earlier always better*[1] if one wants to plant a few seeds? Early adopters may fail[2] 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 RFP[4]), 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 research[5] data[6].

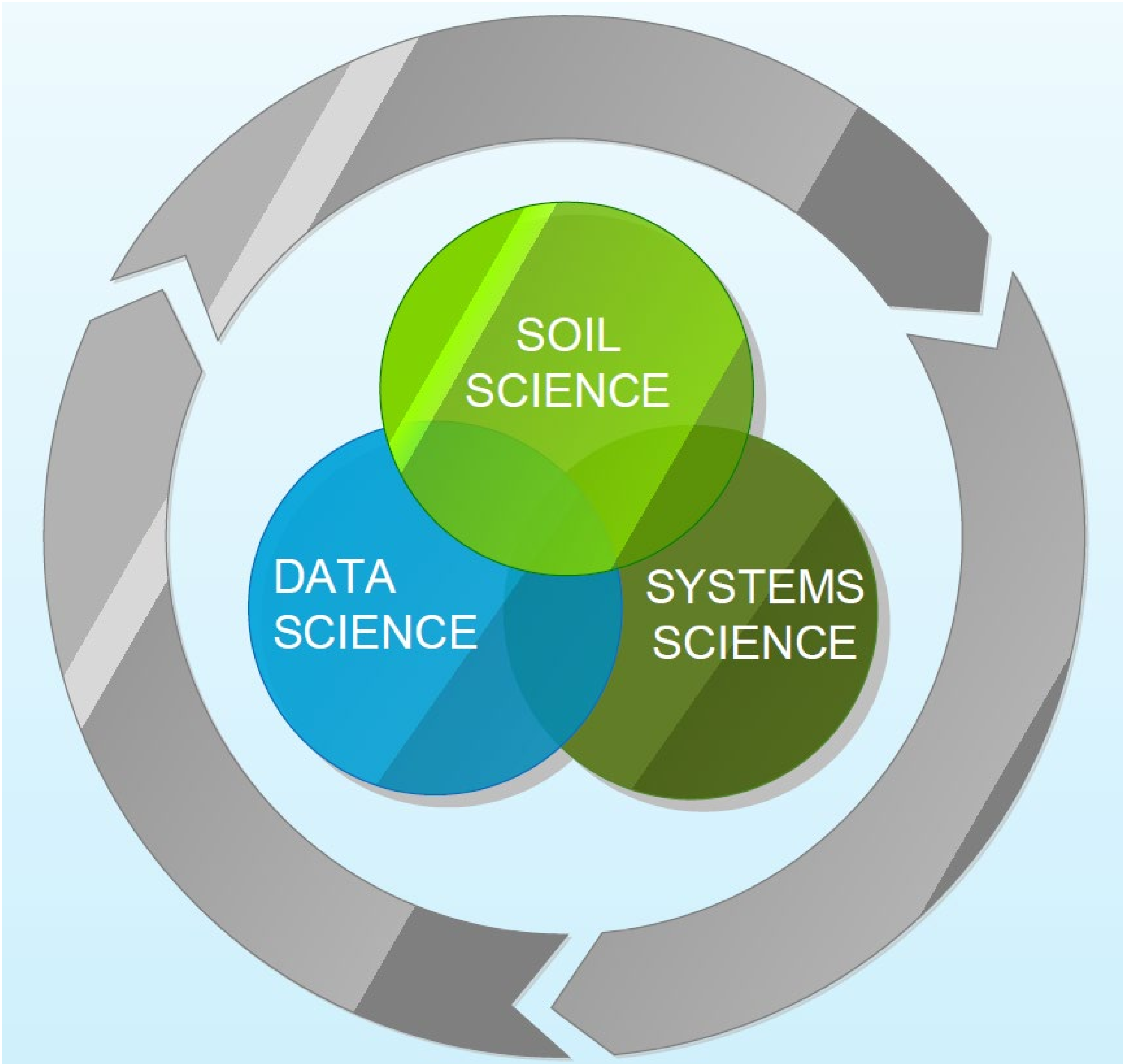
Provided data is sufficiently valid, relatively noise-free and originated from a reliably calibrated instrument, the data may be useful ***agnostic of the source***, instrument or sensor. Acquisition of the data from the source, instrument or sensor, is ***agnostic of the medium*** of the raw data (waves, voltage, chemicals) as long as the signal transduction is reproducible, does not violate laws of physics and its authenticity is uncontaminated by signal processing artefacts[7] (anomalies often 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, which serves as the *context*. When they reach their destination (question is answered), they hop off. The automobile is a convertible, an open platform. Time to time, elements under the bonnet and boot may be serviced, replaced and upgraded. Semantics and ontology are crowdsourced fuels. The (abstract) journey continues irrespective of incongruities, on the unpaved road, to meet unknown challenges.

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[8] 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[9] (**NORA**) may move beyond the ***awareness***[10] realm to empower a new breed of thinkers to ask the correct questions, and tinker with unknown unknowns, in quest of answers, knowledge and/or breakthroughs. The shrink-wrapped version of existing conventional wisdom[11] 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[12] improvement of candles?



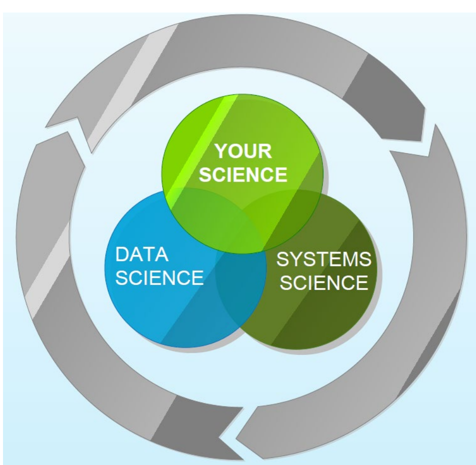


Figure 0 – In the SITS context, the proposal may focus on the trinity of soil science, systems science and data science. Systems science refers to the vast ecosystem accompanying the domain of interest. It must be accounted for in terms of interrelationships relevant to sub-domains [for example, soil 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 (in other words, 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[13] and cross-kingdom tools in plants[14] and humans[15] 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[16] of the virome are currently documented.

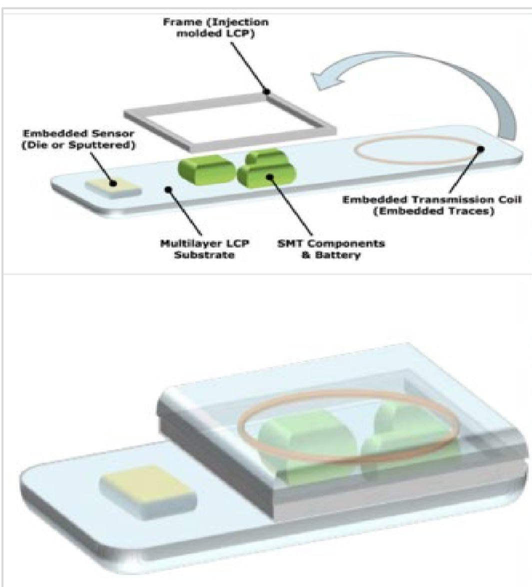
With respect to soil science, whether the outcome of this miniscule slice of the system (SITS) will generate any results, remains to be seen. Developing tools to answer obvious questions generates data. Analytics may reveal cryptic relationships, which may hold clues for molecular interactions (plants, 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 and/or a nugget of *experience* just emerged.

At this time, we cannot answer these questions. Hence, we will focus on outcomes driven by tools to acquire data, to catalog categories, attributes and characteristics. 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. When combined with a thoughtful data management plan, this approach may illuminate the meaning of this data, at the least, with relevance to SITS. This is a journey and the destination is still TBD.

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).

Physico-Chemical Properties and Biological Content: Is this pedology still relevant for SITS ?							
Microbe 1	Microbe 2	Microbe 3	Microbe 4	Microbe 5	Microbe 6	Microbe 7	Microbe 8
pH	moisture	salinity	metals	ions	density	particles	color

	Microbe 1	Microbe 2	Microbe 3	Microbe 4	Microbe 5	Microbe 6	Microbe 7	Microbe 8
pH								
Moisture								
Salinity								
Metals								
Ions								
Density								
Particles								
Color								



	Unit	LCP Flex	Polyimide Flex
Thickness	µm	25, 50, 100	12, 25,50
Electrical properties			
Dielectric constant (10 GHz)	-	2.9	3.2
Dissipation factor (10 GHz)	-	0.002	0.002
Surface resistivity	Ω	1.0 E16	4.0 E13
Volume resistivity	Ω cm	1.0 E18	2.6 E14
Dielectric strength	kV/mil	3.5	7
Mechanical properties			
Young's modulus	GPa	2.3	7.1
Tensile strength	MPa	280	220
Thermal properties			
CTE, x-y	ppm/K	18	20
CTE, z	ppm/K	200	120
Solder float temperature	°C	> 288	> 300
Melting temperature / glass transition	°C	330	343
Other properties			
Moisture absorption (23°C, 24h)	%	0.04	1
Flammability	-	UL 94 VTM-0	UL 94 V0

Figure 1 – Sensor engineering[17] 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).

SCIENCE – USING PRINCIPLES AND PRACTICE OF PHYSICS (OF REMOTE SENSING)

SITS may reveal its value in terms of the creative convergence[18] 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[19] can be used for remote[20] applications in humans. Remote sensing in the context of SITS include [a] concepts[21] for moisture detection by satellites[22] and data[23] from 102GHz frequency, as well as [b] use of 2.45GHz (WiFi) to estimate soil moisture[24] using drones[25] rather than WSN[26] to acquire data from immobile tethered sensors, which pollutes the soil, yet, often promoted by laissez-faire titles and published[27] 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 (0.3 to 3.0) may be useful frequency cluster for detection of radio signals from bio-molecules[28] (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[29] 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[30] in (artificial) dielectrics[31] 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[32] and for many[33] uses. It may hold clues for remote sensing tools to measure pH, salts and enzymes[34] in the soil, as previously[35] suggested. Static dielectric constant of water[36] drops with addition of salts, but the reverse is true, if the complexity of permittivity of water is explored with THz transmission spectroscopy[37] (the value increases, Debye relaxation, with salt concentration).

For SITS, is it possible to exploit any correlation[38] between dielectric permittivity, ionic charge, charge density (phosphates) and chemical structure? Dielectric permittivity[39] is the primary diagnostic physical property[40] for GPR[41] (ground penetrating radar)? Creative use of ISAR[42] (inverse synthetic aperture radar) in the development of WiVi (Katabi *et al*) emerged from radar technologies. Dielectric permittivity is one of many tools from radar technology which may be quite pragmatic in terms of applications for the ag/food industry.

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[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[44] as well as investigate other ion-based detection systems (GC/MS[45]).

The importance of "markers" cannot be overemphasized. Glucose oxidase was the first enzymatic marker[46] (coupled to an amperometric electrode for monitoring oxygen in blood) followed by various forms of spectroscopy. Raman spectroscopy[47] and bio-impedance spectroscopy[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[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^+] and increase in [K^+] 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.

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[50] changed the field of medical imaging. By analogy, it is tempting to speculate that nanoscale vector magnetometry[51] 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[52] 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[53] 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[54] may have appeared as far-fetched "pie-in-the-sky" idea.

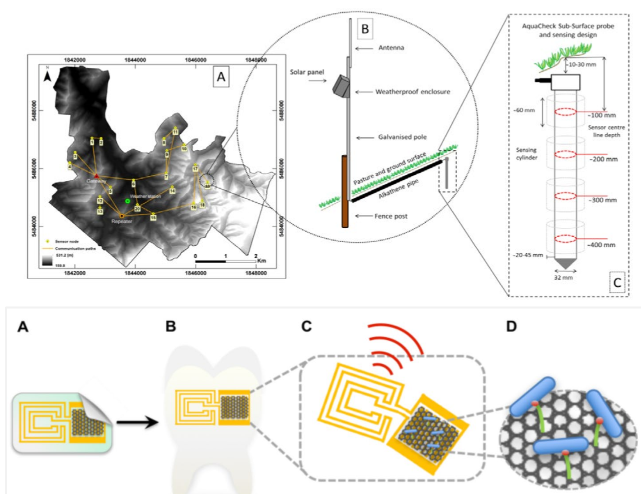


Figure 2: Top - The "bucket brigade" for data collection. Sensors in the soil[26] vs modern wearables[55]. Bottom - Glucose sensor[56]. (A) Graphene printed on bio-resorbable silk with wireless coil. (B) Bio-transfer of sensor on tooth. (C) Non-invasive wireless glucose sensing (D) Self-assembly of pathogenic 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[57] were fabricated. Capacitive micromachined ultrasonic transducers[58] (CMUT) recently reached the summit of the hype curve[59] with respect to medical imaging. However, in this instance (for SITS) air-borne applications[60] may be more relevant. Use of a multi-frequency CMUT device for ultrasound capnography[61] suggests that soil gas analysis using ultrasound[62] may be possible. Advances in precision ultrasound chips[63] treats gas as an interference, to be damped. What if we perhaps reverse the experiment and ***measure*** the gas (multi-frequency ultrasound device for multiple gases to be measured) in the soil? Can we use gas[64] as markers? Gas-as-a-marker may be suitable for determining 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[65].

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[66]), Kalman (1960[67]), Granger (1969[68]), Engle (1982[69]) and Granger and Engle (1987[70]). A recent paper (2019[71]) 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 problems in error correction may concern protein dynamics[72] and implied by Martin Karplus[73] and suggested[74] by others. It appears that proteins are radios. Could we use THz waves to explore[75] protein signatures? The signal will be contaminated by noise from vibration of water molecules (think van der Waals radii).

SCIENCE – PRINCIPLES AND PRACTICE OF CHEMISTRY (OF MOLECULAR ANALYTES)

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 crop-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[86] are one of the widely used sensing tools (in addition to other uses[87] and biological role of lectins[88]). Recent advances[89] using fluorescent emission spectra[90] 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*[91], *Bacillus subtilis*[92], *Pseudomonas fluorescens*[93], *Nitrosomonas europaea*[94], and Cyanobacteria[95] (*Synechococcus sp* and if there is/are soil equivalent of *Prochlorococcus*[96]). 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.

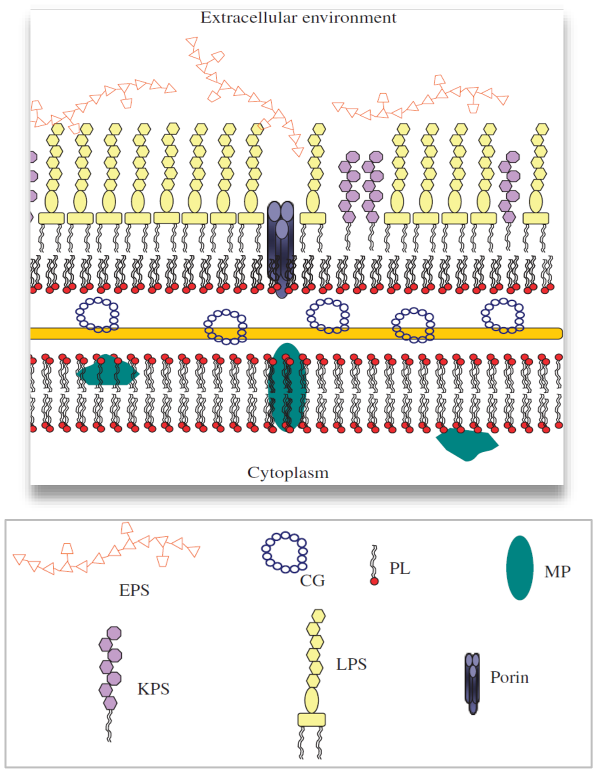


Figure 3 – Rhizobial cell surface[97] cartoon, illustrated with polysaccharides involved in rhizobial attachment to roots. The "accessible" nature of EPS and KPS suggests its potential as molecular targets for binding to sensors. OM-outer membrane; PS-periplasmic space; PG-peptidoglycan layer; PM-cytoplasmic membrane; EPS-exopolysaccharide; CG-cyclic glucan; PL-phospholipid; MP-membrane protein; KPS-capsular polysaccharide (K-antigens); LPS-lipopolysaccharide.

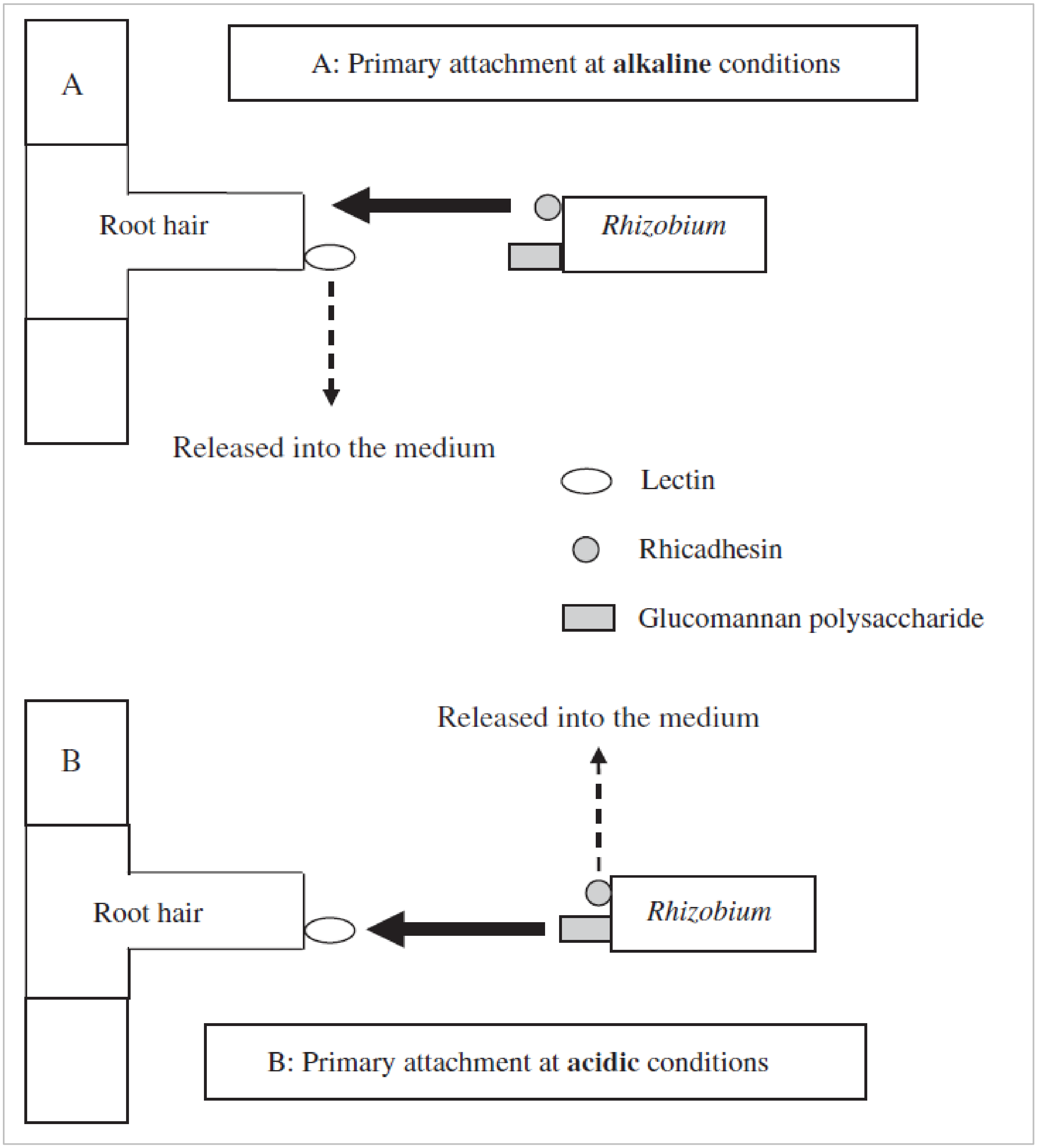


Figure 4 - Attachment of *Rhizobium leguminosarum* to lectin released from pea and vetch root hairs is pH dependent[98]. (Top) Alkaline: Lectin released from root-hair tips. Rhizobial Rhicadhesin mediating attachment. (Bottom) Acidic: Lectin remain anchored. Rhicadhesin released from bacterial surface. 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[99] 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 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[100] 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[101] and natural selection[102] may be unexplored but the role of thermodynamics in chemical reactions and molecular interactions[103] 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[104] in sensing[105] devices[106] may be an important topic[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[108] (perhaps mimic how enzymes are used in the laundry[109] industry).

We will continue to identify molecules sensitive as a detector for sensor development, suitable for signal transduction, amenable to generating43 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 immensely valuable and a worthy sign post for SITS 2.0 proposal of the future.

Sensors based on binding of an analyte follow the "lock-key" model[110] which may be the most widely used analogy[111] in biology. The actual biological mechanism, the induced fit[112] hypothesis, may be relatively unknown. The proposal of induced fit was a ***déjà vu*** moment if we recall that the Periodic Table[113] is not[114] what it used to be[115] at the turn of the 20th century[116], a fact which is rapidly receding from general knowledge, judging from a recent title[117] of at least one global publication, 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[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[119] transitions? The critical role of allosteric ligands[120] and the value of allostery[121] remains under-appreciated. The observation of long-range allosteric effect in hemoglobin[122] is now relegated only to the hidden pages of history.

The chemical identification of an analyte by a sensor (eg sensor microarray[123] on a chip using nanoplasmonics and fluorescence assay[124]) may provide evidence for the presence of a biological species. The conundrum, for ag, is that soil levels of microbes and nutrients, and changes, are ***non-linear*** in terms of crop growth (plant tissue). The *value for growers is* ***that data*** which synthesizes signals not only from a single species as an indicator but provides *real-time* ***information*** about the "jigsaw" of the fauna and the functions[125] of soil biota "guild"[126] which is a confluence of soil biology and chemistry (Table 1). Data about functional guilds and the ecosystem must be combined. The physics and chemistry of sensing as well as the design of sensors for agro-ecosystem management demands a ***mix of signals from the breadth of responses*** from guilds. These guilds may be crop-specific.

DATA SCIENCE – UNDERSTANDING DATA IN THE CONTEXT OF SYSTEMS SCIENCE

Those who cannot remember the past are condemned to repeat it[127] 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[128]. In 2015, IBM's Modha[129] made the idiotic claim to have ***"the brain in a box by 2020"*** and corporate lunacy[130] followed. In 2000, President Bill Clinton[131] declared that the Human Genome Project[132] 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[133] with potential[134] but far from panacea. The hoax[135] and hubris from snake oil salesmen[136] resonates on the Minsky[137] scale[138], *"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[139] of the semantic web[140] as a tool for "understanding" was, albeit, temporarily[141] crushed, a decade ago. Since *Nature abhors a vacuum*[142], hysteria[143] and deadly[144] sins[145] of AI[146] raced to fill[147] the void. Media frenzy and AI stupidity[148] overshadowed[149] the importance of machine learning[150] (ML) and even experts used *the cover of* AI[151] to sell books, rather than inculcating the *society of mind*[152] paradigm. The paradigm has shifted in favour of news cycles as catalysts for polishing the chrome, rather than the ardor for tuning the engine[153].

From the dawn of computing, the role of context[154] in understanding the meaning of data was not a part of the process, for example, the difference engine[155] and what followed thereafter. Yet there is little need to overemphasize that information arbitrage tools[156] 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 also 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[157] and lack of data interoperability[158] is the third leading cause of death in the US[159] (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 *irremediable injustice*[160] perpetrated by the binary system. Proof of concept[161] and examples[162] from late 20th century reveals that we may know[163] how[164] to make computers ***understand***. Fundamental principles of these tools remain in relative obscurity except perhaps one[165] application which was "dumbed down" to serve as a weather[166] "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[167] 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[168] elsewhere and attempt to embrace data, context, and connectivity, combined. The latter implies that tools from the semantic web standards (OWL/RDF) 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 (knowledge graphs) in using semantics rather than remain handicapped by the dead weight of old technology.

Context and understanding are inextricably linked with value. Making sense of data is crucial for adoption. Trust in the value of information increases when it is based on curated (noise-free?) data which can be shared between system of systems using open standards[169] (for example, DDS[170] or data distribution service[171]). However, in attempting to free data from noise, it is possible that we might eliminate "signature" signals about the ecosystem (environment which is generating the noise). Should we be as interested in the associated noise as we are in the signal itself? It will be a mordant irony if the brilliance of science and signals is rejected due to ignorance, if we fail to view noise and errors[172] with new[173] eyes[174].

Hence, data by design ***might separate but include "noise"*** in the context of error correction (correct obvious errors but provide errors in a separate container for analysis by qualified users). Enabling the interpretation of errors, by a host (crowd) of users, can breathe new life into *signals latent in the noise* and new ideas about the concept of flux (data versus fluctuations of noise-types associated with data). The latter may contribute to the design of solutions, because multiple points of correlation could be helpful to provide a better picture of the guild, where the type of noise may be a potent signal for changes in the agroecosystem.

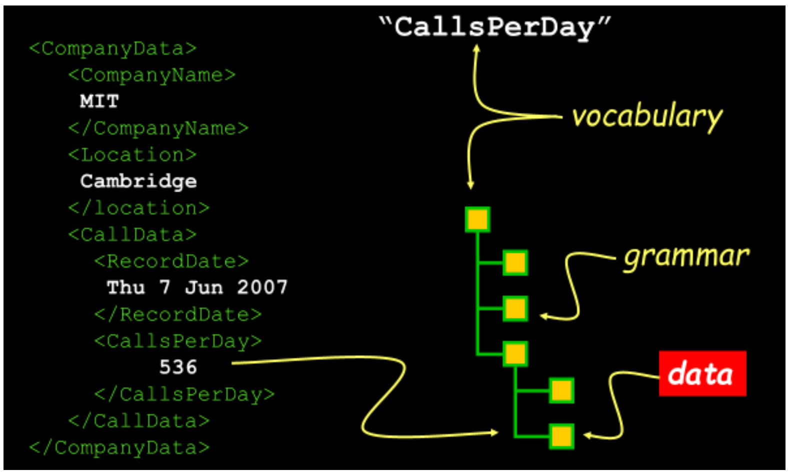


Figure 5 – Problems that can be solved[175] still fuels lack of interoperability even between intra-company databases. The syntax "callsperday" hard-coded in this software lacks any attribute which can be included in any semantic data dictionary because "callsperday" is void of key linguistic structure[176] (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[177] and domain knowledge[178] 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). DDS may offer solutions but vendor-specific implementation of DDS are similar but not identical.

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[179] 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.

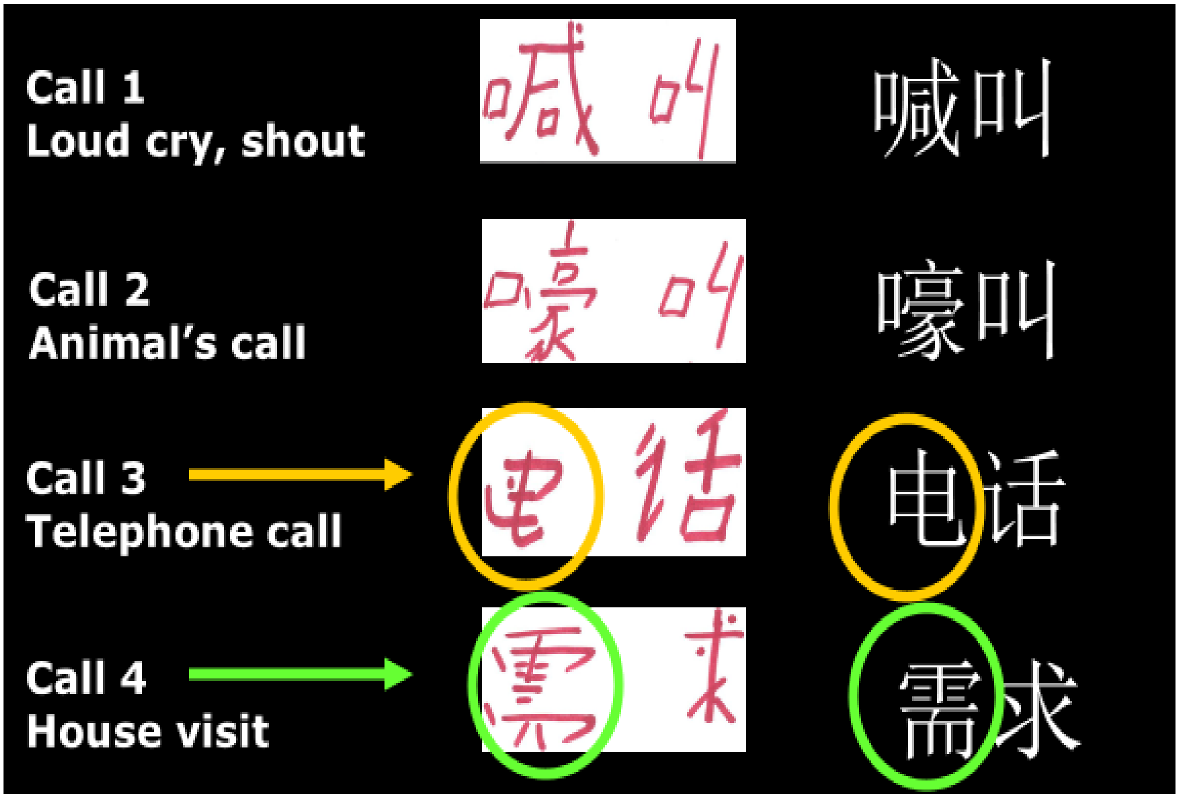


Figure 7 – The task of understanding is far more difficult than perceived by the Anglo-Saxon world. For a binary system agnostic of natural language, the difficulty in understanding the meaning of the word "call" in the context of its use and ***user***, may be difficult (impossible?) to accomplish using traditional[180] semantic web ideas, which lack binary translation features in its toolkit. RDF[181] (resource description framework) is the key to data interchange and relationships between things (triples) which leads to directed graphs (edges of relationships). Other tools, for example, OWL (ontology working language) which builds on RDF, 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[182] of the ***descriptive structure*** with binary translation for ***context*** (call.1, call.2, call.3, call.4) may exceed *machine readable*, it may be ***machine understandable***[183].

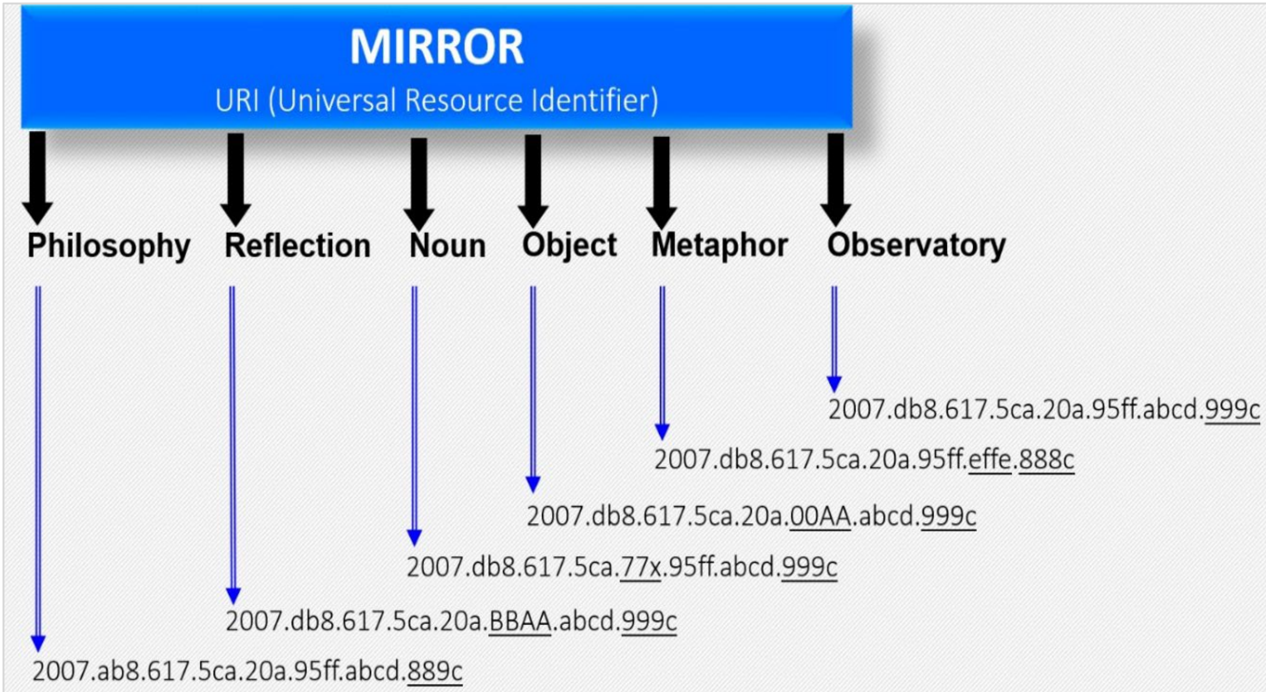


Figure 8 – Digital Semantics – URN – Universal Resource Numbers for ontological frameworks? ***Each context*** has an unique id [numerical (knowledge) representation]. The list can grow (number spaces similar to EPC[184] or UPC[185] barcode) as communities/countries crowdsource and contribute to the "context list" the *descriptive structures* relevant to ***their*** **language and cultural context.** Mapping between a numeric system and categorization (taxonomy[186]) with relevance 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). The statement, "it mirrors my life" conveys the meaning that Mireille is choosing to be "philosophical" about her life[187]. 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 'marketing' departments have 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[188], its adoption, especially in business and industry, is solely a matter of economics. Exploring corn[189], dynamo[190] and electricity[191] have 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[192] 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[193]. Each edge of a graph (the line connecting the circles, the nodes), is a relationship. The internet[194] is an example of a directed graph[195]. 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[196] which users can access, analyze and interpret, to derive value.

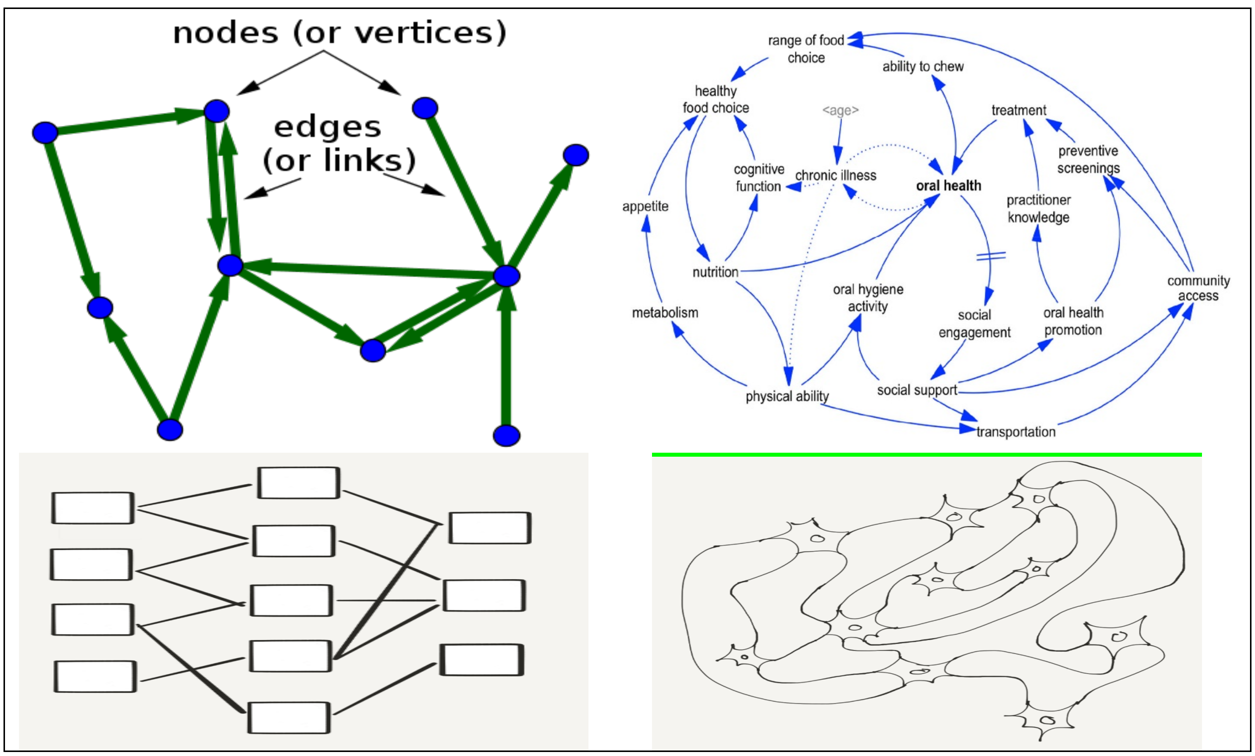


Figure 9 – Example of a graph (top, L), a system dynamics model[197] (top, R), cartoon of an artificial neural network (b, Left) and sketch of a biological network of neurons (b, Right).

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[198] 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[199] and traversed[200] in relational databases, hence, the question, do we need graph databases?

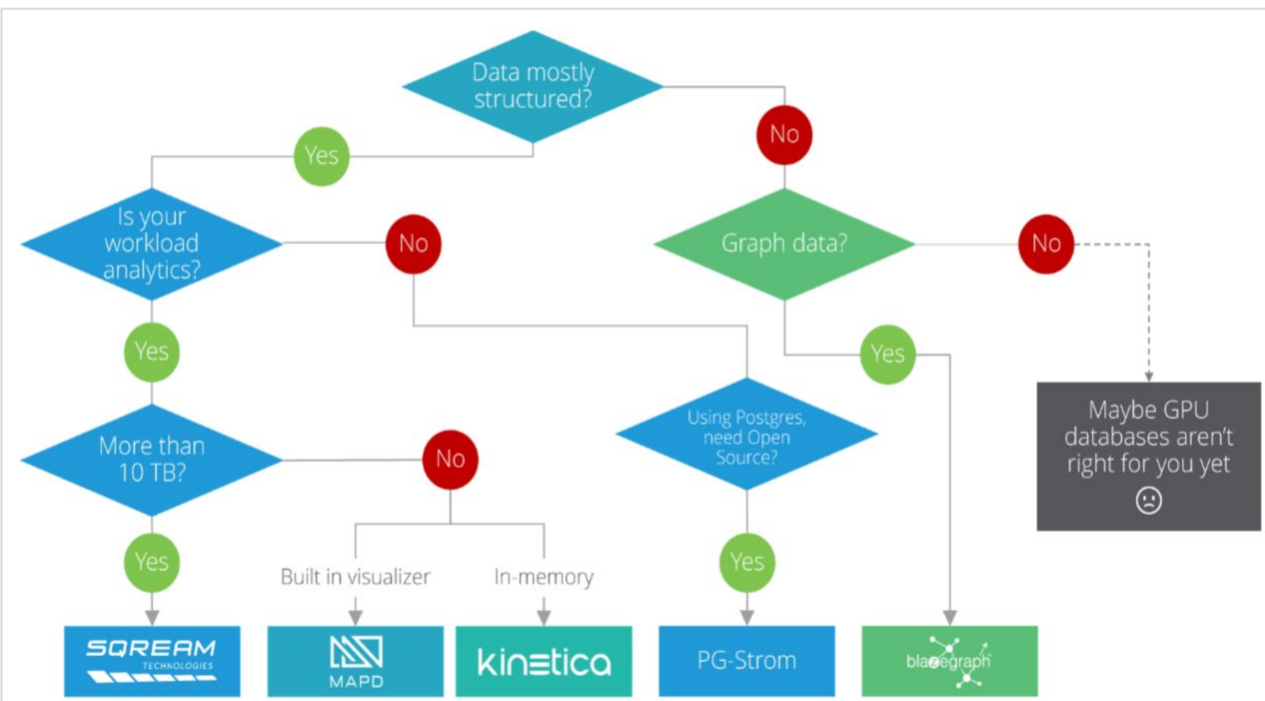


Figure 10 – Graph DB: a future forward strategy? Is graph database[201] right for SITS?

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[202]), is mandatory. Graph databases are likely to handle multiple query languages. Can this be an opportunity for more research and innovation in semantic query languages?

In the context of the data, context determines the users focus on the data and *that* 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, < soil pH 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 ask whether 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, perhaps better captured in knowledge 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. Can users "drag and drop" this information in the database? Open interfaces (API) to import sensor data (local, global) in the knowledge 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[203] 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 already occupying the database.

One obvious approach may involve creation of an automated feature[204] selection tool. It may be downloaded by labs interested to share sensor data. Feature[205] 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 easier or a less arduous task.

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[206] (such as Farm[207] IoT) 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, linked to an open source time series database[208]).

Importing raw sensor data and populating structured data fields may be sufficient in relational databases but in SITS graph database, setting up the graphs (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[209] 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, ML, algorithms). This may not happen on ***day 1*** but must be a salient feature for knowledge platforms of the future (KIDS in PEAS).

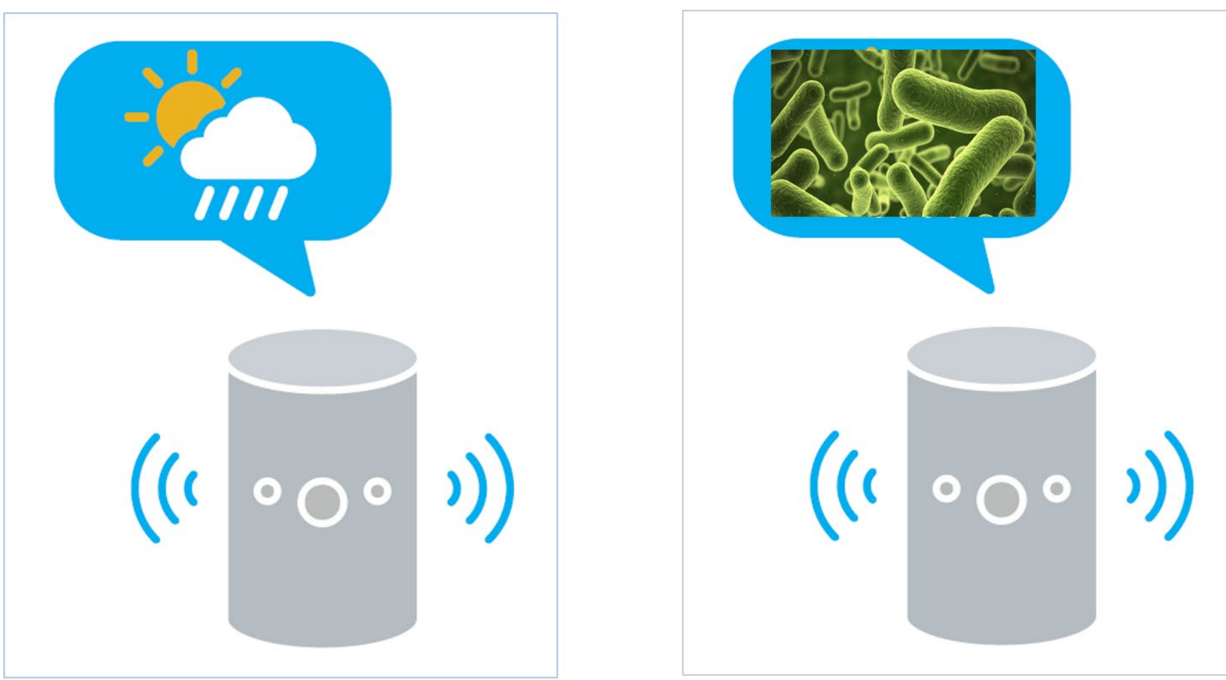


Figure 11 – Knowledge graphs come alive when we query "Alexa" what is the weather? (Left box). What we propose to achieve is the SITS version for ag. We may ask "Hypatia - which types of bacteria are in the soil sample from Alexandria?" (Right box). Note: re-visit Figure 1

TEMPORARY CONCLUSION

The suggestions here are amorphous but enabling mass adoption[210] 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 "Google of ag" which may be 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[211]) which is one reason for suggesting open source[212] GPU[213] database in the design of SITS.

Figure 12 – Viewing the ancient library card catalog with ***"new eyes"*** in the Semantic Data Catalog. SDC contains links (URL, URI, URN) to resource and information/connections to other SDCs, to navigate across information spaces. 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[214] 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[215] are increasingly using semantic graphs to keep up with the edge[216] of connectivity (explore ART and PEAS).

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[217] 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 "Google of ag" may be an easy undertaking.

In attempting to provide my two cents on this *threepenny opera*[218] I attempted a task much too great for my abilities, the extent of which I didn't perceive when I started[219]. We choose to pursue these ideas "***not because they are easy, but because they are hard***"[220].

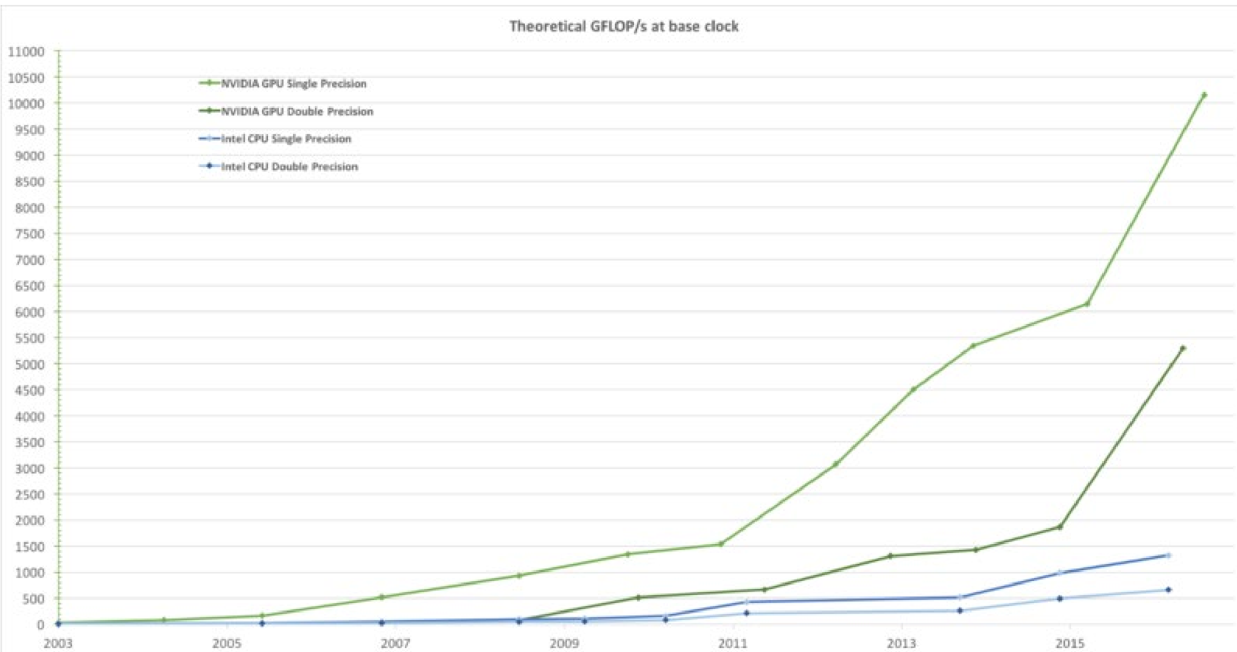


Figure 13 – Signals[221] in the soil databases will not only provide access to the data but also display GIS maps to show source of the signal (and environment, for example, tomato growers in Florida or rice paddy on terrain in Yunnan Province, China). To render realistic views of images at a high frame 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[222] compares GPU vs CPU trends.

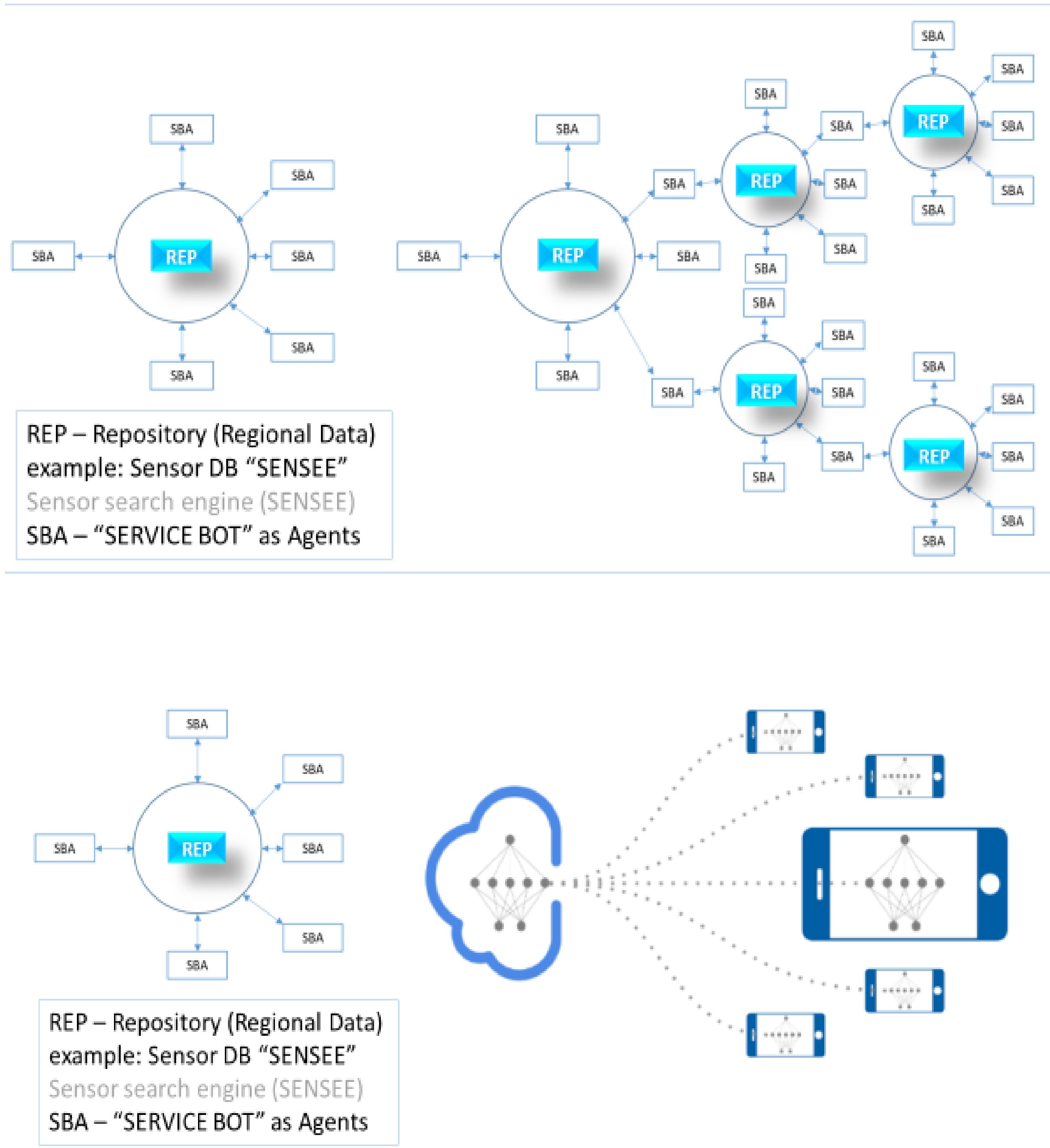


Figure 14 – Locally distributed repositories [SENSEE] containing data and questions (relevant to users and growers by crop or environment) can be globally connected by Agent-based communication (top panel) and consumed by any system, in near-real time, due to interoperability of open source mobile platforms (bottom panel). Explore ART and PEAS.

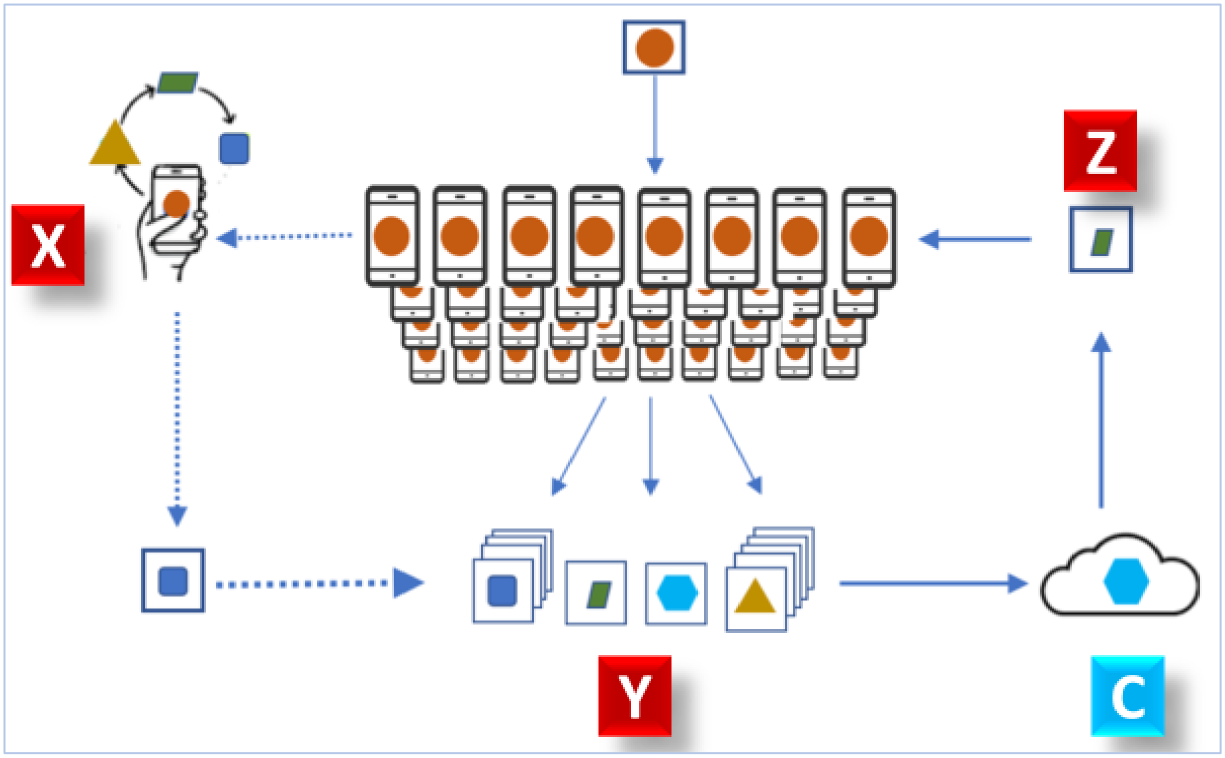


Figure 15 – The value of the REP concept (SENSEE) may be enhanced by coupling publish-subscribe modes with crowd-sourced data adoption/dissemination. User "X" (edge) may update data, recommend tools or techniques or share outcomes/outputs (for example, growers can share photographs of infected produce or sinfully delicious tomatoes). Thus, local user personalization (point X, in the crowd - edge) is sent/stored to the analytical platform (engine Y). An emerging consensus from contributed data (for example, improved technique or data with incorrect units or better use of a tool) is sent to cloud C for expert evaluation and critical analysis. Verified change Z is communicated to all subscribers, globally. This process repeats, to enhance open models and enrich common goals for public goods, using distributed data from crowdsourcing (experts, users, farmers, growers, scientists, engineers, academics, politicians) but deploying a neutral/trusted analytical evaluator (cloud C) to deconstruct/reconstruct, aggregate/disaggregate data and models, to serve the best interest of the system. It may prevent data pollution, act to neutralize cyberthreats and stop, if at all possible, attacks perpetrated by GAN (general adversarial network) as infectious agents.

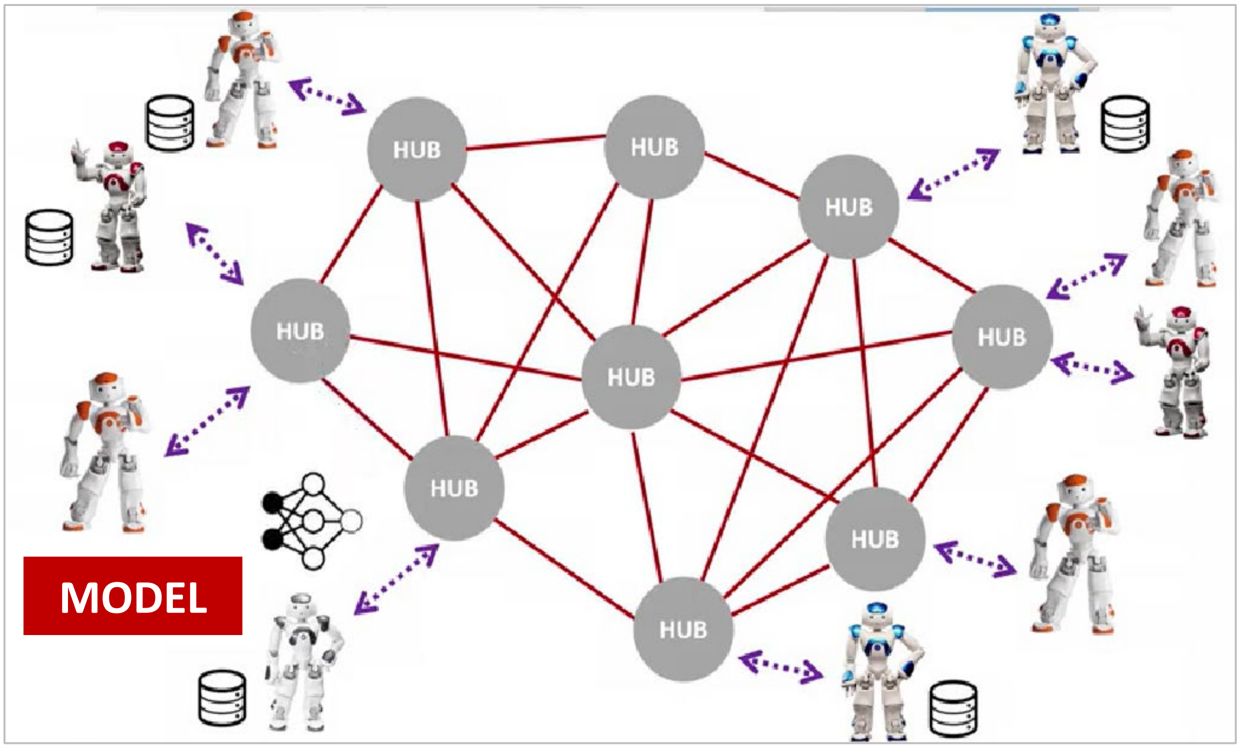


Figure 16 – The SENSEE suggestion increasingly *reverses* the paradigm where the edge (the user X in Figure 15) becomes the point of information, analysis, feedback and actuation, because it is the user at the edge (grower, farmer) who must benefit from data arbitrage. In this approach, we envision that the "model" (see cartoon, lower left corner) which is created at the edge, locally, by one or more (crowd) of users, is of value to that user or user group. But the learning from this edge (the cartoon) may be applicable in educating other users (other edges in the illustration). Imagine the cartoon "model" to travel over the "network tracks" (hubs and spokes in the illustration) and serve other users. They may use "as is" or modify or reject the model. ***In other words, we deliver the analytical engines, models and algorithms to the data, rather than the classical concept of sending the data to a hub or cloud for analysis and analytics.*** These ideas as based on "federated learning models" commonly used by financial institutions and banks to train fraud detection models without sharing their sensitive customer data. Popular frameworks now include TensorFlow Federated, an open source framework for experimenting on decentralized data. PySyft is a open source library built on top of PyTorch for encrypted, privacy preserving deep learning.

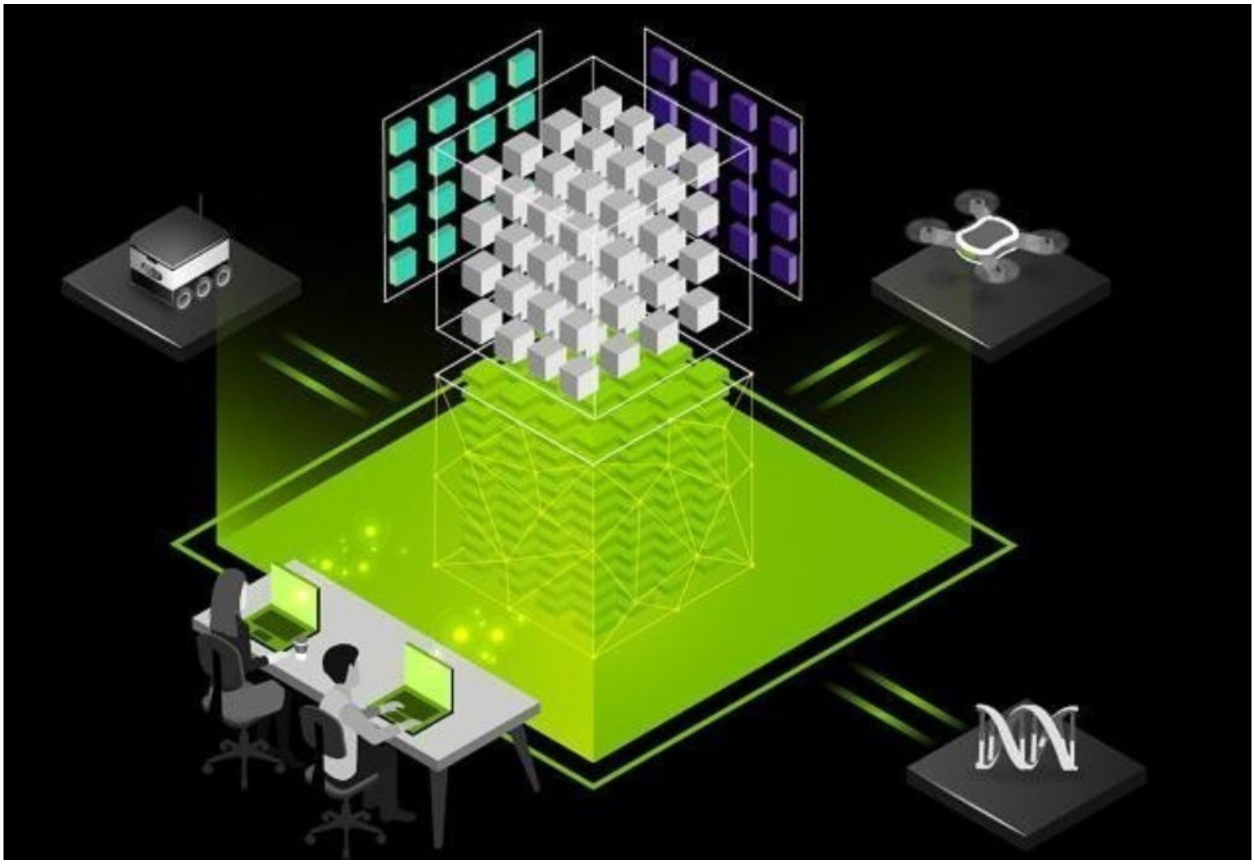
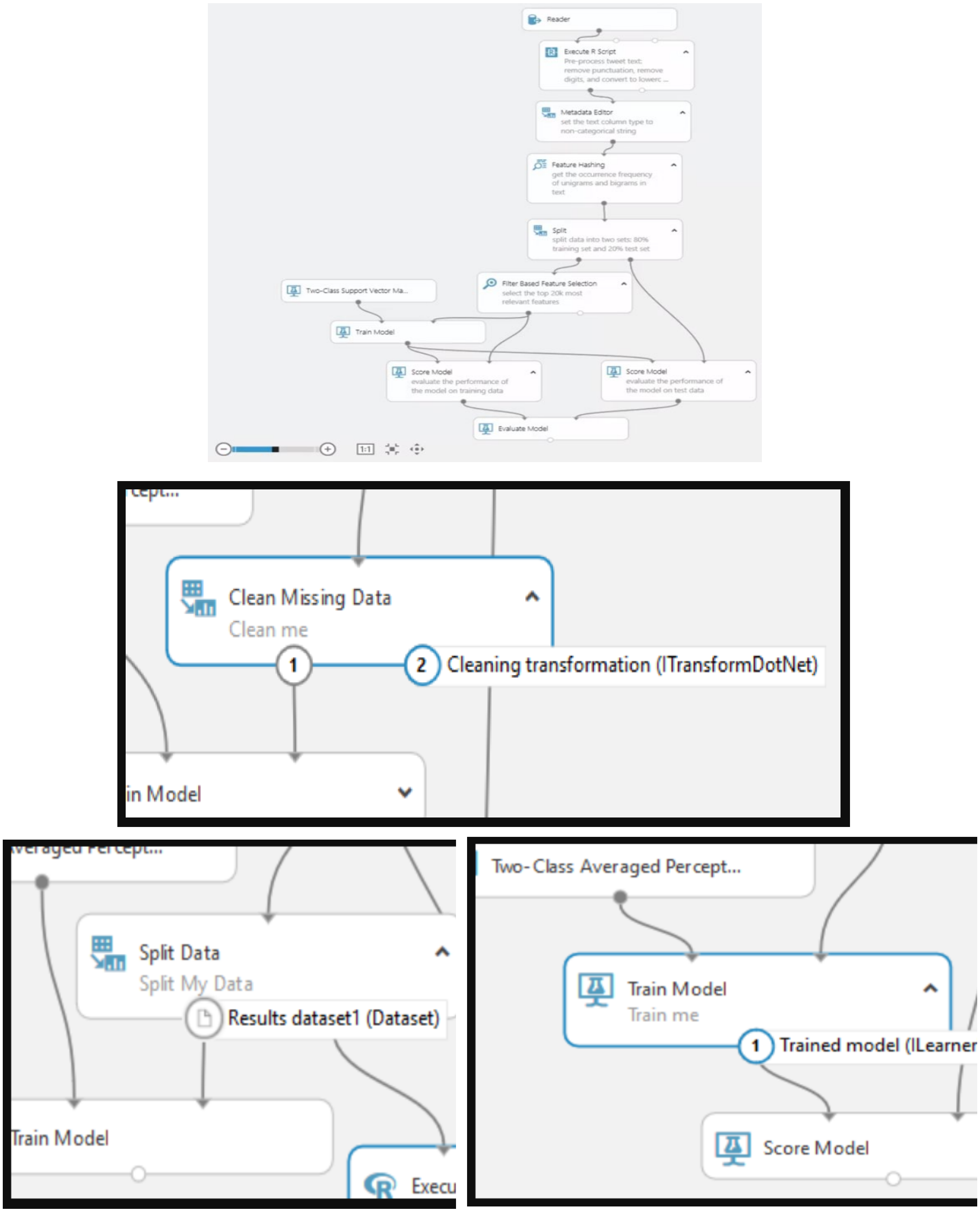


Figure 17 - Federated AI Technology Enabler (FATE) is an open-source project initiated by Webank's AI group to provide a secure computing framework to support the Federated AI ecosystem. Despite the hype whipped up by the glib, snake oil salesmen of AI, there is value in this approach, if and when rationally analyzed, for specific purposes, using bonafide tools, which may be customized for specific applications and are based on rigorous mathematics and statistics. It may be useful for SITS and its ecosystem to explore these advanced tools of the future and adapt enterprise solutions around federated learning for the agro-ecosystem. At present, the primary deployment challenge may be the computational constraint of edge devices (smartphone, tablet) to perform local training, cloud consultations and inferencing. However, smartphones and IoT devices are increasingly equipped with GPUs or sufficient computing hardware to run CNN/RNN and other ANN/AI models at the edge to augment near-real time "intelligent" decision support systems, at the point of use. In Figure 14 the idea of REP/SENSEE may be the SITS "smart path" approach to harvest these ideas and convert them into actionable transactions to help the ag industry in the pursuit of food.

Figure 18 – Sense of the future – “DRAG AND DROP” interfaces of partially automated processes which will be agnostic of the user’s knowledge of programming (akin to the next generation of Lego Mindstorms). Cartoon source https://github.com/hning86/azuremlps



REFERENCES

[1] http://news.mit.edu/2019/securing-internet-things-in-quantum-age-0301

[2] https://www.researchgate.net/publication/228136384_First_to_Market_First_to_Fail_Real_Causes_of_Enduring_Market_Leadership

[3] https://www.nobelprize.org/prizes/physics/1991/summary/

[4] https://www.nsf.gov/pubs/2019/nsf19556/nsf19556.pdf

[5] http://www.soilhealthinstituteresearch.org/Home/Search

[6] http://www.soilhealthinstituteresearch.org/Content/DataDictionaryv2.pdf

[7] http://bit.ly/DATA-AA

[8] https://canvas.harvard.edu/files/3486763/download?download_frd=1&verifier=qRAG5jUIbng83WzEl8BnQQop9GgeXfw0hCkeMYiJ

[9] http://www.cs.utsa.edu/~bylander/cs1023/SRD-IBM-Entity-Analytics-9-15-05.ppt

[10] https://jeffjonas.typepad.com/jeff_jonas/2010/07/hell-with-rules.html

[11] https://www.nap.edu/download/25059

[12] http://www.matshore.com/849-2/

[13] https://onlinelibrary.wiley.com/doi/full/10.1111/cei.13195

[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

[17] https://www.techbriefs.com/dl/2019.php?i=34095&eid=376590803&bid=2405616&d=1

[18] https://arxiv.org/pdf/1805.04011.pdf

[19] https://people.csail.mit.edu/fadel/papers/wivi-paper.pdf

[20] http://eqradio.csail.mit.edu/files/eqradio-paper.pdf

[21] https://smap.jpl.nasa.gov/files/smap2/c1_Njoku_Entekhabi.pdf

[22] http://smap.jpl.nasa.gov/files/smap2/SMAP_Apps_Plan_120706.pdf

[23] https://smap.jpl.nasa.gov/system/internal_resources/details/original/484_L2_SM_P_ATBD_rev_D_Jun2018.pdf

[24] https://www.microsoft.com/en-us/research/uploads/prod/2018/10/SMURF_TR-1.pdf

[25] https://www.youtube.com/watch?v=BCIohEJYxdY

[26] http://bit.ly/WSN-MASSEY

[27] http://robotics.sciencemag.org/content/4/27/eaaw1875

[28] http://bit.ly/SIN-BTH

[29] https://www.nature.com/articles/s41928-018-0173-2

[30] http://www.wave-scattering.com/pbgheadlines.html

[31] https://apps.dtic.mil/dtic/tr/fulltext/u2/601325.pdf

[32] https://polymerscience.physik.hu-berlin.de/docs/manuals/dielectric.pdf

[33] https://arxiv.org/ftp/cond-mat/papers/0612/0612570.pdf

[34] http://jgp.rupress.org/content/jgp/44/1/19.full.pdf

[35] http://bit.ly/SARS-AG

[36] https://pdfs.semanticscholar.org/f363/6deed27227fa5a52603ccf16a5127f48f085.pdf

[37] https://pdfs.semanticscholar.org/839b/dbc3ce47df32b04b9722aa492d51b3921268.pdf

[38] https://core.ac.uk/download/pdf/61188787.pdf

[39] https://www.lehigh.edu/imi/teched/JapanWS/WS11AB_Lanagan.pdf

[40] https://www.mdpi.com/2072-4292/6/10/9712/pdf

[41] http://onlinepubs.trb.org/Onlinepubs/trr/1992/1355/1355-012.pdf

42 https://www.mdpi.com/1424-8220/18/10/3333
43 https://arxiv.org/pdf/quant-ph/0411002.pdf
44 https://arxiv.org/pdf/hep-ph/0406008.pdf
45 https://dspace.mit.edu/openaccess-disseminate/1721.1/69931
46 https://www.ncbi.nlm.nih.gov/pubmed/14021529
47 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5896772/pdf/nihms955832.pdf
48 https://www.ncbi.nlm.nih.gov/pubmed/17644693
49 https://www.amjmed.com/article/S0002-9343(99)00055-8/fulltext
50 https://www.ias.ac.in/article/fulltext/reso/021/08/0741-0747
51 https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.100501
52 https://arxiv.org/pdf/1604.01677.pdf
53 http://web.mit.edu/ashokaj/www/
54 https://arpa-e.energy.gov/?q=slick-sheet-project/magnetic-resonance-imaging-root-growth
55 https://pubs.rsc.org/en/content/articlelanding/2017/lc/c7lc00192d#!divAbstract
56 https://www.mdpi.com/1424-8220/17/8/1866
57 http://stanford.edu/group/khuri-yakub/publications/94_Haller_01.pdf
58 https://iopscience.iop.org/article/10.1088/0960-1317/21/5/054004/meta
59 https://www.technologyreview.com/s/532166/with-100-million-entrepreneur-sees-path-to-disrupt-medical-imaging/
60 http://repository.bilkent.edu.tr/bitstream/handle/11693/27202/bilkent-research-paper.pdf?sequence=1&isAllowed=y
61 https://erj.ersjournals.com/content/46/suppl_59/OA301
62 https://patentimages.storage.googleapis.com/7c/77/4e/867b47a5072065/US8689606.pdf
63 https://www.nature.com/articles/s41467-018-08038-4
64 https://www.mdpi.com/2079-6374/6/4/60
65 https://www.nist.gov/news-events/news/2019/02/floc-takes-flight-first-portable-prototype-photonic-pressure-sensor
66 http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf
67 https://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf
68 https://www.jstor.org/stable/1912791
69 http://www.econ.uiuc.edu/~econ536/Papers/engle82.pdf
70 https://www.jstor.org/stable/1913236
71 https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.080504
72 https://www.nature.com/articles/267585a0
73 https://www.nobelprize.org/prizes/chemistry/2013/karplus/facts/
74 http://orca.cf.ac.uk/58996/
75 https://www.nature.com/articles/415152a
76 https://link.springer.com/article/10.1007/s11104-007-9514-z
77 https://jb.asm.org/content/jb/2/1/35.full.pdf
78 https://link.springer.com/chapter/10.1007/978-981-10-5708-3_7
79 https://www.nature.com/articles/s41598-018-20256-w.pdf
80 https://jcm.asm.org/content/39/5/1840
81 https://aem.asm.org/content/aem/72/3/1719.full.pdf
82 http://bit.ly/EDWARD-TUFTE
83 https://pubs.rsc.org/en/content/articlelanding/2018/an/c8an00065d#!divAbstract
84 https://www.law.uchicago.edu/files/file/coase-nature.pdf
85 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC440936/pdf/bactrev00182-0090.pdf
86 http://docsdrive.com/pdfs/ansinet/biotech/2010/428-443.pdf
87 https://smartech.gatech.edu/bitstream/handle/1853/55276/g-32-629_399496.pdf

88 https://pubs.acs.org/doi/abs/10.1021/jf025660x
89 https://doi.org/10.3390/ma8125478
90 https://link.springer.com/content/pdf/10.1186%2Fs11671-018-2772-2.pdf
91 https://jb.asm.org/content/jb/190/13/4706.full.pdf
92 https://www.jstor.org/stable/3625499
93 https://naldc.nal.usda.gov/download/IND43966191/PDF
94 https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.1890.0005
95 https://www.frontiersin.org/articles/10.3389/fmicb.2016.00529/full
96 https://www.nature.com/articles/334340a0
97 https://academic.oup.com/femsle/article/272/2/127/637339
98 https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1365-2958.2006.05057.x
99 https://pubs.acs.org/doi/10.1021/nn2042384
100 https://smartech.gatech.edu/handle/1853/45965
101 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1085052/pdf/pnas01891-0031.pdf
102 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1085053/pdf/pnas01891-0035.pdf
103 https://core.ac.uk/download/pdf/55640662.pdf
104 https://www.springer.com/us/book/9783662530665
105 https://pubs.rsc.org/en/content/articlelanding/2016/qi/c5qi00240k
106 https://www.mdpi.com/journal/chemosensors/special_issues/nanozyme-based-sensors
107 https://doi.org/10.1016/B978-0-323-50878-0.00007-0
108 https://doi.org/10.1016/j.yrtph.2012.06.016
109 https://www.statnews.com/sponsor/2018/02/22/clean-laundry-enzyme-science/
110 https://onlinelibrary.wiley.com/doi/abs/10.1002/cber.18940270364
111 https://www.medschool.lsuhsc.edu/biochemistry/docs/Lichtenthaler.pdf
112 https://www.nature.com/articles/432447a
113 http://www.sussexvt.k12.de.us/science/The%20Periodic%20Table/Seaborg.htm
114 https://arxiv.org/ftp/physics/papers/0411/0411080.pdf
115 https://link.springer.com/article/10.1023/A:1025673206850
116 https://pdfs.semanticscholar.org/0379/02a15832359ff6527db3e949ceadb8d2dd16.pdf
117 https://www.economist.com/science-and-technology/2019/02/28/the-periodic-table-is-150-years-old-this-week
118 https://doi.org/10.1073/pnas.0604452103
119 https://doi.org/10.1016/S0022-2836(63)80091-1
120 http://www.unige.ch/sciences/biochimie/Edelstein/Monod%20Wyman%20Changeux%201965.pdf
121 https://doi.org/10.1016/j.jmb.2013.02.033
122 https://www.physiology.org/doi/pdf/10.1152/ajplung.00527.2018
123 https://pubs.acs.org/doi/abs/10.1021/acsphotonics.8b00440
124 https://www.osapublishing.org/boe/abstract.cfm?uri=boe-9-11-5735
125 https://www.frontiersin.org/articles/10.3389/fenvs.2014.00007/full
126 www.researchgate.net/publication/222791698_Soil_fauna_guilds_functional_groups_and_ecosystem_processes
127 https://www.wikipremed.com/reading/philosophy/The_Life_of_Reason.pdf
128 https://www.electronicdesign.com/defense/1950s-transistors-fill-vacuum-digital-age-begins
129 http://www.research.ibm.com/articles/brain-chip.shtml
130 https://www.nextgov.com/emerging-tech/2015/02/ibm-tries-sell-government-brain-box-technology/105975/
131 https://clintonwhitehouse3.archives.gov/WH/New/html/20000626.html
132 https://biology.mit.edu/profile/eric-s-lander/

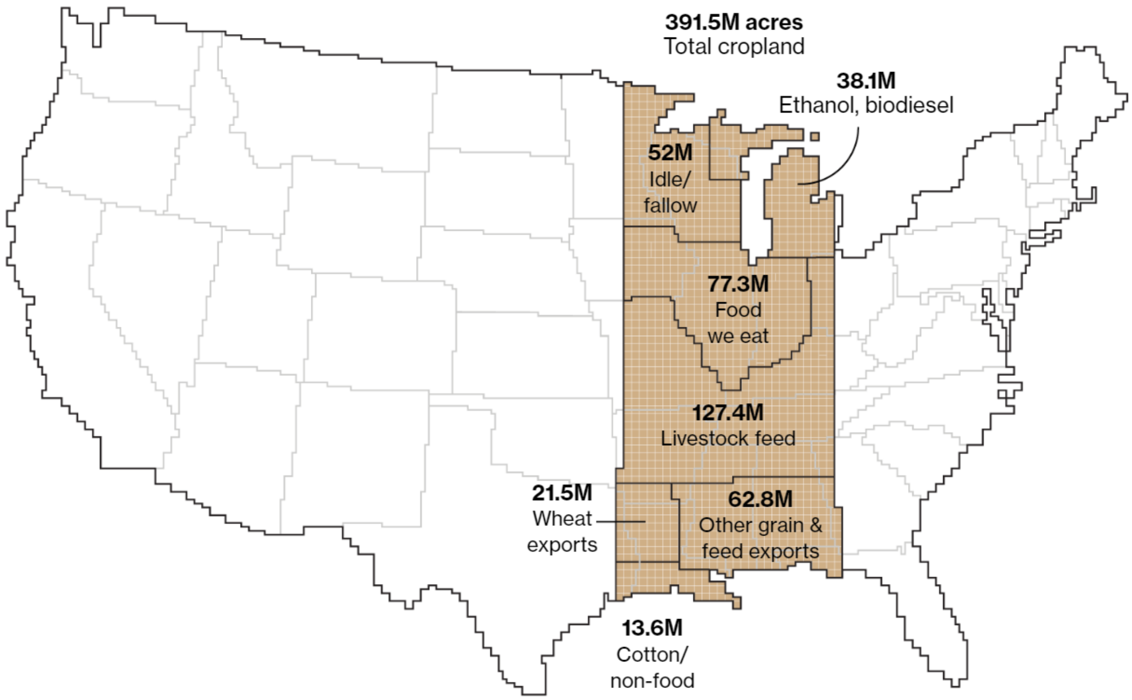
133 https://eccc.weizmann.ac.il/report/2018/107/
134 https://www.pnas.org/content/114/29/7555
135 https://www.nature.com/articles/s41598-019-40765-6
136 https://www.consulting.us/news/1485/accenture-wins-quantum-computing-related-patent
137 https://www.jstor.org/stable/e24931041
138 https://www.siliconrepublic.com/machines/marvin-minsky-ai-predictions
139 https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf
140 https://www.w3.org/
141 https://www.researchgate.net/publication/237288554_Knowledge_Web_Technology_Roadmap_The_Technology_Roadmap_of_the_Semantic_Web
142 www.newscientist.com/article/mg21228390-800-nothingness-the-turbulent-life-of-empty-space/
143 https://rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/
144 https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/
145 https://www.telegraph.co.uk/technology/2018/07/27/ibm-watson-ai-criticised-giving-unsafe-cancer-treatment-advice/
146 https://www.healthnewsreview.org/2017/02/md-anderson-cancer-centers-ibm-watson-project-fails-journalism-related/
147 http://www.cs.virginia.edu/~robins/The_Rise_of_AI.pdf
148 https://www.advisorperspectives.com/commentaries/2019/02/23/ai-will-add-15-trillion-to-the-world-economy-by-2030
149 http://sprott.physics.wisc.edu/Chaos-Complexity/allard17.pdf
150 https://www.deeplearningbook.org/
151 http://aima.cs.berkeley.edu/
152 http://www.acad.bg/ebook/ml/Society%20of%20Mind.pdf
153 https://arxiv.org/ftp/arxiv/papers/1610/1610.07862.pdf
154 https://www.nature.com/news/when-google-got-flu-wrong-1.12413
155 https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/babbage/diff.htm
156 https://www.gao.gov/assets/680/677454.pdf
157 https://www.chistlukeshealth.org/sites/default/files/CMS%20Survey.pdf
158 https://www.govinfo.gov/content/pkg/FR-2019-02-15/pdf/2019-02519.pdf
159 https://www.bmj.com/content/353/bmj.i2139
160 https://dutraeconomicus.files.wordpress.com/2014/02/amartya-sen-the-idea-of-justice-2009.pdf
161 http://www.ai.mit.edu/projects/oxygen/oxygen-book2001/oxygenbook2001.pdf
162 http://oxygen.csail.mit.edu/videosketching.html
163 http://oxygen.csail.mit.edu/Overview.html
164 http://www.ai.mit.edu/projects/oxygen/oxygen-jan02-slides/
165 http://oxygen.csail.mit.edu/videomultilingual.html
166 https://www.siriuserguide.com/article/how-to-check-the-weather-forecast-with-siri
167 http://bit.ly/IOT-METAPHOR
168 http://bit.ly/SARS-AG
169 https://www.dre.vanderbilt.edu/~schmidt/PDF/dds-sos.pdf
170 https://www.omg.org/spec/DDS/About-DDS/
171 http://www.laas.fr/files/SLides-A_Corsaro.pdf
172 https://www.sciencedirect.com/science/article/pii/0304407681900798
173 https://www.jstor.org/stable/1913236

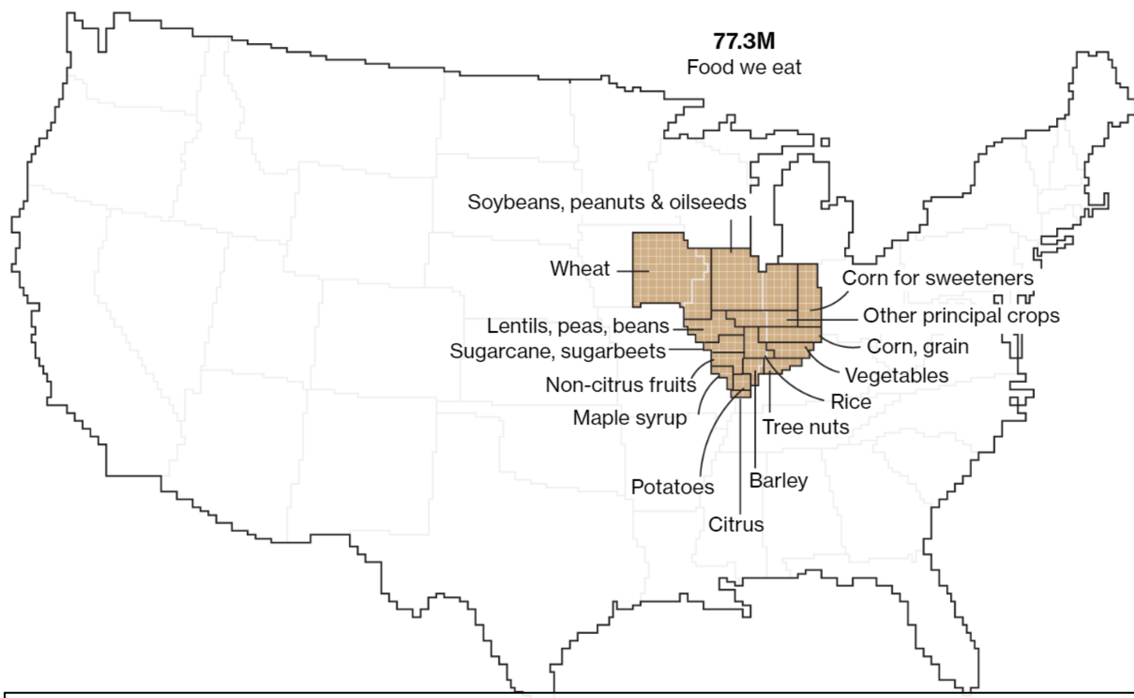
[174] http://www-stat.wharton.upenn.edu/~steele/Courses/434/434Context/Co-integration/GrangerNobelAddress.pdf
[175] http://unitsofmeasure.org/ucum.html
[176] https://chomsky.info/wp-content/uploads/195609-.pdf
[177] https://www.w3.org/standards/semanticweb/ontology
[178] https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
[179] http://xml.coverpages.org/pml-ons.html
[180] https://www.w3.org/RDF/Metalog/docs/sw-easy
[181] https://www.w3.org/RDF/
[182] https://dspace.mit.edu/handle/1721.1/111021
[183] https://dspace.mit.edu/handle/1721.1/41902
[184] https://www.gs1.org/standards/epc-rfid
[185] https://www.gs1.org/standards/barcodes/ean-upc
[186] https://www.forbes.com/sites/cognitiveworld/2019/03/12/taxonomies-ontologies-and-machine-learning-the-future-of-knowledge-management
[187] https://www.tandfonline.com/doi/abs/10.1080/2287108X.2012.11006070
[188] https://www.economics.ox.ac.uk/materials/working_papers/paper084.pdf
[189] https://pdfs.semanticscholar.org/02be/4bd7dc9f91a20fc522edd041ac6708889dc7.pdf
[190] https://warwick.ac.uk/fac/soc/economics/research/workingpapers/1989-1994/twerp339.pdf
[191] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.2728&rep=rep1&type=pdf
[192] https://www.tandfonline.com/doi/pdf/10.1080/13102818.1997.10818934
[193] https://www.cs.le.ac.uk/people/terlebach/talks/haifa06internet.pdf
[194] https://www.mdh.se/polopoly_fs/1.34346!/Menu/general/column-content/attachment/MAA507_L1_vt13.pdf
[195] http://snap.stanford.edu/class/cs224w-readings/broder00bowtie.pdf
[196] https://neo4j.com/developer/graph-database/
[197] http://dx.doi.org/10.4135/9781526410993
[198] https://www.youtube.com/watch?v=6Z_XVbECJzw
[199] https://www.fusionbox.com/blog/detail/graph-algorithms-in-a-database-recursive-ctes-and-topological-sort-with-postgres/620/
[200] http://people.apache.org/~dongsheng/horak/100309_dag_structures_sql.pdf
[201] https://hackernoon.com/which-gpu-database-is-right-for-me-6ceef6a17505
[202] https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html
[203] http://bit.ly/SENSOR-LIBRARY-ERIC-MCLAMORE
[204] http://www.jmaxkanter.com/static/papers/DSAA_DSM_2015.pdf
[205] http://bit.ly/BOOKS-FEATURES
[206] https://aioti.eu/wp-content/uploads/2017/03/AIOTIWG06Report2015-Farming-and-Food-Safety.pdf
[207] https://ieeexplore.ieee.org/document/8081906
[208] https://outlyer.com/blog/top10-open-source-time-series-databases/
[209] https://aip.scitation.org/doi/full/10.1063/1.4999209
[210] http://science.sciencemag.org/content/363/6430/912
[211] http://science.sciencemag.org/content/363/6430/917
[212] https://eng.uber.com/aresdb/
[213] https://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Programming_Guide.pdf
[214] http://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-001.pdf

215 www.technologyreview.com/s/601013/the-internet-of-things-roadmap-to-a-connected-world/
216 http://eyeriss.mit.edu/
217 https://www.nature.com/articles/s41598-018-36973-1
218 http://ciml.250x.com/archive/communists/brecht/english/brecht_the_threepenny_opera.pdf
219 https://www.earlyprintedbooks.com/ames_typographical-antiquities_1749_a1r/
220 https://er.jsc.nasa.gov/seh/ricetalk.htm
221 http://www.css.cornell.edu/extension/soil-health/manual.pdf
222 https://eng.uber.com
223 http://www1.cs.columbia.edu/~ji/F02/ir02/p44-perlman.pdf
224 https://www.bloomberg.com/graphics/2018-us-land-use/

RADIA PERLMAN[223]

```
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.

“SIGNALS” contains a series of essays spewing amorphous thoughts:

1. SITS
2. SIP-SAR -------- **You are here**
3. SARS♠AG
4. ART
5. PEAS

Please review “SIGNALS”

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)

“IoT is a Metaphor” (see Commentary E)

Please review “IoT is a Metaphor”

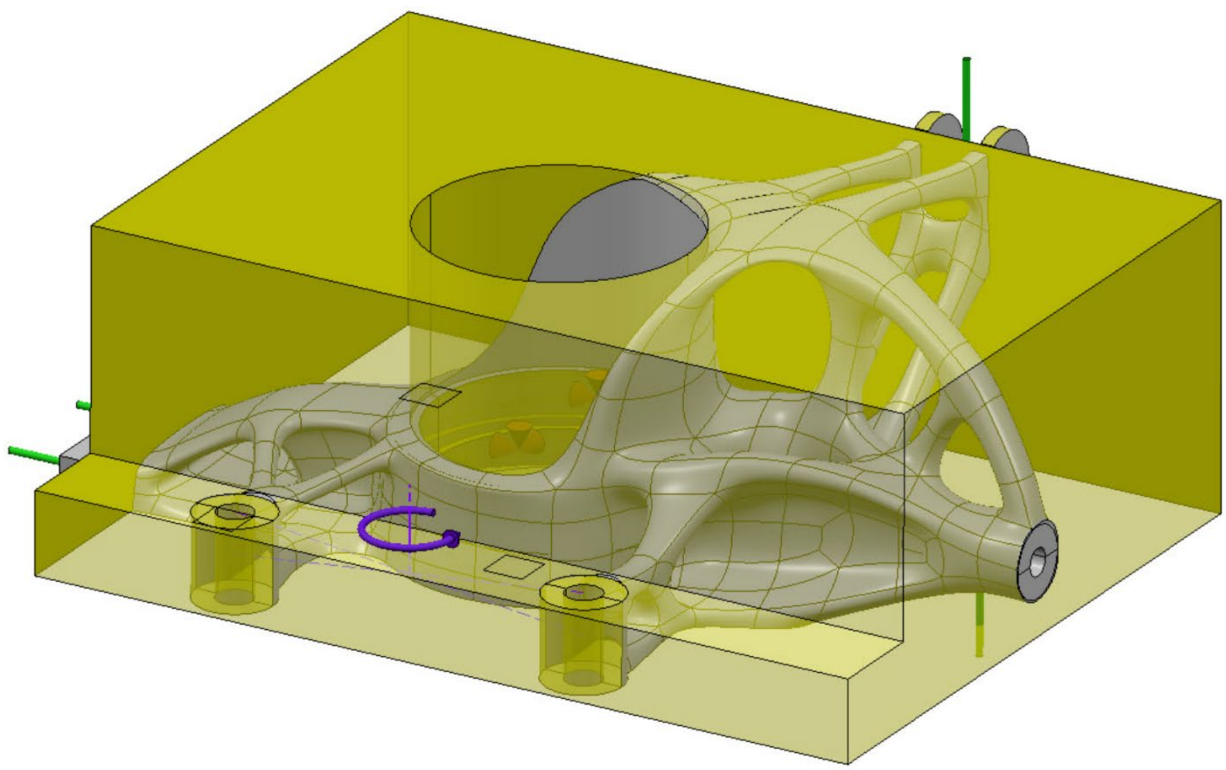
● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.

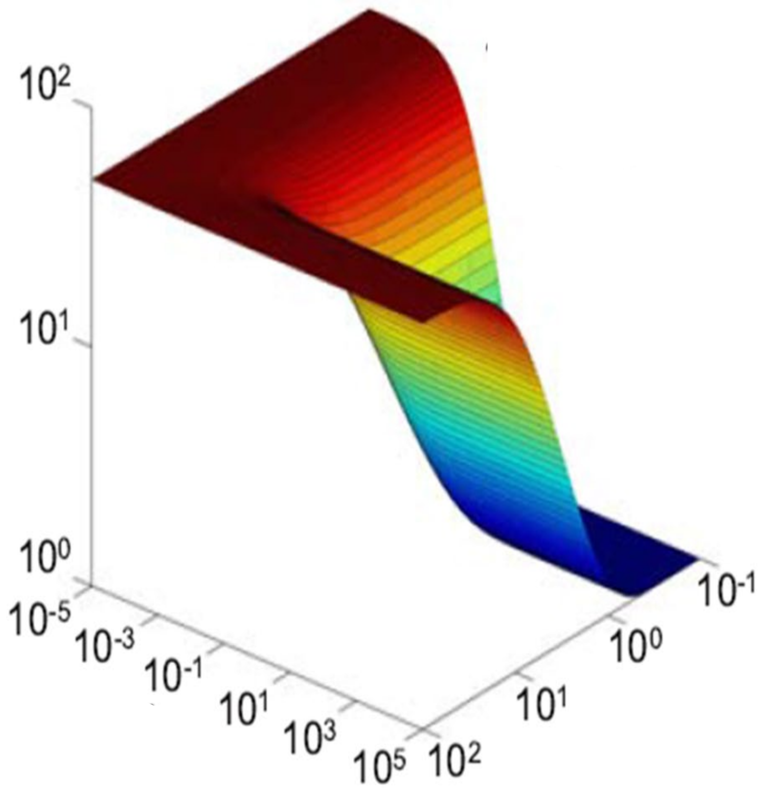
RELEVANCE OF SIP-SAR IN CONTEXT OF "SIGNALS"

The unexpected interjection of product lifecycle management (PLM) related to machines in "SIGNALS" is based on the fact that machine tools play a critical role in the agro-ecosystem. Data from machines, devices and tools (which are not sensors in soil or water) must be combined in decision support systems, to extract the full picture. In addition, sensors in machines, devices and tools are also generating data, which are valuable for the operation of the system.



SIPSAR

Shoumen Palit Austin Datta
Email - shoumen@mit.edu



Platform for Synergistic Integration of Predictive Maintenance of Machines with Maintenance and Repair Supply Networks: Virtual Mobile Dashboard for Near Real-Time Transparency and Operations Management Optimization

Shoumen Datta, Auto ID Labs, Massachusetts Institute of Technology and MDPnP Lab, Massachusetts General Hospital, Harvard Medical School

PREFACE

In the US, a quarter million patients die each year due[1] to preventable medical errors. One leading cause is the lack of medical device interoperability[2] 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 advocate 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.

These suggestions are neither unique nor novel. The strength of this note is grounded in an *open-source, modular* approach, which supports scalability, interoperability and real-time communications 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 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.i2139
[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052027/

FRAMEWORK

The concept of ***synergistic integration platform***, may be illustrated by the old telephone switchboard (Fig 1) as well as **PEAS** (last item in this PDF). The “plug-n-play” input resembles the potential to aggregate data streams from sensors, actuators, RFID tags and other tools. In the “logic tools” 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[3] and other virtual machines, which may be on-site or sourced, *ad hoc*, from clouds, fog, or mist computing repertoires. Automated programming[4] and feature engineering[5] may be included with ML.

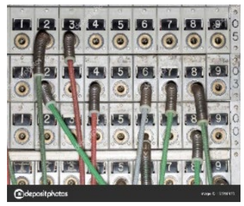


Fig 1: Old telephone switchboard – conceptually analogous to future ‘plug-n-play’ synergistic integration platforms.

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[6] 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.

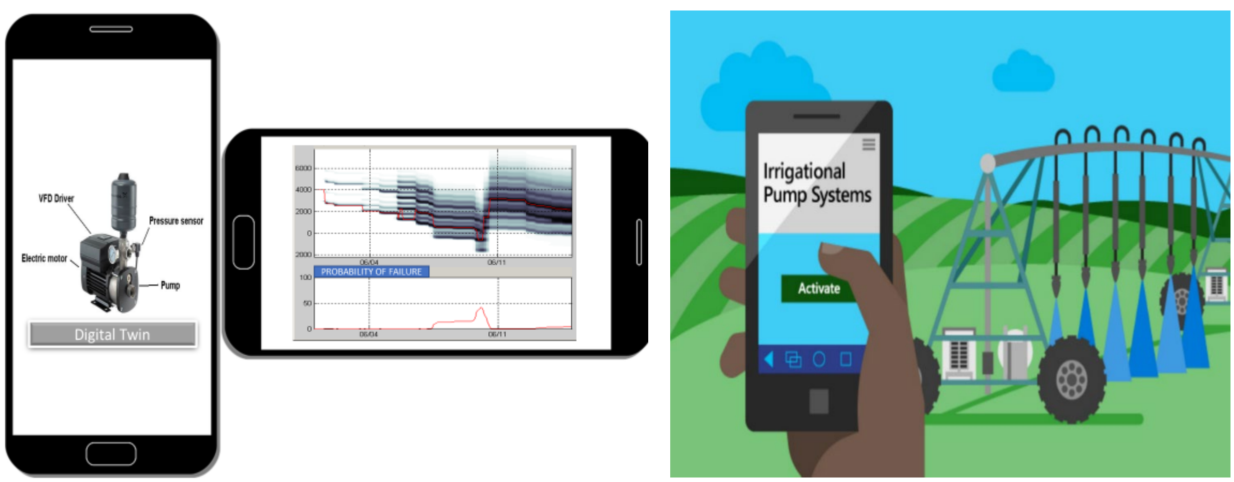


Fig 2: Monitoring mean time between failure (MTBF) from sensor data on a mobile smartphone (L). With decreasing residual lifecycle (upper panel) the probability of failure (broken part may precipitate unplanned downtime) increases (lower panel). Just in time spare parts supply chain coordination with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

[3] https://ieeexplore.ieee.org/document/7506647/
[4] https://arxiv.org/pdf/1807.02816.pdf
[5] https://www.featurelabs.com/wp-content/uploads/2017/12/DSAA_DSM_2015.pdf
[6] http://dspace.mit.edu/handle/1721.1/111021

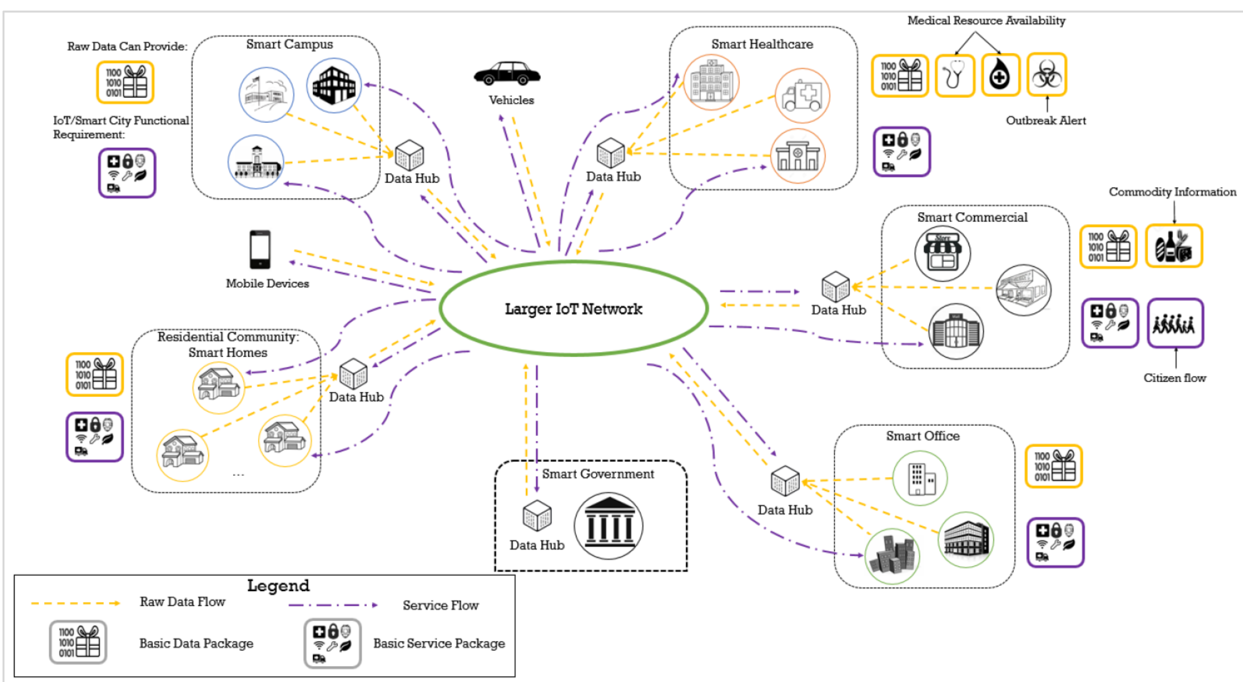
replaced. This example of an agent-based system uses MTBF metrics and ML to guide preventive maintenance. (Right) Cartoon of a partially automated irrigation system (see last PDF - PEAS Platform for the Agro-Ecosystem).

IMMENSE SCOPE – SENSE & RESPONSE

Data from sensors integrated into[7] sensing-compatible components of machines (eg belts, hoses, seals, pumps) 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 (eg Wabash), farm operations (eg Fair Oaks Farms) and robotics (eg Automotive Robotics). The scope may extend to system of systems including machines, buildings, irrigation and emergency response systems. In other words, 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[8] 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[9] are making better use of connectivity.

One purpose of this note is to identify gaps which may be addressed through research to improve the connectivity and, in turn, productivity. In parallel, if formed, PPP may 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[10] where the duality of the physical, and digital, converges into a fluidic continuum, at least theoretically, in power-point presentations.



[7] https://dl.acm.org/citation.cfm?doid=3131672.3131702
[8] http://bit.ly/CDAIT-IoT
[9] https://newsroom.mastercard.com/press-releases/mastercard-introduces-mastercard-track-to-make-the-business-of-doing-business-easier/
[10] http://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-001.pdf

Fig 3: IoT is a metaphor without borders. It represents[8] a continuum made up of discrete objects, decisions, processes and outcomes, each of which, in turn, impacts and influences, man, woman and machines, life and living.

CYBERMANUFACTURING

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[11] to riches story[12] 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 Mars or on a military[13] aircraft carrier). 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[14] 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[15] that may be a partially hollow mesh integrated with sensors (gyroscopic sensor, glucose sensor, CTx-1 collagen sensor[16] 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 Austin, TX as well as the less fortunate people in Addis Ababa, Ethiopia.

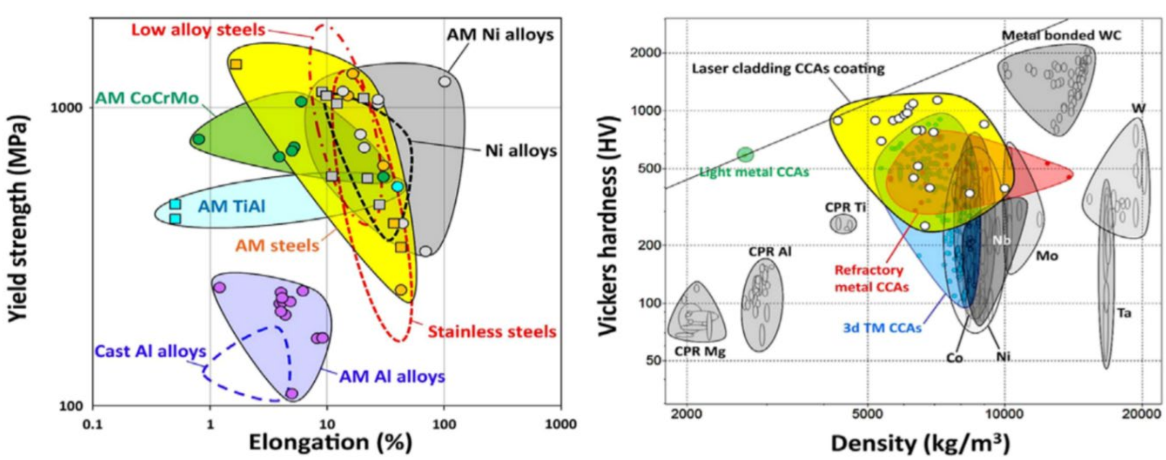


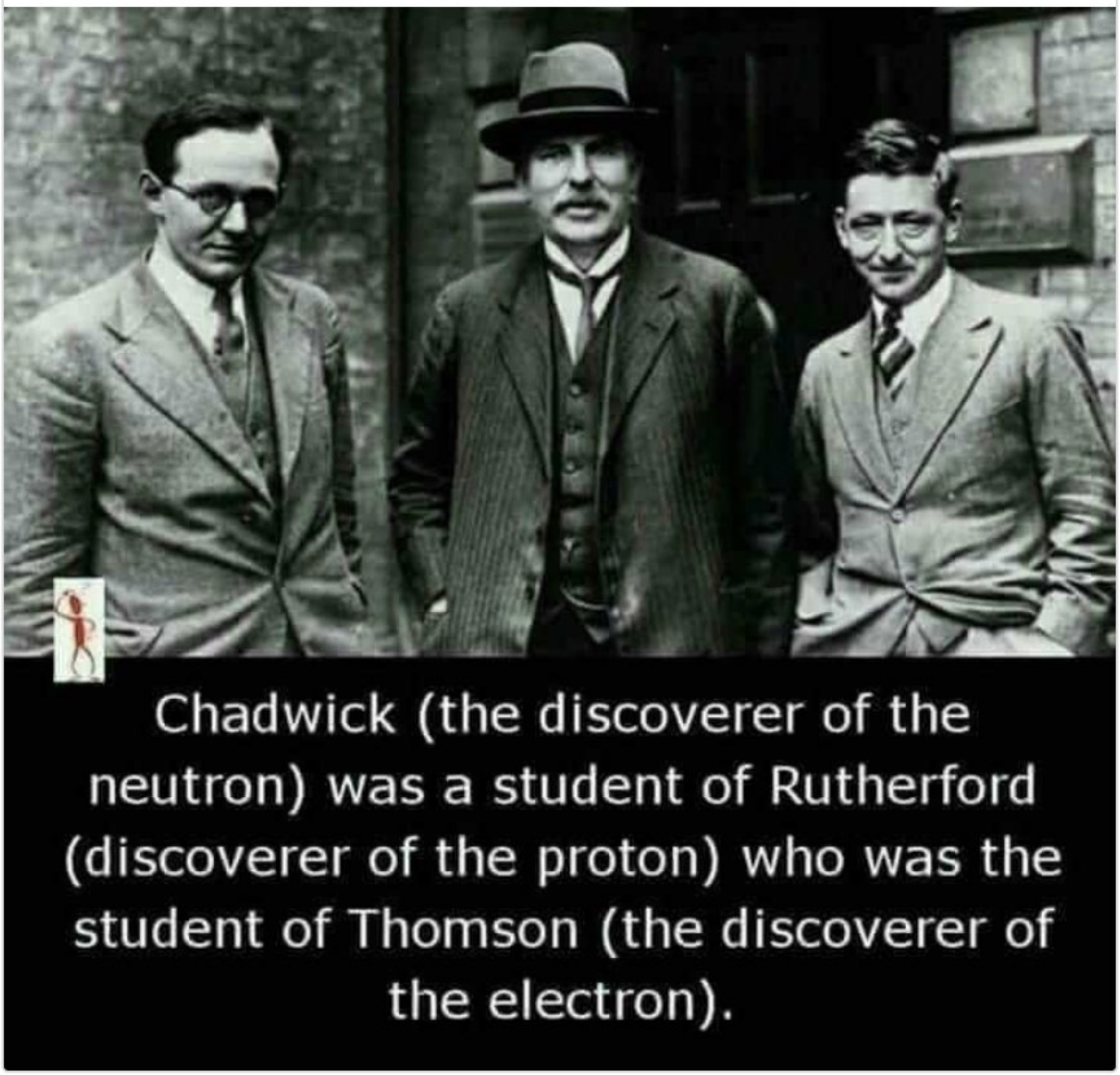
Fig 4: Mechanical properties of additive manufactured alloys[15] compared to conventionally processed counterparts.

[11] https://bases-brevets.inpi.fr/fr/document/FR2567668/publications.html
[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
[15] https://www.tandfonline.com/doi/full/10.1080/14686996.2017.1361305
[16] http://www.mdpi.com/2224-2708/7/1/10

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

TRANSCENDENTAL CONNECTIVITY



"SIGNALS" contains a series of essays spewing amorphous thoughts:

1. SITS
2. SIP-SAR
3. SARS♠AG -------- **You are here**
4. ART
5. PEAS

Please review "SIGNALS"

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)

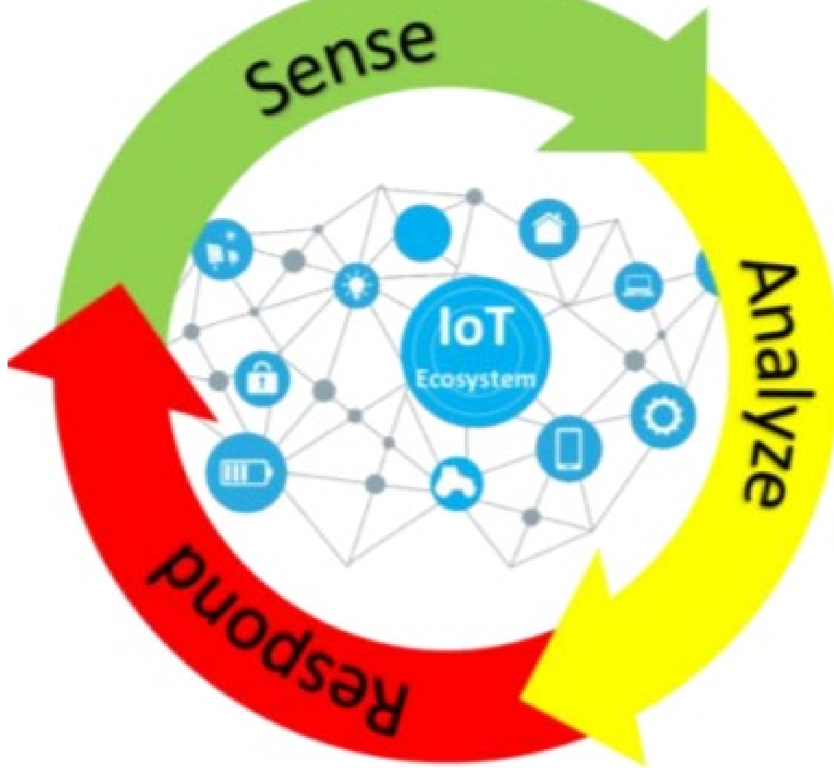
"IoT is a Metaphor" (see Commentary E)

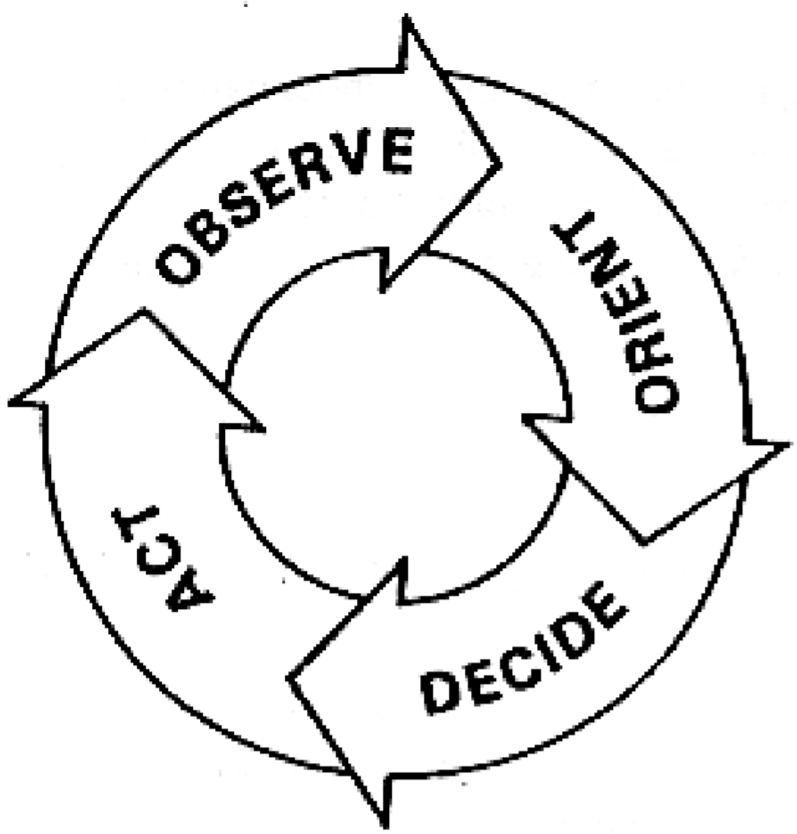
Please review "IoT is a Metaphor"

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.


Sense
Analyze
Respond
IoT
Ecosystem


OBSERVE
ORIENT
DECIDE
ACT

SARS ♦ AG

Preface to the PEAS Platform for the Agro-Ecosystem ● Essay by Shoumen Palit Austin Datta (shoumen@mit.edu)

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 if user-centric design delivers value for users. However, often people do not know what they need. Hence, user-centricity is a suggestion, not a doctrine. SARS for ag (agriculture) advocates ideas with the end user in mind.

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

SARS is **not** a paradigm shift. It is an intrinsic concept which is feasible due to decreasing cost of computation, storage, telecommunications and systems. The principle of SARS existed in parallel with human thought, procedural sequencing and the organization of logic. Feeling the water temperature of the sea water by dipping a finger is a form of sensing. The "sense" of the temperature (analysis), influences the response (decision to swim or not). 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 musculoskeletal 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 vast volumes of data. Depending on complexity, the tool-based outcome feeds decision support systems. Humans in the loop, generally, 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 to your physician or the smoke alarm in your home. SARS in the context of internet of things (IoT[2]) may be machine maintenance[3] 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 and extract (hopefully) 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
[3] http://bit.ly/SIP-SAR is now a part of http://bit.ly/SARS-AG
[4] https://arxiv.org/abs/1610.06467

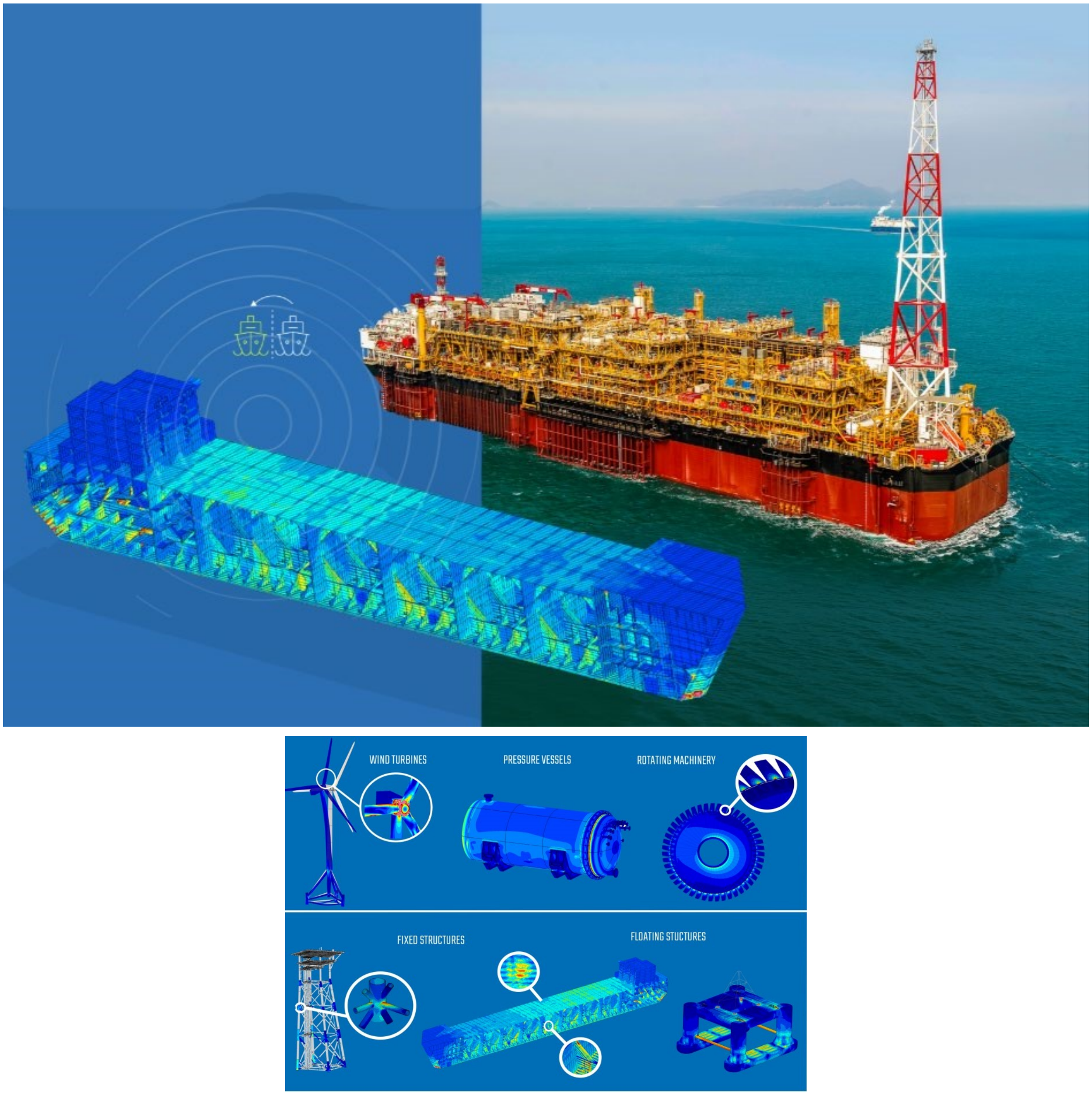


Figure 0: The concept of SARS is embedded in the metaphor of Digital Twins. Shell's Floating Production Storage Offloading unit (Bonga FPSO, top, has a storage capacity of 2 million barrels of crude oil). It is a digital twin project (left cartoon, top panel) executed by Akselos, a MIT spinout, a high-tech 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♦AG – 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 essay refers to a proof of concept (PoC), which may be achieved using a sub-set of data (for example, see **SENSEE**, sensor data repository). The limited scope of the PoC is discussed in **ART** and the bigger picture is in **PEAS** (in this collection).

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 datastreams, from a variety of sources (including, environmental data, and, 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 **"logic tool"** for agriculture must be an open **platform**. Other groups and consortia can add to this data platform, leading to growth of the tool for specific or general purposes.

SARS PLATFORM – TOOL DESIGN AND PROOF OF CONCEPT TASK

The PEAS platform (PoC sub-set SENSEE) is expected to connect different components, subsystems and domains[7] to add value to the outcome (pragmatic solution). 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 instruction. The description of the outcome starts with the end user. The outcome, the end user anticipates, is the "deliverable" scenario.

The design of the SARS logic tool starts from the end, that is, the outcome in the context of the user's question. 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 re-assembled, synthesized and harmonized, creates a functional system or a solution which may involve 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.

---

[5] https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/

[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, for example: [a] "blood glucose = 157mg/dl" "what does it mean?" [b] "mercury in water = 85 micrograms per liter" "suggest water treatment" for decontamination.

In future, we may move forward by going back, almost a hundred years. SARS proactive system push ("Hello Rebecca – I have new soil spectroscopic[8] sensor data, click on this link"). Bioinspired "cybernetics[9] of ag" is bidirectional, discussed by Weiner[10] in the context of sensors at the edge of the eye, which vie for the attention of the analytical core in the *fovea centralis*. SARS PoC may focus on [a] water/soil sensors and [b] water treatment[11]. In that context the platform rests on two key sources of data (data issues are discussed in SENSEE 2.0 in ART).

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[12] algorithms, feature[13] 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 simplification of the PEAS PLATFORM metaphor and potential connectivity to data (graph databases or datastores may be real or virtual, that is, sourced from cloud, fog or mist). Platform "aggregation" on 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
9 https://uberty.org/wp-content/uploads/2015/07/Norbert_Wiener_Cybernetics.pdf
10 https://libraries.mit.edu/archives/research/collections/collections-mc/mc22.html
11 https://oaspub.epa.gov/tdb/pages/general/home.do
12 https://arxiv.org/pdf/1807.00401.pdf
13 https://dai.lids.mit.edu/wp-content/uploads/2017/10/1407.5238v1.pdf

The view of the 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 platform tool may be created and delivered with only two sources of data, illustrated by two red strings (illustration source[14]).

The software architecture of the open platform, its connectivity to data and access to cloud storage, may be designed to embody the broad-spectrum vision of the 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-VO[15] model.

Open architectures may enhance or catalyze “plug and play” operation when integrating other data sources (for example, weather, crop, fertilizers) from different clouds, to feed the 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 very limited types of questions in narrow domains, for example, [a] water/soil sensor data and [b] wastewater treatment protocols.

The **potential** for dynamic composability needs an open architecture which can act as a **scalable scaffold**. The nature of data, dependencies between data, and application logic, that is suitable to connect / mine / combine, to generate useful information, are software functions which can be added/subtracted to the application design environment. Modularity and fluidity demands the use of agent-based systems (ABS) and agent-based architectures[16] in the platform.

[14] https://www.miliashop.com/es/mesas/4187-org-mesa-rectangular-cappellini.html
[15] https://phys.org/news/2011-04-global-portal-cyber-physical.html
[16] https://dl.acm.org/citation.cfm?id=122367

User adoption may be directly proportional to the functional strength of the logic tool in delivering the **value of the output**. The strength 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 from ART to DIDA'S to KIDS to PEAS platform (see other PDFs). Knowledge as a key performance indicator (KPI) for the PEAS platform may 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 (and attempt to approach knowledge-informed decisions).

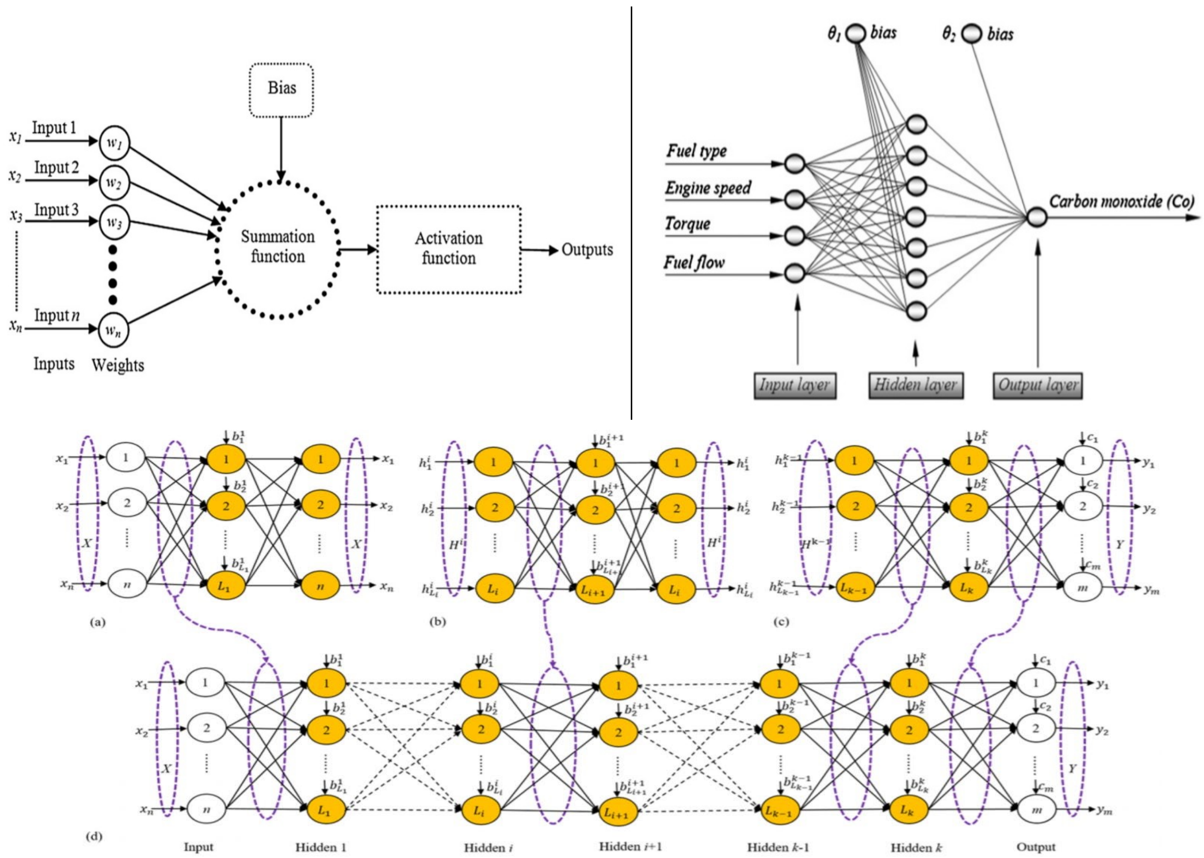


Fig 3: Neural nets optimize outputs[20] (above) and deep ***bidirectional*** learning machines[21]

---

[17] https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html
[18] https://patents.google.com/patent/US20180170575A1/en
[19] http://dx.doi.org/10.1016/j.energy.2012.10.052
[20] 10.1021/acs.energyfuels.7b01415
[21] https://arxiv.org/pdf/1404.7828.pdf
[22] http://www.jmaxkanter.com/static/papers/DSAA_DSM_2015.pdf

SARS PLATFORM – PHYSICAL COMPONENTS AND RELEVANCE TO CYBERPHYSICAL SYSTEMS

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 (sensor data). Therefore, SARS is also a part of the internet of things (IoT) metaphor. The platform may benefit from CPS[25] and IoT[26] architectures, and generic[27] framework. 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 IoT connectivity between systems of systems (internet of systems). The SARS platform may immensely benefit from the CPS 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.

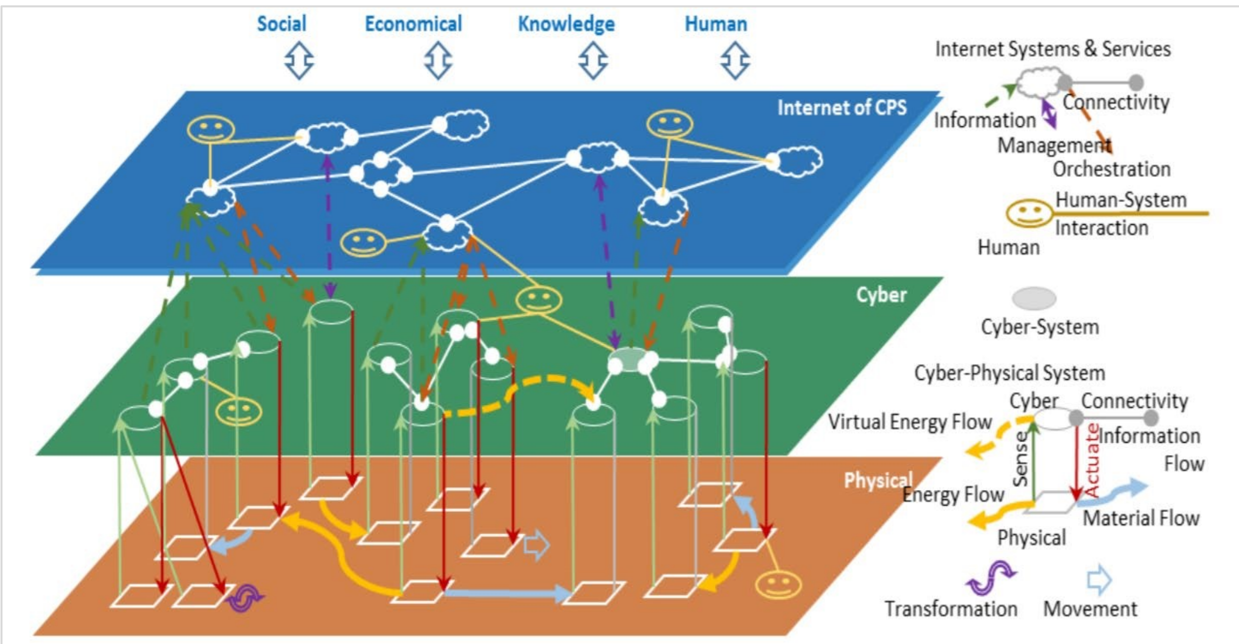


Figure 4: Architectural framework[25] for cyberphysical systems proposed by NIST. Most platform paradigms (for example, digital twins, SARS) may be a subset[29] of CPS (cyberphysical systems).

[23] https://doi.org/10.1007/s10845-018-1433-8
[24] https://cps-vo.org/
[25] https://s3.amazonaws.com/nist-sgcps/cpspwg/files/pwgglobal/CPS_PWG_Framework_for_Cyber_Physical_Systems_Release_1_0Final.pdf
[26] https://iotforum.org/wp-content/uploads/2014/09/D1.5-20130715-VERYFINAL.pdf
[27] https://arxiv.org/ftp/arxiv/papers/1708/1708.04560.pdf
[28] https://www.nccoe.nist.gov/
[29] https://ptolemy.berkeley.edu/projects/cps/

PLATFORMS DEPEND ON DATA

A cursory survey[30] of types of sensors (bundled by loose similarity[31]) is overwhelming, simply due to volume[32] and variability. The inordinate complexity of sensor types in agriculture (weather, water, soil, pathogens, chemicals) presents a curse of dimensionality[33] with respect to signals. The latter presents an almost insurmountable barrier to make ***<u>collective</u>*** sense of data, from different sensors and network of sensors.

One approach to streamline the quagmire is a call for protocols[34] and standards[35] which offers economic[36] incentive. Agnostic of verticals, standards accelerate[37] adoption if industry values such non-competitive agreements. This effort may not possess the resources to organize the sensor industry[38] to begin any discussion on standards. Even the FDA had to resort to legislation[39] just to *define* what is a medical device. These issues are detrimental to ***<u>collective</u>*** sense of data, unless ***all*** data sources were transmitting data with common characteristics and handled by a ***single*** analytical engine. That would be a sole source, proprietary product, closed to open innovation and an anathema for distributed systems (for example, agriculture).

Making ***<u>collective</u>*** sense of data is crucial, even without sensor standards for specific analytes (eg manufacturers may not 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**[40] and tools[41] for **feature engineering**[42] are emerging as vital tools to identify 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 determining unknown unknowns. Foresight is key to aggregate, curate and extract value from structured and unstructured data.

---

[30] http://bit.ly/PLASMONICS
[31] http://bit.ly/GAS-SENSORS-01
[32] http://www-analytik.chemie.uni-regensburg.de/wolfbeis/P2-www.pdf
[33] https://doi.org/10.1155/2018/7467418
[34] http://www.ieee802.org/
[35] https://www.gs1.org/standards/epc-rfid
[36] http://www.econ.yale.edu/growth_pdf/cdp984.pdf
[37] https://www.futuremedicine.com/doi/10.2217/pgs-2018-0028
[38] http://www.sens2b-sensors.com/directory
[39] www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM401996.pdf
[40] http://dx.doi.org/10.1016/j.compeleceng.2013.11.024
[41] https://people.eecs.berkeley.edu/~dawnsong/papers/icdm-2016.pdf
[42] https://www.kdnuggets.com/2018/02/deep-feature-synthesis-automated-feature-engineering.html

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[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. 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 platform infrastructure. We plan to embrace unstructured data and build capacity for change control[48] to deal with big data (bad word), 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 ⇨ 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
46 http://blogs.worldbank.org/ic4d/agriculture-20-how-internet-things-can-revolutionize-farming-sector
47 http://www.europarl.europa.eu/RegData/etudes/STUD/2016/581892/EPRS_STU(2016)581892_EN.pdf
48 http://vcaf.berkeley.edu/sites/default/files/change_control_process_aa.pdf
49 https://journals.aom.org/doi/10.5465/amj.2014.4002
50 http://tarjomefa.com/wp-content/uploads/2017/08/7446-English-TarjomeFa.pdf

STRUCTURE OF THE INFRASTRUCTURE – ARCHITECTURE, DATA HANDLING, OPEN TOOLS

In the short term, it may be difficult to substantiate convergence of deep[51] systems[52] thinking[53] and system dynamics[54] with lightweight but agile IoT **connectivity** services, IFTTT[55]. To catalyze the diffusion of these ideas, convergence may benefit from a simulation[56] to display its potential value[57] for users and/or socialize parts of the platform concept via extension[58] and education[59] programs for dissemination[60]. In reality, the tool may be a limited sandbox option for [a] water/soil sensor data and [b] wastewater treatment, from an "user push" approach.

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 ***unstructured*** data. Structure/schema rigor must co-exist with *ad hoc* event driven architectures, eg, using containers for portability on multi-cloud platforms *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 focus on 2 issues: [1] feature selection, and [2] software tools for data.

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[61] support (for example, operations or supply chain[62] decision support).

For the PoC, if we focus on two types of data ([a] water/soil sensor data repository and [b] waste water treatment repository), we can address the various characteristics of the data in these repositories (for example, SENSEE) and explore how to select key features.

---

[51] https://www.thinkmind.org/download.php?articleid=bustech_2015_1_30_90048
[52] http://jasss.soc.surrey.ac.uk/17/4/2.html
[53] http://www.sfu.ca/~vdabbagh/Forrester68.pdf
[54] https://en.wikipedia.org/wiki/System_dynamics
[55] https://doi.org/10.1080/01639269.2014.964593
[56] http://web.mit.edu/jsterman/www/SDG/MFS/simplebeer.html
[57] https://ctl.mit.edu/sites/ctl.mit.edu/files/attachments/Josue%20-%20Beer%20Game%20Run%20%20Ex%20Ed%20January%202017%20sub.pdf
[58] https://beergameapp.firebaseapp.com/
[59] http://web.mit.edu/jsterman/www/SDG/beergame.html
[60] http://www.beergame.org/
[61] https://www.theseus.fi/bitstream/handle/10024/132787/Thesis_Bohne_DM%20and%20IoT.pdf?sequence=1
[62] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-smart-sensors.pdf

⇨ 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[63] of inactivation (kd, μ_{max} and K_s)
- 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 logic tools, 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[64]. The ability to create combinations, enhances the potential to ***meaningfully*** merge[65] data.

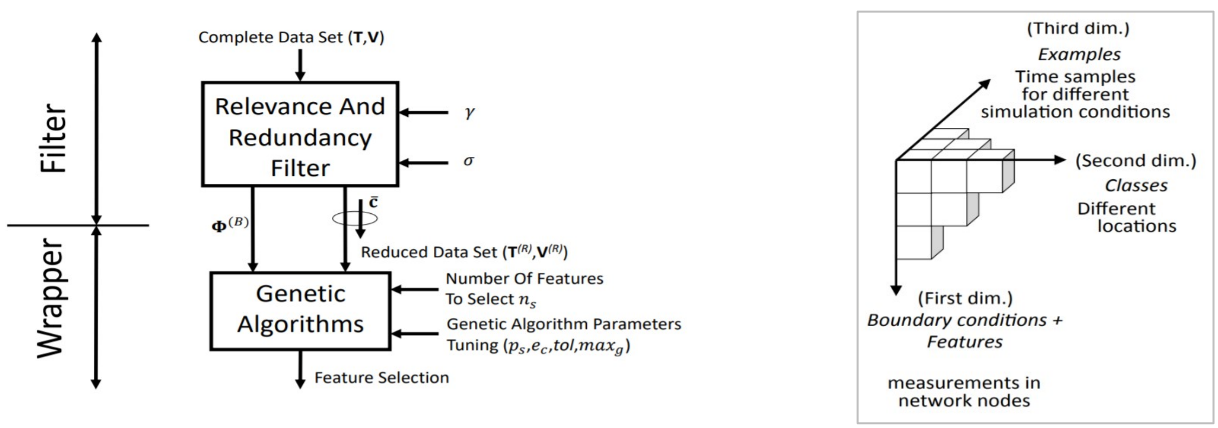


Figure 5: Feature selection (left panel) helps to generate data format (right panel). From ref 64.

[63] http://www.bioline.org.br/pdf?se11003
[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
[65] 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[66] may play a substantial role in data merger between classes of sensor networks[67].

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, the logic tools (ART, SENSEE) 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 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[68] representation of data may prevent merger or data fusion[69], 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. The quality, specificity and value of any answer, depends on the data, design, 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 tools and open platforms. The imagination about the "big picture" must be in focus to anticipate future design needs.

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. Logic 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
67 https://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/
68 http://pages.di.unipi.it/bacciu/wp-content/uploads/sites/12/2016/04/nca2015.pdf
69 https://www.mdpi.com/1424-8220/9/10/7771

The following is an incomplete list of concerns, software suggestions, architecture notes and data issues, to be considered when designing open PEAS Platform for the Agro-Ecosystem:
[a] Time assurance of sensor data (variable). Limited bandwidth or transmission cost are reasons to compress[69] data[70] 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
[e] Time Series DB: Storage of ingested sensor data in TSDB is increasingly[71] popular (InfluxDB[72] or Prometheus, open[73] 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 be a routine operation.
[f] Specialized Databases: Applications must deal with variety of data, including data embedded in text (environmental inspection notes). Open[74] Elasticsearch[75] with support for search primitives (topic modeling, synonym resolution, fuzzy search) and associated open services[76] are recommended. RDBMS may not answer "How often after heavy rainfall are mercury concentrations elevated?" or geospatial[77] queries[78] "Are there mercury levels above 100ppb within a 10km radius of "this" plot of land?" NoSQL[79] databases (MongoDB[80] or Elasticsearch) offer flexible and expressive query languages, do not require defined schemas.
[g] 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[81] offers flexibility[82] for unknown query patterns. PostgreSQL[83] is a credible[84] and robust option.

---

69 https://www.itworldcanada.com/article/canadian-telematics-firm-geotab-doubles-down-on-data-services-as-itfinds-u-s-growth/408647
70 https://www.nrel.gov/docs/fy18osti/70223.pdf
71 https://db-engines.com/en/ranking_categories
72 https://www.influxdata.com/time-series-database/
73 http://opentsdb.net/
74 https://qbox.io/blog/what-is-elasticsearch
75 https://www.elastic.co/
76 https://aws.amazon.com/elasticsearch-service/
77 http://toblerity.org/shapely/manual.html
78 https://www.omnisci.com/
79 http://nosql-database.org/
80 https://www.mongodb.com/nosql-explained
81 https://blog.timescale.com/why-sql-beating-nosql-what-this-means-for-future-of-data-time-series-database348b777b847a
82 https://pdfs.semanticscholar.org/3635/12f3dc6659d05eba4b6f07339f8542bd2737.pdf
83 https://www.postgresql.org/
84 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[85] with additional metadata describing its content[86] (eg, Einstein AutoML system[87]) [iv] aligning geospatial data to one or more known grids (eg, Uber's H3[88]).

At the architecture level, separating the data persistence (store) and retrieval[89] (query) functions of databases are recommended. Common data stores include open source distributed file systems - HDFS[90], GlusterFS[91] and Amazon S3[92] (but life of HDFS is in question). These can be paired with open analytics/query layers for which the underlying storage is swappable (Apache Spark[93], Apache Drill[94], Impala[95] and Presto[96]). 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).

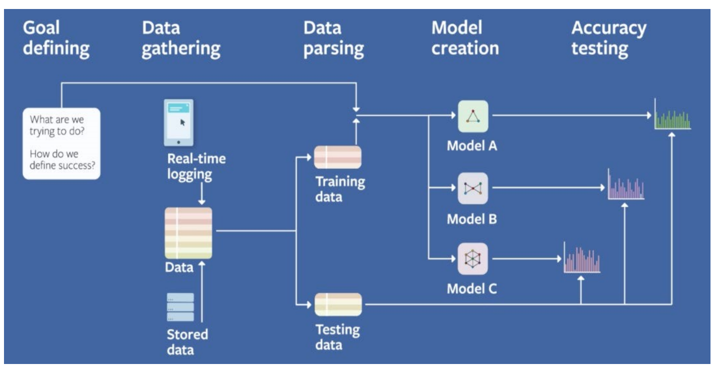


Figure 6: Asking correct questions may better define our goals[97] and extract value from data.

[85] https://github.com/salesforce/TransmogrifAI
[86] https://engineering.salesforce.com/open-sourcing-transmogrifai-4e5d0e098da2
[87] https://www.salesforce.com/video/1770953/
[88] https://github.com/uber/h3
[89] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4528360/pdf/13321_2015_Article_81.pdf
[90] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
[91] https://docs.gluster.org/en/latest/
[92] https://aws.amazon.com/s3/
[93] https://spark.apache.org/
[94] https://drill.apache.org/
[95] https://impala.apache.org/
[96] https://prestodb.io/
[97] https://research.fb.com/category/machine-learning/

TRANSFORMATION OF SARS TO *SARA* - PARADIGM *SHIFTS*

Basic research[98] in advanced[99] material[100] science may be the catalyst for forthcoming paradigm shift from SARS to SARA by using materials that not only sense but actuate[101] 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** ⇨ sense, analyze, response, actuate). Using radio frequency signals to modulate molecules, such as DNA[102], may be used in emerging sensing tools which uses aptamers. The structure-function[103] relationship of DNA can be tuned by remote RF signals. With the emergence of the IoT[104] concept formalized by Sanjay Sarma[105] (marketing[106] term IoT was coined by Kevin Ashton, the concept of IoT has a rich[107] history[108]), the radio frequency identification tool, the RFID tag, evolved as a target for disenfranchised individuals. Functional modification of RFID tags frequently used parasitic backscatter[109] attack, a tool used by cryptanalysts (power analysis attacks). The EPC standard[110] 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, mobile communications, nanobiotech, molecules, biochemistry of infectious disease, and device systems, may enable us to actuate[111] functions and modify sensors using smartphones.

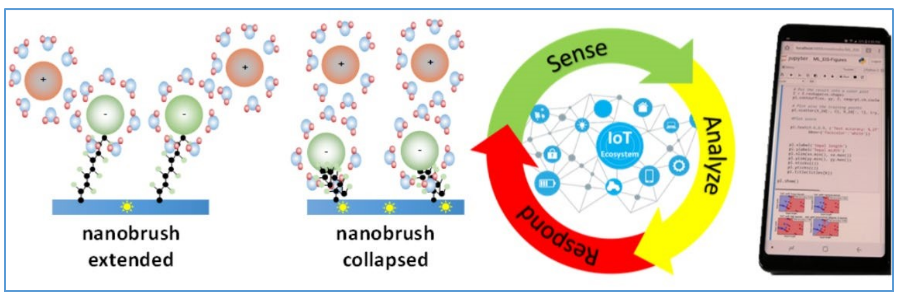


Figure 7: Smart materials can be monitored and actuated with a smart phone (ref 112).

---

[98] https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.044302
[99] http://science.sciencemag.org/content/354/6317/1257
[100] http://advances.sciencemag.org/content/4/8/eaat4634
[101] http://science.sciencemag.org/content/347/6228/1261689
[102] https://www.nature.com/articles/415152a
[103] https://patentimages.storage.googleapis.com/18/41/67/64c9c21a955380/US20020061588A1.pdf
[104] http://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-001.pdf
[105] https://openlearning.mit.edu/about/our-team/sanjay-sarma
[106] https://www.postscapes.com/internet-of-things-history/
[107] http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=34057
[108] https://www.fantasticfiction.com/a/isaac-asimov/sally.htm
[109] https://iss.oy.ne.ro/RemotePasswordExtractionFromRFIDTags.pdf
[110] https://www.gs1.org/sites/default/files/docs/epc/uhfc1g2_1_0_9-standard-20050126.pdf
[111] 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).

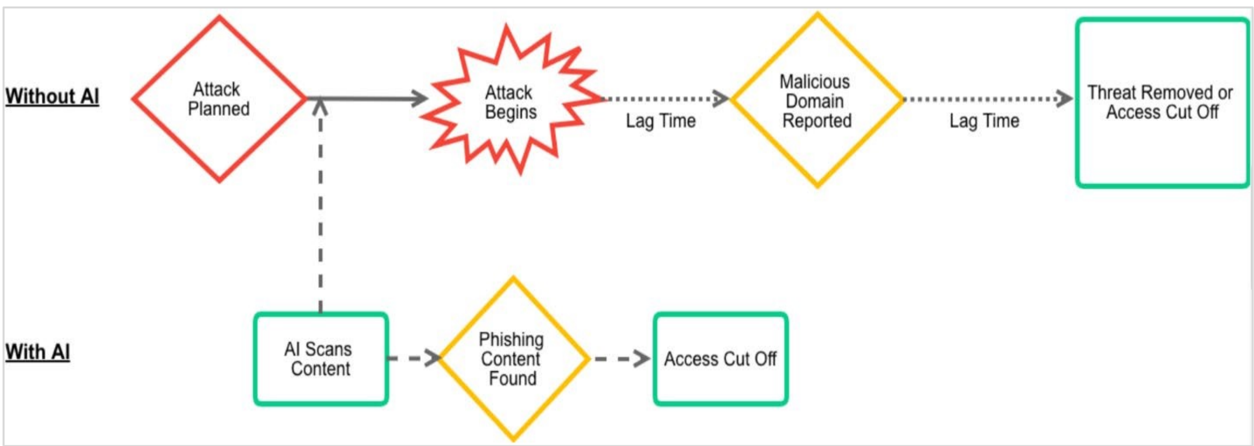


Figure 8: Paradigm of SARA illustrated in cybersecurity (phishing). Agents[112] may sense content, find evidence of potential phishing fraud upon analysis, responds by actuating termination.

To substantiate ideas latent in SARS/SARA we need a compass to continuously map need-based innovation. Future-ready architecture must overlay all computation and complexity with simple forms of knowledge representation, such as "drag & drop" icons[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.

[112] https://www.technologyreview.com/s/608036/we-need-to-talk-about-the-power-of-ai-to-manipulate-humans/
[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?

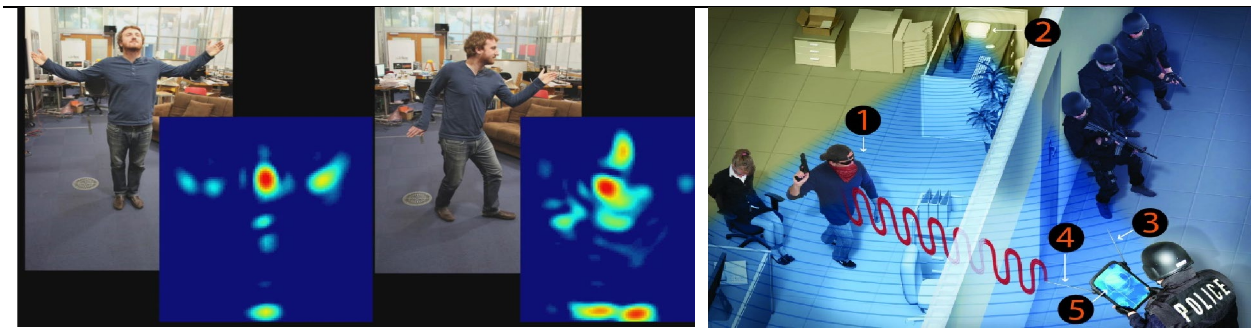


Figure 9: Dina Katabi[113] and Fadel Adib[114] and others have demonstrated how reflection of RF waves[115] may be used to detect objects behind opaque barriers. Using the ability of frequency modulated carrier waves (FMCW, a radar technique) to separate reflections from different objects, it is possible to detect[116] with accuracy, respiration and heartbeat from individuals. Applications[117] 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[118] 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[119] algorithms, and use machine learning[122] techniques to detect "needles" from many haystacks!

[113] http://people.csail.mit.edu/dina/
[114] http://www.mit.edu/~fadel/
[115] https://people.csail.mit.edu/fadel/papers/wivi-paper.pdf
[116] http://www.mit.edu/~fadel/papers/vitalradio-paper.pdf
[117] http://openaccess.thecvf.com/content_cvpr_2018/CameraReady/2406.pdf
[118] https://pure.coventry.ac.uk/ws/portalfiles/portal/7646003/tancomb.pdf
[119] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.888&rep=rep1&type=pdf
[122] http://bit.ly/BOOKS-ML

It is interesting to note that we can identify between Coke and Pepsi using RF[120] signals. Using UWB (ultrawideband) radios, we differentiate between liquids using permittivity[121] of liquids. RF waves travel slower in soil with higher permittivity. By measuring[122] 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[123] as well as simple construction material[124] (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 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[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 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 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" in the Wi-Vi paper (ref 116). The kinetics of molecular motion must be accounted in the design of detector, but are these motions (motion of object vs molecules) comparable?

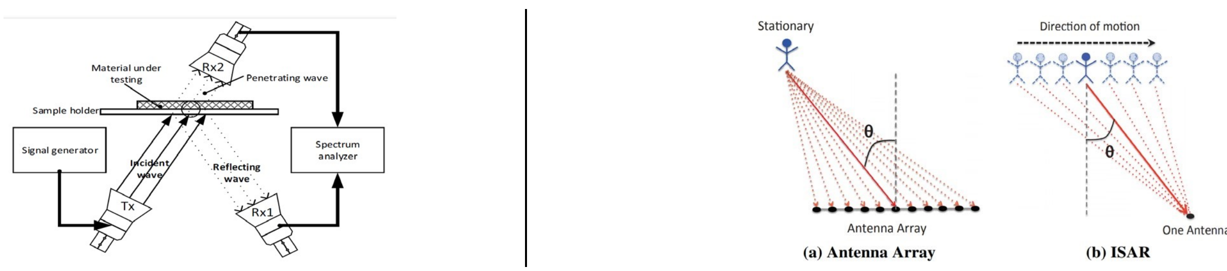


Figure 10: Rational basis for convergence of ideas? Can we combine NDT[127] and ISAR[117] types of tools & techniques to create sensors, based on differential reflection of waves (RF, EM, sound)?

[120] https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4878&context=sis_research
[121] https://synrg.csl.illinois.edu/papers/liquid_mobisys2018.pdf
[122] https://www.microsoft.com/en-us/research/uploads/prod/2018/10/SMURF_TR-1.pdf
[123] https://www.mdpi.com/1424-8220/18/2/609
[124] 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?

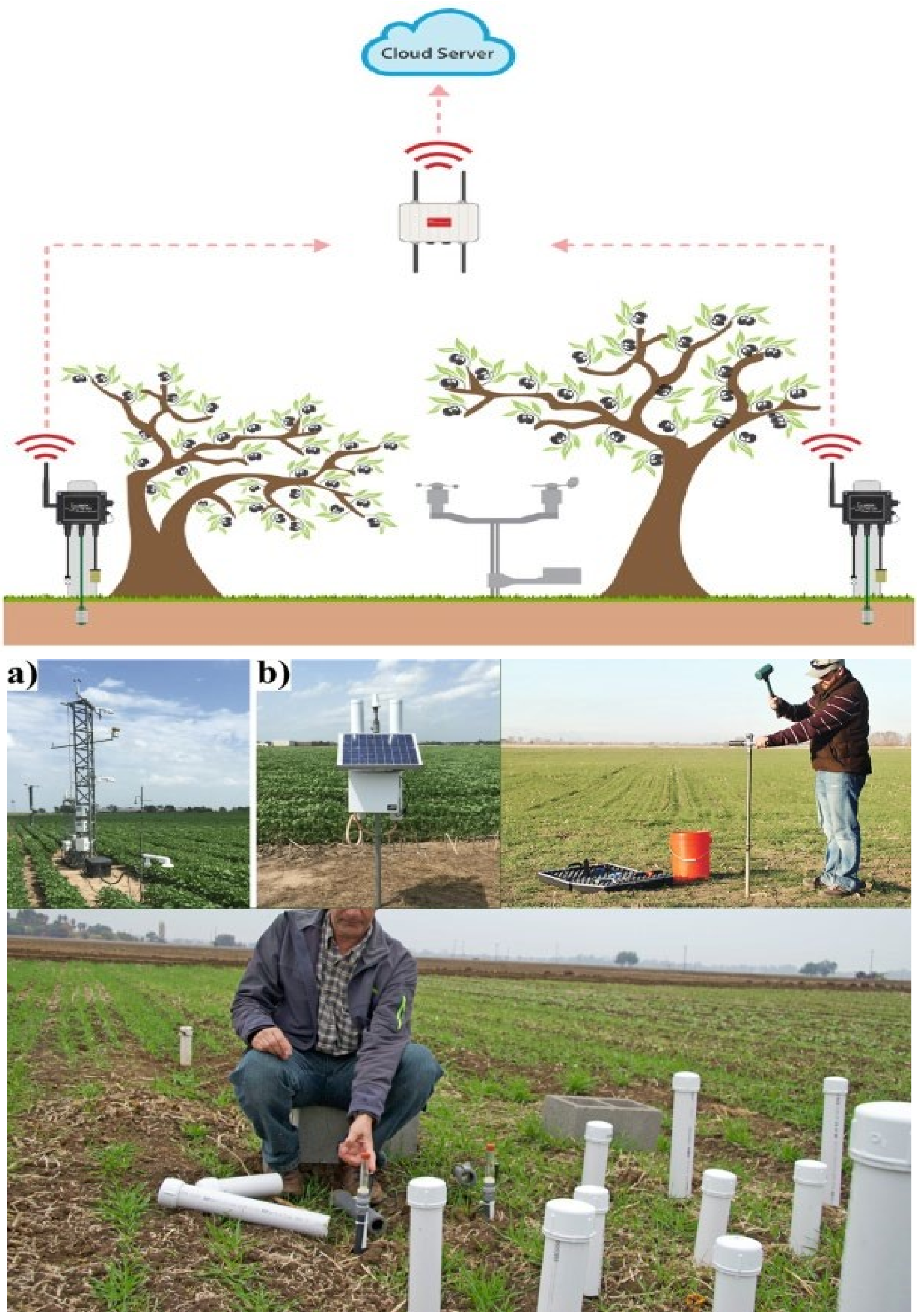


Fig 11: The conventional practice of immobilizing sensors may restrict their dynamic potential. It adds enormous burden of pollution when sensors are non-functional and not biodegradable.

With advances in RF sensing, the use of drones as a gateway[125] 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. The infectious inclination to claim advances by power-point (Fig 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 (Fig 14). It is easy to illustrate what "should/could be" (this essay) and "simulate" what is looks like (Fig 14) but it is a herculean task to transform the vision into reality.

Transformation cannot ***stop*** at decision support. Detecting harmful pathogens or metals in a sample, calls for treatment. ***Suggesting*** treatment without ***implementation***, does not remedy the problem. Decision support is not a solution.

[125] https://www.mdpi.com/1424-8220/18/2/624

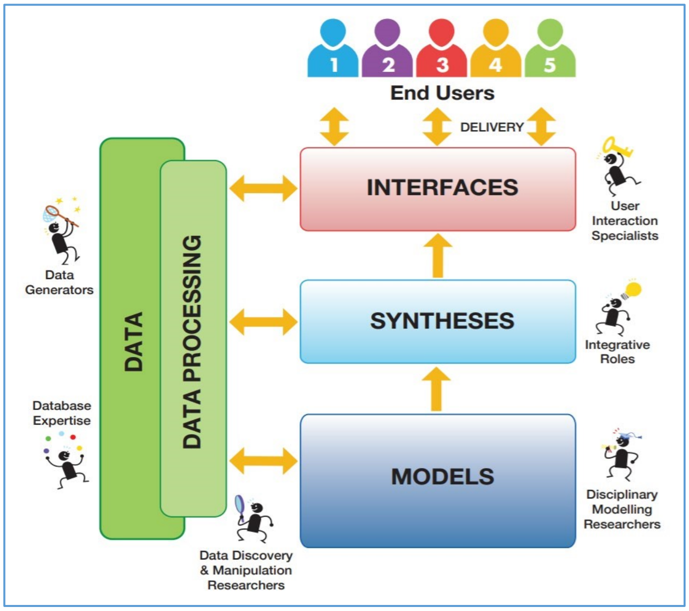


Fig 13: Data-driven intelligent information arbitrage has advanced by power-point[126] but the real task of driving the data, and merging it to make sense, is the task of feature engineering[127] which few can accomplish (and rarely seem to surface in august monographs, reference 129). The illustration (on the left) may take into consideration "discovery" processes[128], include principles of dynamic model-based enterprise[129] 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[130] which can update values (state, parameter), as and when, new data arrives.

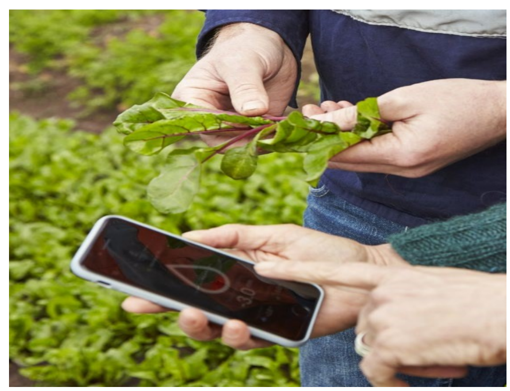


Fig 14: Sensors[131] to detect pathogens and transmit data to smartphones[132] are initial steps to remedial action. All "ideas by power-point" must translate to mobile systems (right) but we still need to "move the needle" by implementing solutions. Knowledge of contaminants must not generate complacency. 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.

[126] www.agmip.org/wp-content/uploads/2015/04/Towards-a-New-Generation-of-Ag-Systems-Complete.pdf
[127] http://bit.ly/BOOKS-FEATURES
[128] https://dl.acm.org/citation.cfm?doid=2836075.2822529
[129] https://www.nist.gov/sites/default/files/documents/2018/11/05/programsummary_mbe.pdf
[130] https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-017-2726-9
[131] https://doi.org/10.1016/j.ebiom.2018.11.031
[132] 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[133] soil moisture sensor uses 2.4GHz WiFi band (Fig 16). NASA's SMAP[134] 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[135] 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?

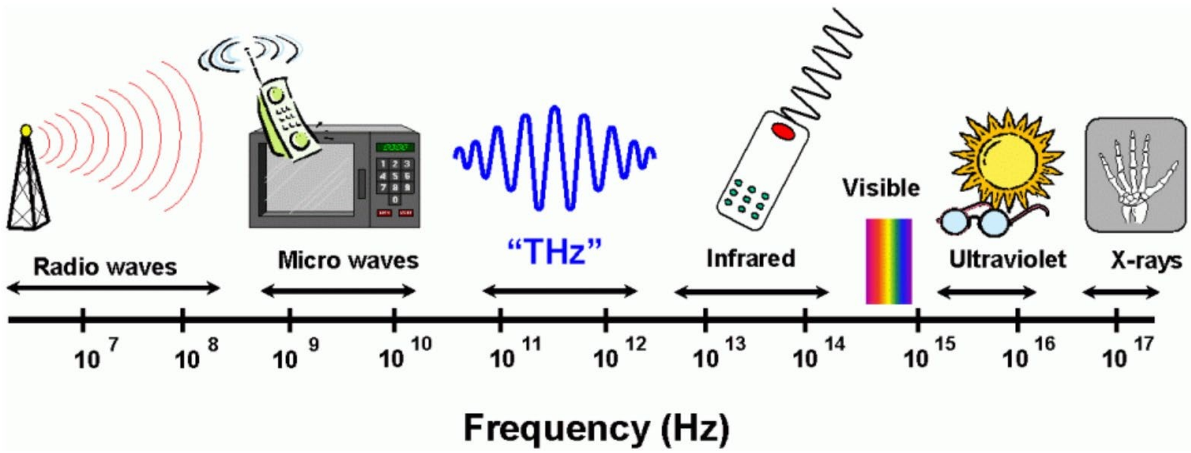


Fig 15: Sign posts of our spectrum: MHz (megahertz, 10^6 Hz), GHz (microwaves are in gigahertz, 10^9 Hz) and THz (sub-millimeter waves are in terahertz, 10^{12} Hz).

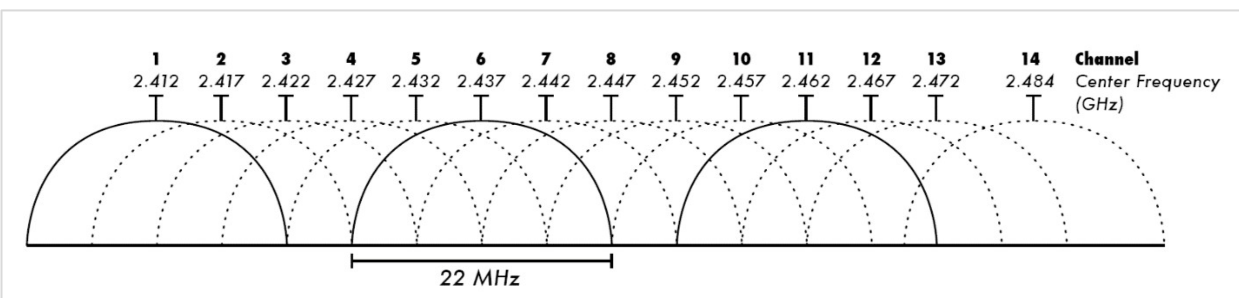


Fig 16: IEEE 802.11 standard protocol provides several distinct radio frequency ranges for use in WiFi communications: 900 MHz, 2.4 GHz, 3.6 GHz, 4.9 GHz, 5 GHz, 5.9 GHz, 60 GHz bands. Each range is divided into a multitude of channels[136] (illustrated here for the 2.4GHz band).

---

133 https://www.microsoft.com/en-us/research/uploads/prod/2018/10/SMURF_TR-1.pdf
134 https://smap.jpl.nasa.gov/faq/
135 https://doi.org/10.1038/s41928-018-0186-x
136 https://upload.wikimedia.org/wikipedia/commons/8/8c/2.4_GHz_WiFi_channels_%28802.11b%2Cg_WLAN%29.svg

Terahertz (THz) sensing and advances in terahertz integrated hybrid electronic-photonic systems[137] 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[138] 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.

[137] https://www.nature.com/articles/s41928-018-0173-2
[138] https://dspace.mit.edu/handle/1721.1/9134

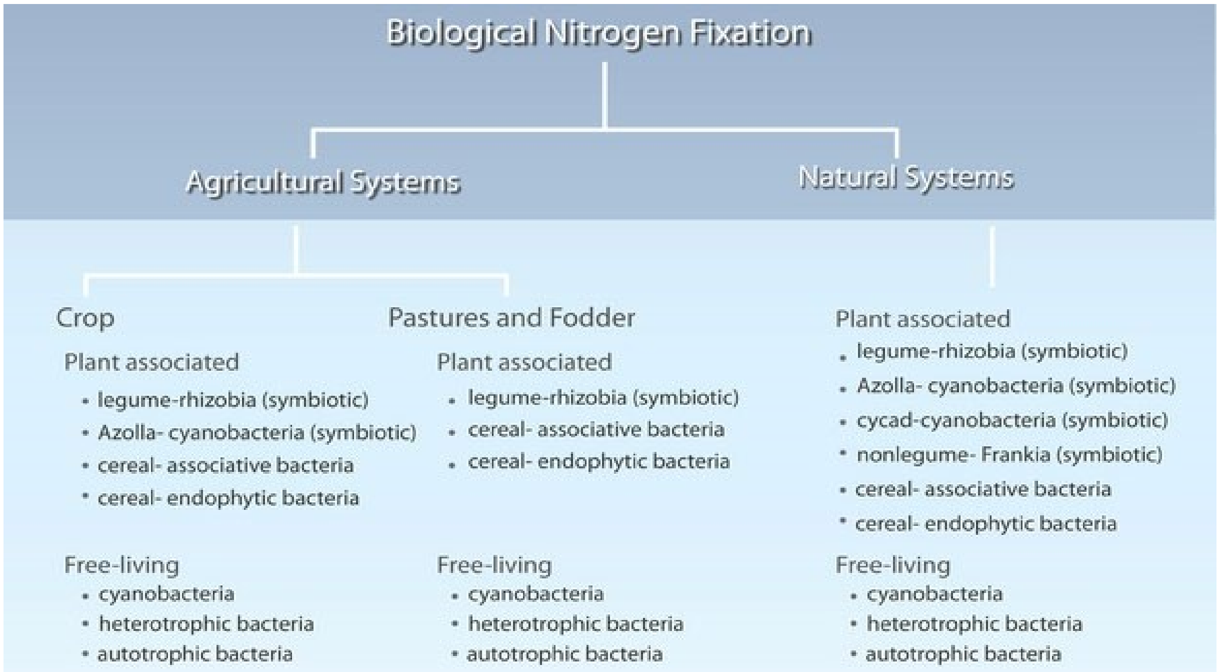


Fig 17: Nitrogen fixation (Beijerinck, 1901), is carried out by a special group of prokaryotes.[139] The central metalloenzyme[140] for nitrogen fixation is a dimeric protein, nitrogenase, with a moiety of Molybdenum (Mo) and Iron (Fe), one for each of its monomers, but with minor exceptions[141] where Vanadium (V) is the metal of choice.[142]

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[143]) 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[144] to buffer oxygen concentration and by a class of bacteroid cytochrome c oxidases[145] with a high affinity for oxygen (in a classic[146] feedback loop, low oxygen signals the transcriptional activation of nif[147] and fix[148] genes in nitrogen fixing bacteria, eg, *Rhizobium sp*).

---

139 https://www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419
140 https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/nitrogenase
141 https://www.cell.com/action/showPdf?pii=0968-0004%2889%2990271-5
142 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1135386/pdf/biochemj00219-0188.pdf
143 https://www.ncbi.nlm.nih.gov/pubmed/22633062
144 https://www.ebi.ac.uk/chebi/searchId.do?chebiId=35144
145 https://www.pnas.org/content/113/45/12815
146 http://www.gs.washington.edu/academics/courses/braun/55106/readings/jacob_and_monod.pdf
147 https://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1007629&type=printable
148 https://www.frontiersin.org/articles/10.3389/fmicb.2016.01343/full

The recent focus[149] 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[150] almost half century[151] ago and patented[152] a quarter century ago. Modifying non-invasive hemoglobin measurement[153] tools may be useful but some the of tools available may not be reliable[157] 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[154] which is increasingly popular as electron paramagnetic resonance (EPR) spectroscopy[155] and used[156] 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.

FUTURE SENSE

In one segment of this meandering collection, 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[157] 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.

PEAS Platform for the Agro-Ecosystem (in this collection) outlines the journey from data to knowledge, in stages, using artificial reasoning tools (ART) to help users, followed by data-informed decision as a service (DIDA'S) and knowledge-informed decision as a service (KIDS).

---

[149] https://www.fda.gov/downloads/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/UCM620362.pdf
[150] http://www.jbc.org/content/148/1/173.full.pdf
[151] https://academic.oup.com/ajcp/article-abstract/29/5/403/1768152?redirectedFrom=PDF
[152] https://patentimages.storage.googleapis.com/87/df/6e/daf7dfd6227ac8/WO1993012712A1.pdf
[153] https://doi.org/10.1016/j.ijsu.2015.11.048
[157] https://journals.lww.com/anesthesia-analgesia/fulltext/2016/02000/Continuous_Noninvasive_Hemoglobin_Monitoring___A.37.aspx
[154] http://www.its.caltech.edu/~derose/labs/exp6.html
[155] https://doi.org/10.1016/B978-0-444-63776-5.00003-6
[156] https://www.ias.ac.in/article/fulltext/reso/020/11/1017-1032
[157] Russell S and Norvig P (2009) *Artificial Intelligence: A Modern Approach* (3rd ed) http://aima.cs.berkeley.edu/

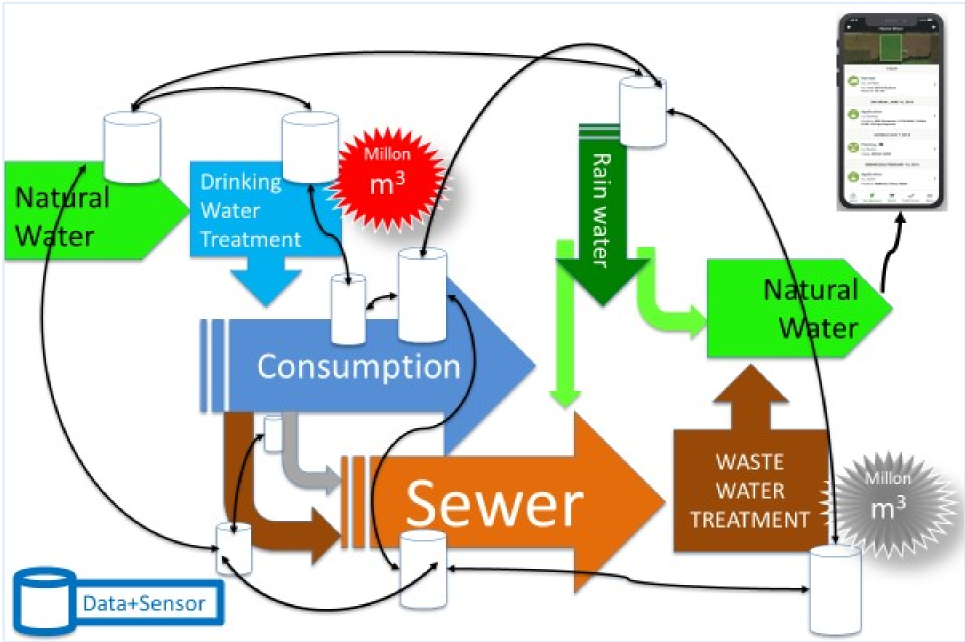


Fig 18: Start with data (SARS) or commence with a goal (PEAS) – The world (of water IoT) needs both. Balance shifts depending on factors determined by the user and/or the user environment.

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[158] 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.

[158] https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf

Conceptually, the SIP-SARS-SARA thinking seeks to merge with the conceptual OODA[159] framework and tangles with the six (0-5) levels of autonomy[160] highlighted in automotive automation, and occasionally used as a management tool[161] 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 it is 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.

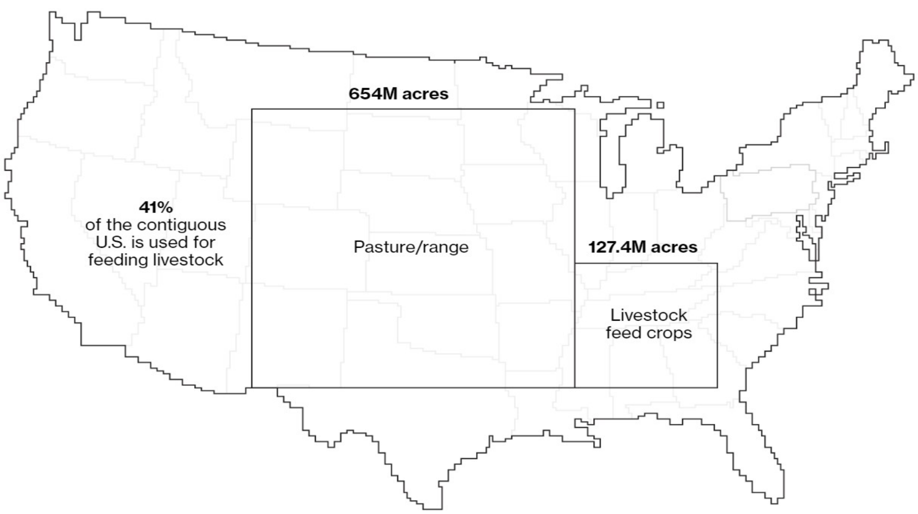
The industry-academia partnership suggested in Figure 14 is not only between like-minded companies in the same business segment but must involve the value chain partners. In the context of machine tools (SIP-SAR) and applications related to food (SARS-AG), we must engage with behemoths and SMEs, including 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.

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 laws of physics governing the spin of oxygen and relevant properties of nitrogen, cannot change in the NICU or in the plant root nodule.

[159] https://apps.dtic.mil/dtic/tr/fulltext/u2/a590672.pdf
[160] https://journals.sagepub.com/doi/abs/10.1177/1555343417726476
[161] https://www.maths.tcd.ie/~nora/FT351-3/DSS.pdf

The desire to usher mobile tools for continuous remote sensing is the basis for the purely speculative suggestion about combining ultrasound sensing on a chip[162] with terahertz sensing[163] with hybrid photonic-electronic systems which can take advantage of RF reflection[164] to create new applications. The use of WiFi[165] and NASA SMAP[166] 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).



There's a single, major occupant on all this land: cows. Between pastures and cropland used to produce feed, 41 percent of U.S. land in the contiguous states revolves around livestock.

Fig 19: US - the land of cows[171] - may need remote sensing tools for external and *in vivo* 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 (eg 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 (eg animal farming, agriculture, water, energy, sanitation and public health).

---

[162] https://www.nature.com/articles/s41467-018-08038-4.pdf
[163] https://www.nature.com/articles/s41928-018-0173-2
[164] http://people.csail.mit.edu/fadel/wivi/
[165] https://www.microsoft.com/en-us/research/uploads/prod/2018/10/SMURF_TR-1.pdf
[166] https://smap.jpl.nasa.gov/files/smap2/SMAP_Handbook_FINAL_1_JULY_2014_Web.pdf
[171] https://www.bloomberg.com/graphics/2018-us-land-use

SARS♦AG - ANIMAL♦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 *we are what we eat*[167] and the vast plethora of efforts[168] to bring home[169] this message. In this section, we attempt to capture a few ideas and suggestions embedded in SARS♦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[170] mixture, the animal is eating (think about ruminants, cows). The biochemistry[171] 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[172] and needs little introduction or justification of its importance. We may address this case, as follows:

[1] Identify Indicators	[3] Sensing for Measurement
[2] Engineering to Sense	[4] Measurement to Applications

In the biochemical lifecycle of a bolus, can we list [1] the key molecules/pathways which may serve, *in sequence*, as molecular indicators (sugars, fats, acids) of progress (*time series*), from ingestion of the feed to digested products, *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 *sequence* of steps.

In part [3] we focus on closed loop[173] sensing, sensor engineering and sensor form and factor. The potential for use of soft robots[174] and closed-loop sensing is key to wireless signal acquisition from *in vivo* measurements. Calibration and *in vivo re-calibration* of sensors are

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

central to data reliability. Calibration and mechanism of sensing (binding, signal, reusability, saturation, sensitivity, mobility, degradation, excretion) are key areas for innovative R&D.

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[175] to reduce methane emissions.

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.

---

[175] https://www.technologyreview.com/s/612458/exclusive-chinese-scientists-are-creating-crispr-babies/
[181] http://www.acad.bg/ebook/ml/Society%20of%20Mind.pdf

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[181] 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).

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).

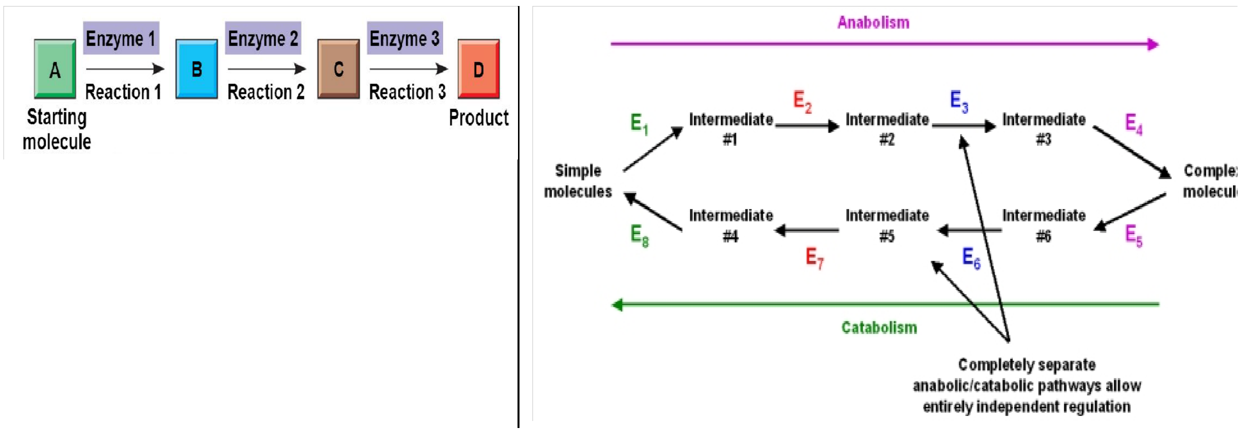


Fig 20: The feed (bolus) is made up of several complex materials which must be digested in a series of biochemical steps to extract nutritional value for the animal. Time series data from measurement of intermediates, or activity of enzymes, are indicative of the progress of these reactions. In the segment [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 20) 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.[176]

The presentation "tool" may resemble a mobile application which may draw on some of the ideas in Lego MindStorm [177] (where players may cooperate[178] 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[179] which are dynamic[180] compared[181] to the Jay Forrester[182] 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.

[176] https://www.nature.com/subjects/nutrigenomics
[177] https://www.lego.com/en-us/mindstorms
[178] https://dam-prod.media.mit.edu/x/files/thesis/2008/sylvan-phd.pdf
[179] https://preventioncentre.org.au/wp-content/uploads/2018/08/080818_Diabetes_FactSheet.pdf
[180] https://peerj.com/articles/5012.pdf
[181] https://pdfs.semanticscholar.org/286e/6e25a244d715cf697cc0a1c0c8f81ec88fbc.pdf
[182] https://executive.mit.edu/faculty/profile/11-jay-forrester

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[183] 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.

CONCLUDING COMMENTS

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.

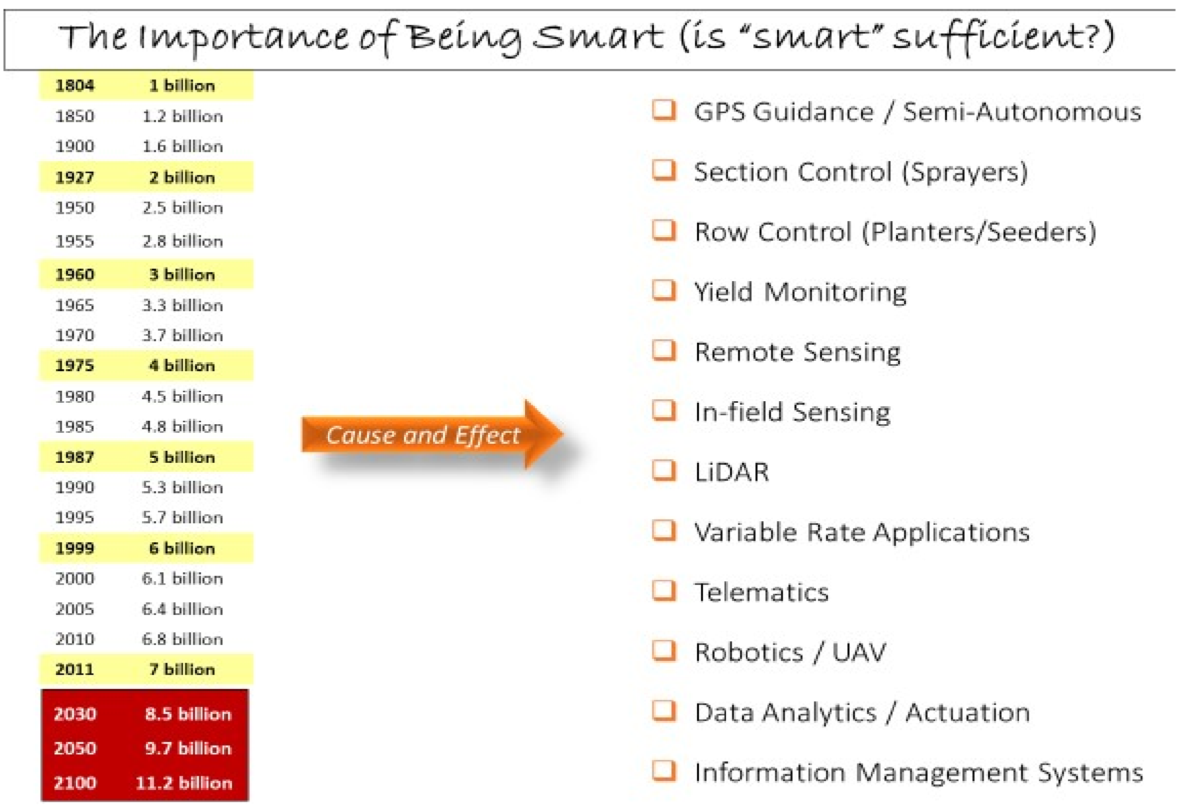


Fig 21: Necessity is the mother of invention. The demand for food and economic development makes it imperative that we search deep within ourselves, and science, to better serve society.

[183] 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[3] 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://ncase.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
[4] https://arxiv.org/ftp/arxiv/papers/1108/1108.3235.pdf
[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://cfpub.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://thesystemsthinker.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/scresponse/repository/Rice_SCResp_Article_SCMR.pdf
[18] https://rmas.fad.harvard.edu/pages/change-control
[19] https://dspace.mit.edu/handle/1721.1/41914
[20] https://link.springer.com/content/pdf/10.1007%2Fs13753-017-0154-5.pdf
[21] https://doi.org/10.1007/s13753-018-0190-9
[22] http://science.sciencemag.org/content/359/6373/325
[23] https://www.researchgate.net/publication/267635903
[24] https://issues.org/esty-2/
[25] https://www.brettonwoodsproject.org/2005/08/art-320747/
[26] https://www.ebrd.com/downloads/research/guides/finance.pdf
[27] https://nautilus.org/gps/applied-gps/global-public-goods/what-are-global-public-goods/
[28] http://dx.doi.org/10.1109/JIOT.2017.2755620
[29] http://www.bookmetrix.com/detail/book/7153d069-09a3-4211-a6cf-34588ef367e9#reviews
[30] http://supplychain.mit.edu/supply-chain-games/beer-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
[34] https://preventioncentre.org.au/wp-content/uploads/2018/08/080818_Diabetes_FactSheet.pdf
[35] http://links.jstor.org/sici?sici=0012-9682%28196908%2937%3A3%3C424%3AICRBEM%3E2.0.CO%3B2-L
[36] https://www.amazon.com/Freakonomics-Economist-Explores-Hidden-Everything/dp/0060731338

“SIGNALS” contains a series of essays spewing amorphous thoughts:

1. SITS
2. SIP-SAR
3. SARS♠AG
4. ART -------- **You are here**
5. PEAS

Please review “SIGNALS”

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)

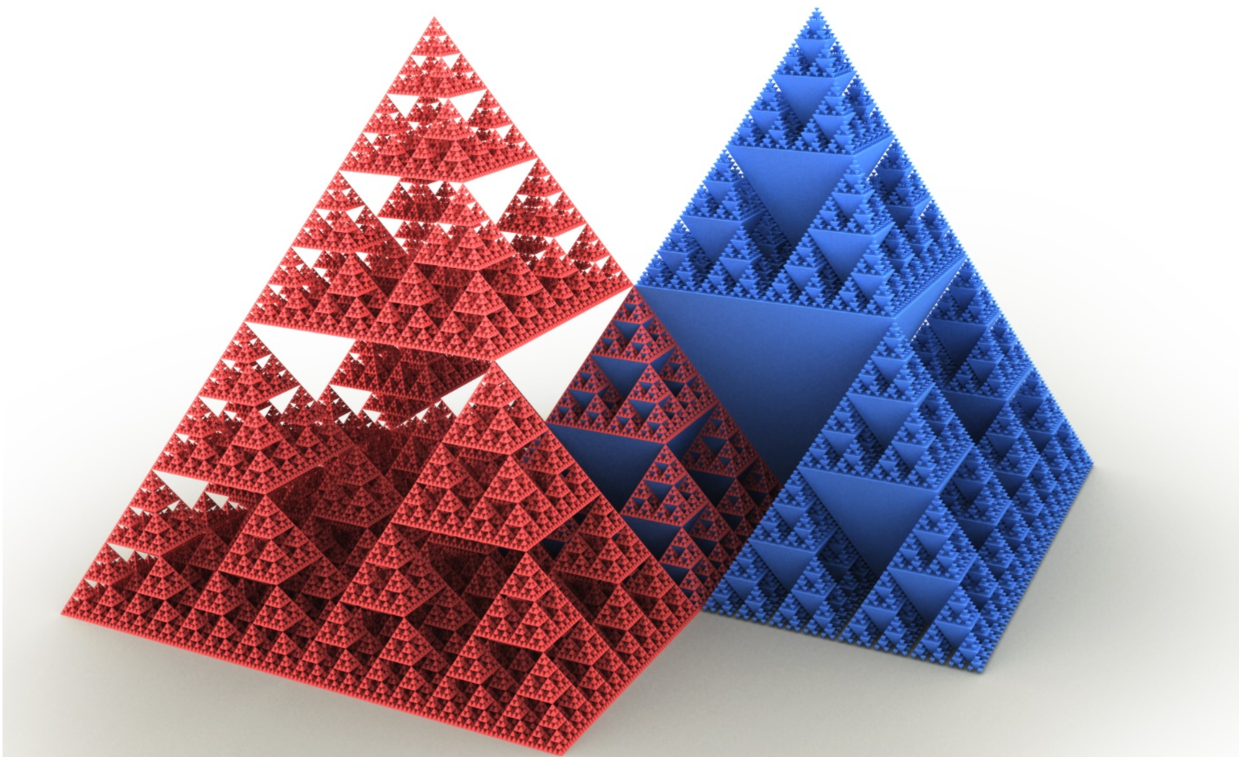
“IoT is a Metaphor” (see Commentary E)

Please review “IoT is a Metaphor”

● Download PDF from MIT Library

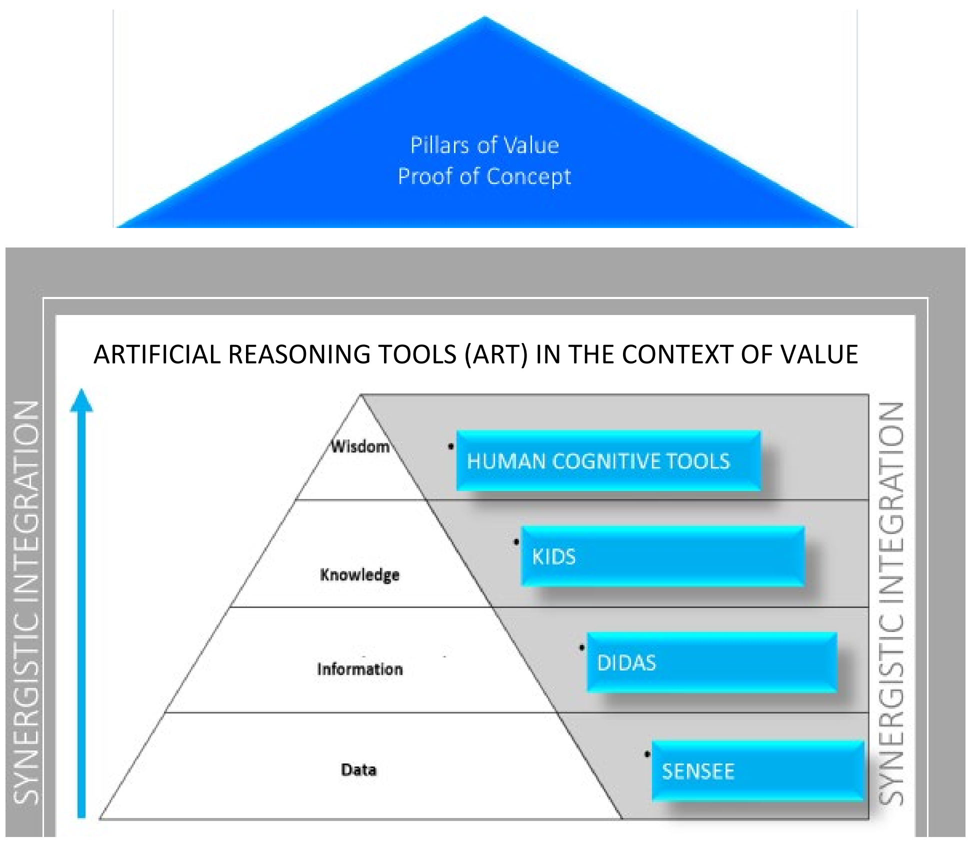
https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.



The **ART** of transforming these ideas into reality?

Shoumen Datta, MIT



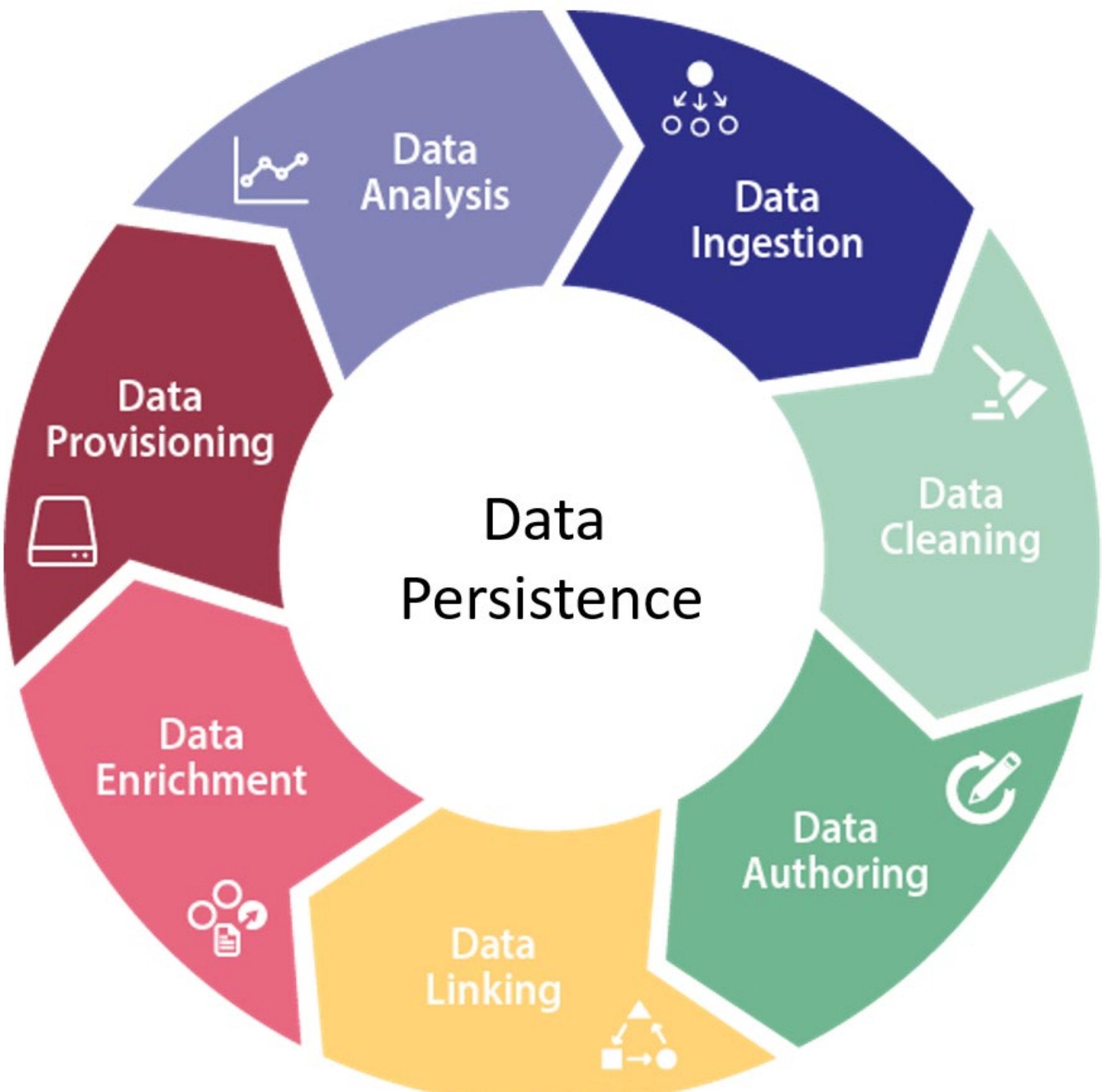
SIGNALS

For the context of this discussion, please download "SIGNALS"

- PDF from https://dspace.mit.edu/handle/1721.1/111021
- Alternate http://bit.ly/SIGNALS-SIGNALS

1 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

The hypothetical ideas in this document outlines data related processes, **not** under implementation, yet. It suggests how we may start phase one (1.0): to create a proof of concept for SENSEE (**SEN**sor **SE**arch **E**ngine). *Draft will be frequently updated and may be downloaded from AWS using this short URL* http://bit.ly/SENSEE-PoC



2 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

This project is a mini proof of concept (PoC) for "SENSEE 1.0" (excludes DIDA'S KIDS).

See "*PEAS Platform for the Agro-Ecosystem*" and essays http://bit.ly/SIGNALS-SIGNALS

What do we want for SENSEE 1.0 demonstration purposes?

A short URL, which enables an app equivalent (web-service) on a smartphone. This app, when it opens on the phone, will reveal a dialog box. Users will type questions in the dialog box.

What is the purpose of the app for SENSEE 1.0 demonstration purposes?

Reply to queries related to the questions the user types in the dialog box. The questions will be of the type suggested (below and elsewhere). The answers will be sourced from the xl spread sheet, indicated below as "source" (please download xl from URL provided below). Decisions about sensor types is the expected outcome (at this phase, SENSEE 1.0 may only help sensor experts).

Type of questions that users may wish to ask using the dialog box (on a mobile device):

Deliverable for the mini proof of concept (PoC): answer natural language questions, for example:

[a] what is the LOD score for ionic mercury / mercury ?

[b] can I use graphene paper to detect E. coli ?

[c] is the ammonium ionophore liquid or solid ?

[d] what type of recognition tool do you have at hand for detecting imidacloprid ?

[e] what is the response time for superoxide dismutase ?

[f] what is the phase of the nitrate ionophore ? *Where semantics will come in (not now, in future).*

A detailed data dictionary (semantics) is not required. Query language will be restricted, for PoC.

Answers are in this spread sheet. The "raw data" pertains to sensor categories, attributes.

SOURCE - download xl from - http://bit.ly/SENSOR-LIBRARY-ERIC-MCLAMORE

FAQ - What is considered the primary key for the dataset?

This is NOT a data set. This is a reference for type of "tools for detection of molecules" that we refer to as sensors. Therefore, in database terms, the entire set of columns uniquely identifies rows in the table (but we can drop a few less populated columns, in the initial response. It is going to be useful when we move to knowledge graphs). At this time we are not presenting any data for the type of sensors in the xl sheet. When we have a "data set" which represents logged data from a specific sensor (SENSEE 2.0), then the ***time stamp*** on that data logger may be one of the primary keys to create an unique identifier. As this point we have a set of columns which are ***all*** key attributes for a sensor (perhaps think column headings as "features" for future phases.)

FAQ - Is there a codebook for column name headings (definitions / standards understood by non-expert users)?

No. This is not the vernacular that farm-users are likely to know or use. For the deliverable proof of concept, ignore the semantics of the column headings (it is a future task when we begin to use graph databases / semantic data catalogs - Fig 12 on page 29 http://bit.ly/SIGNALS-SIGNALS).

FAQ - Is there a specific metadata standard for this community of practice (e.g. common naming for data elements)?

[a] Explore SensorML and StarFL – review – https://pubag.nal.usda.gov/catalog/1229146

[b] Early XML implementations
https://www.isprs.org/proceedings/xxxv/congress/comm4/papers/516.pdf
https://www.iitk.ac.in/nicee/wcee/article/13_956.pdf

[c] STANDARDS
https://geo-ide.noaa.gov/wiki/index.php?title=SensorML_and_ISO_Metadata
https://www.w3.org/TR/vocab-ssn/

[d] REVIEW
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.562.8057&rep=rep1&type=pdf
https://publik.tuwien.ac.at/files/PubDat_208567.pdf
https://www.tandfonline.com/doi/pdf/10.1080/17538940902866195?needAccess=true
https://pdfs.semanticscholar.org/0999/630530af38c85f31ef2a6e2bb3f701da582b.pdf

[e] THE ROAD AHEAD – FUTURE CONSIDERATIONS
https://www.w3.org/2018/03/wot-f2f/slides/Mdata-WoT-2018-03-26MM.pdf
https://www.usgs.gov/land-resources/nli/landsat/science

FAQ - How to document missing data/null values in columns MW [Da], LOD [M], Max range [M], etc. (e.g. "9999" or N/A)?

When SENSEE 1.0 is a working platform, these values may be contributed. The "open platform" and "open port" approach necessary in the SENSEE information architecture, in order that data can be ingested when we have raw data. SENSEE 1.0 is a database ***for sensor device*** reference, at this time. In SENSEE 2.0, we will offer a library of interfaces (APIs) that a 3rd party can download to upload ***sensor-specific data***. SENSEE 2.0 expects to ingest ***case-specific*** sensor data (not there yet). In the context of SENSEE 2.0, feature engineering and feature selection ("selecting a few things that are most important, given that only a few can be sustained") will guide selectivity of data ingestion depending on compatibility between SENSEE 2.0 metadata vs external sources. Publishing "libraries of tools to enable interoperability between databases and data formats" as downloadable tools from the SENSEE portal may help to ensure that we are providing users the ***ability to collaborate*** even if [a] they are in Cairo, Cardiff, Cali or Calcutta and [b] their data schema, style sheets and data holders may not match SENSEE. Open APIs, open platforms and interoperability must be an integral part of the data management plan. Open information architecture may catalyze distributed data collection to strengthen SENSEE, ART and build toward a "Google of ag" approach, as we move beyond SENSEE (1.0, 2.0) and ART to meet the challenges of DIDA'S KIDS (please see "PEAS Platform for the Agro-Ecosystem").

Additional information to address potential question - What are the common data elements for column standardization [e.g. Is the LOD [M] the same as Range (LOD)?]

The value is a concentration expressed in the standard form. Range (xl sheet) should be bounded or display the lowest detectable. Limit of detection (LOD) is defined as the lowest concentration (hence, value in nanomoles, micromoles, millimoles) at which 95% of positive samples are detected. LOD is not necessarily within the linear range of an assay. LOD can be lower than Lower Limit of Quantification (LLOQ), defined as the lowest standard on the calibration curve. For further details, explore: https://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf

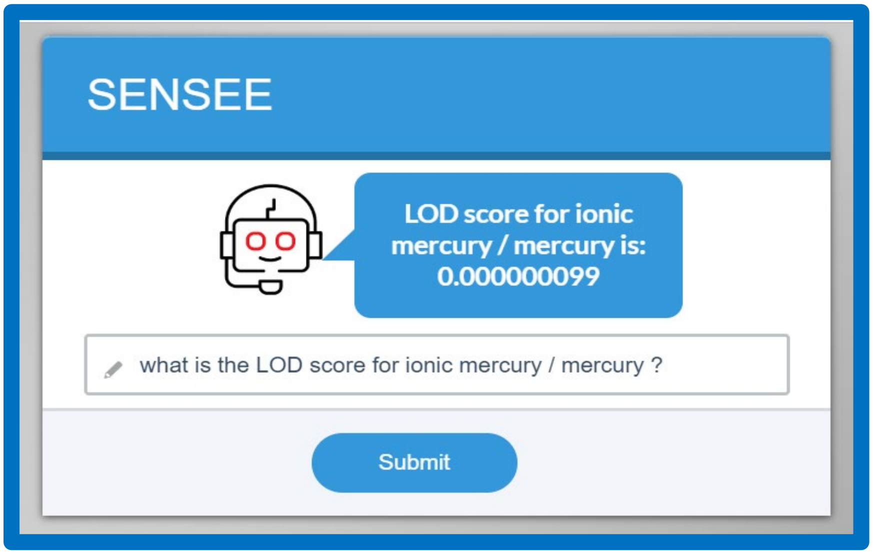
ELEMENTS OF A BIGGER PICTURE – IDEAS BEYOND THE PROOF OF CONCEPT

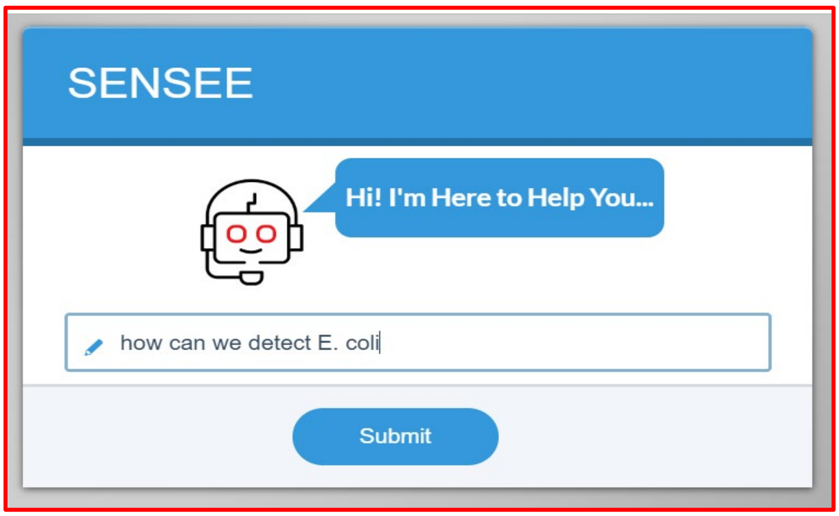
Information arbitrage (PEAS Platform for the Agro-Ecosystem), ART, data-informed decision support for the agro-ecosystem (DIDA'S) and the food industry are expected outcomes. A few steps of this scenario are outlined. Target (?) is to deliver [I] through [V] for FY 2019-2020.

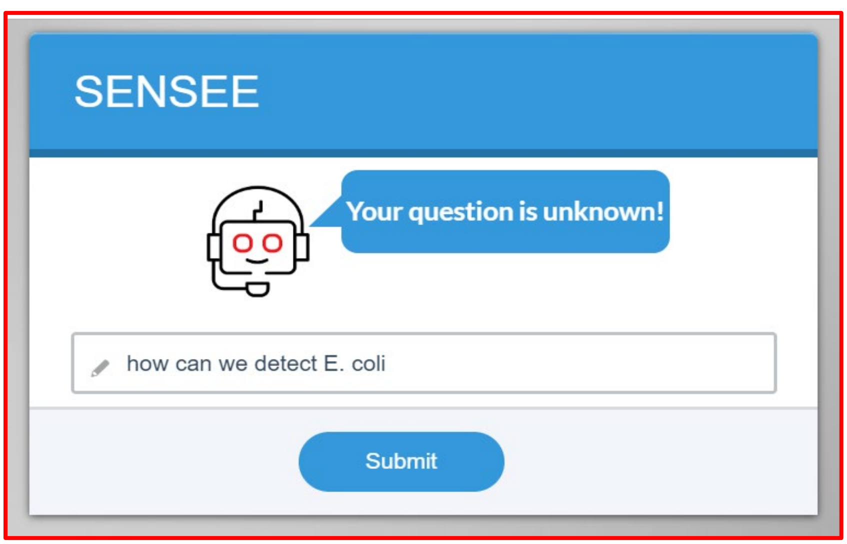
[I] PoC delivers app to return limited number of queries based on sensor description (xl sheet)
Result: http://146.185.133.187/SENSEE1/ ● http://139.162.7.63/SENSEE/
Task: App responds to query about a few sensor descriptions. Host and maintain app and DB. Provide URL to download web service and continue to bolster search functions (SENSEE 1.0).
Comments: Hard coding exact questions (syntax) is inadequate. Elasticsearch and NLP, for example, BERT (Bidirectional Encoder Representations from Transformers) is preferred and/or necessary (https://arxiv.org/pdf/1706.03762.pdf, https://github.com/google-research/bert and https://arxiv.org/abs/1905.05950). NLU/NLP engine may be trained to search keywords in the user's question and may eliminate the need for users to abide by restrictive syntax. At this time, the extent of the library and framework is extremely limited (only one xl sheet provided, others are expected). Therefore, the demand for NLP techniques may be rudimentary and limited to effective text representations and extraction of keywords from natural language queries. Advanced NLP expertise, semantic extraction techniques, data structures and modeling, will be required when volume and variety of sensor descriptions are likely to increase. The UI for Q&A using a web service (http://appinventor.mit.edu) is expected to remain simple. For more on BERT NLP explore: http://mlexplained.com/2017/12/29/attention-is-all-you-need-explained/

[II] Create auto-config xl tool which can be downloaded to upload sensor details (100-1000 labs)
Target: 2019
Task: Expand and scale the system to contain other types of sensors created by other labs.
Comments: How will these other labs accomplish this task? How will the other labs add to the DB? How will they create and populate new columns if the descriptor/characteristic is not present in the current DB? Most labs maintain sensor type descriptions as tables (CSV, xl). Provide a short URL which will lead to a "tool" which can be downloaded and serve as a document management system (DMS) to accept the xl document with sensor descriptions. The uploaded (ingested) document will be parsed and analyzed by system "software" using keywords (think search engine optimization). The existing DB will be updated if the table/column headings are a match. Consider these tasks: [i] Online job application sites where the resume is uploaded to a site and the portal populates its "boxes" (fields) by extracting information from the uploaded resume using metadata tools. The vast number of errors in this process often requires editing (by the user) because the rules of the parser are shoddy.

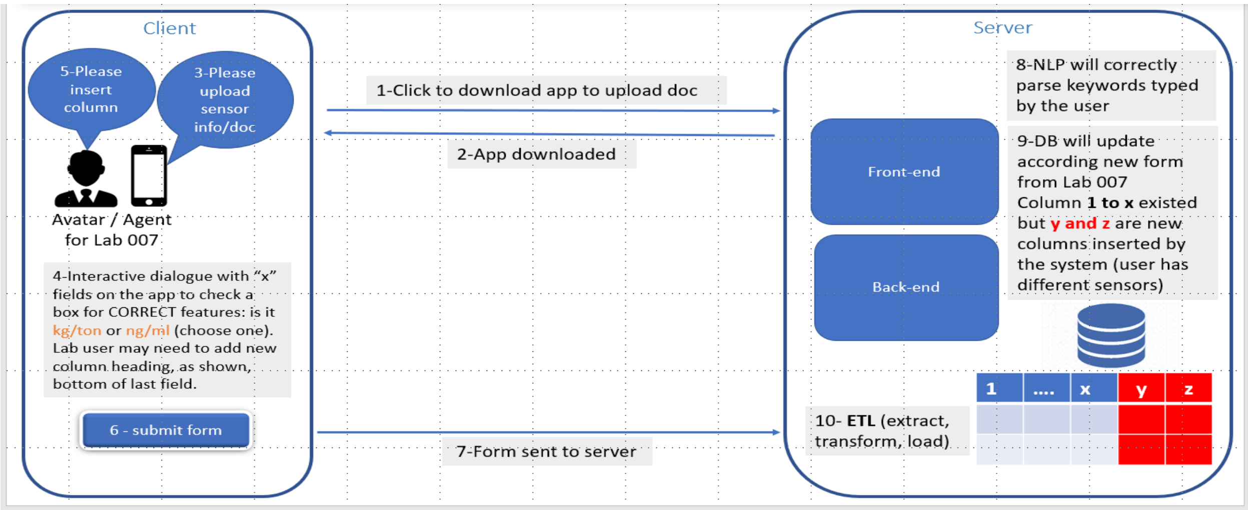
Step I – Dialog Box – http://146.185.133.187/SENSEE1/ ● http://139.162.7.63/SENSEE/

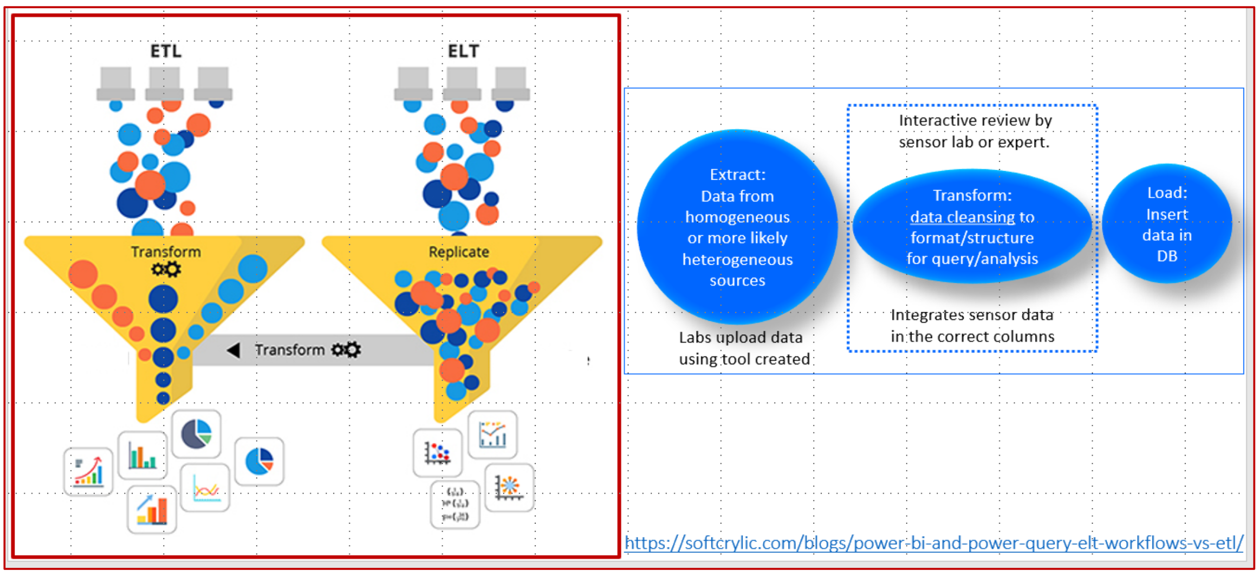


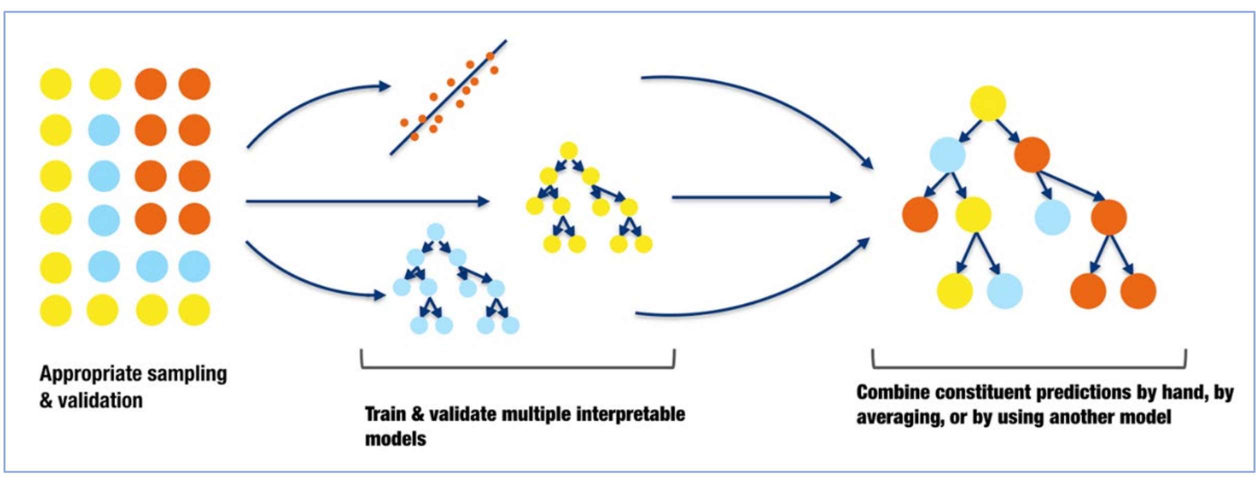




Suggested cartoon for Step II where "tool" may be downloaded via web service/mobile app



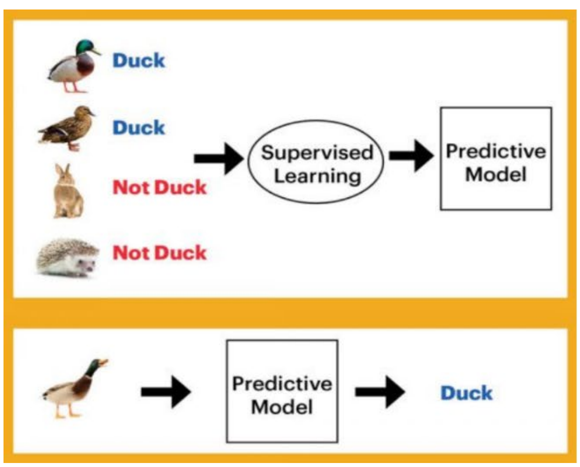


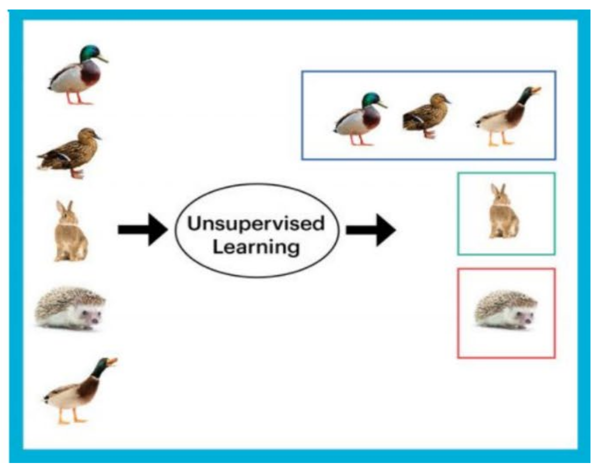


[ii] Online tax preparation systems cannot afford to be shoddy and has developed far better protocols to insert parsed data from uploaded documents (eg W2, W4, 1099 forms). In addition, the software (eg https://turbotax.intuit.com/) may serve as an example of (workflow) interacting with automatically-generated features and interpretation of incoming unstructured (data) sensor description. The value of this software is in the underlying dependencies and UI prompting users to accept or reject the suggestions (see https://www.youtube.com/watch?v=WLaXFYIF4x8).

SENSEE 1.0 tool in step II must be better than resume uploaders and approach the sophistication of the tax preparation packages. The example of tax prep software is relevant for the application in this PoC because it will allow labs to verify uploaded descriptions and characteristic values. It will be helpful (and perhaps necessary) if the DB can be auto "expanded" to create/include new column headings if a new criteria/characteristic is uncovered.

The interactive Q&A type "accept/reject" option in the tax prep software, if implemented in the tool to ingest sensor description data (in SENSEE 2.0 we will ingest and aggregate actual sensor data), will reveal a dataset of "wrong" answers when the user rejects the suggested field value. It is crucial to capture this "what was predicted" vs "what was accepted" "what was rejected" because the correct, and the incorrect input, ***both***, are useful to "classify" what is correct and what is incorrect. Hence, the strength of this approach is in creating datasets for supervised learning (classification algorithms, for example, in the illustration below – duck / not duck).





When using this tool to aggregate SENSEE 1.0 sensor description data (1,000 labs?), it will be important that the application captures every recommendation shown to a user, and the outcome. The major challenge in this approach is the demand for domain specific knowledge (in this case, sensor engineering) and the ability of the individual(s) with domain knowledge to work with a software specialist who has some understanding about the domain. In order to recommend options to the user, the software must rely on a set of rules, dependencies and logic structures, which are relevant to the context of the use (in this case, sensors). For example, the units for the concentration of mercury in a sample may be in ng/ml or in mg or ppb but not in kg/ton or cubic feet. To reduce computational load, perhaps kg/ton or cubic feet, will be excluded as options, to reduce search space, and compress time to search (these metrics will be useful for evaluation).

8 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

It may be informative to review Claude Shannon's "Programming a Computer to Play Chess" (https://vision.unipv.it/IA1/aa2009-2010/ProgrammingaComputerforPlayingChess.pdf) where he coded into the program, knowledge of "weak positions" to limit the search space. Could we use techniques with unstructured documents with sensor information and limit the search space, for example, using topic modeling? In trying to implement these techniques, one cannot escape the need for convergence of domain knowledge with software and principles of machine learning, in tool design. (www.historyofinformation.com/detail.php?entryid=4364 and https://journals.sagepub.com/doi/10.1177/0306312711424596).

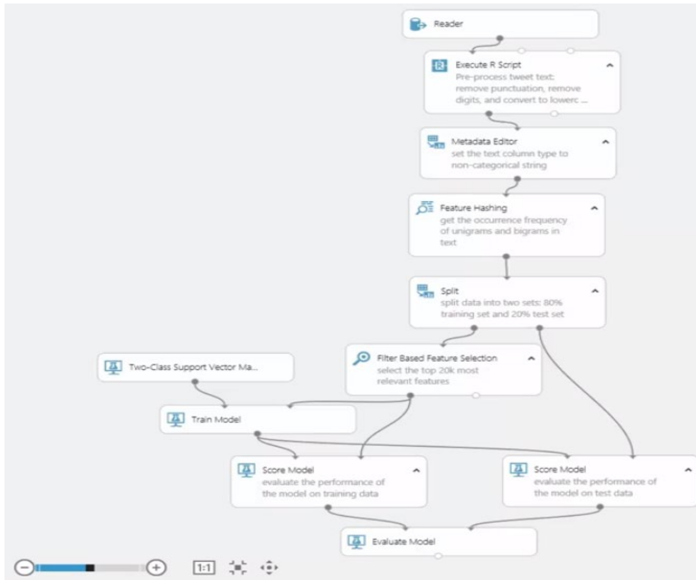
Recommended skill development: To combine sensor data domain expertise with machine learning. Train sensor experts to understand software and machine learning principles and vice versa (machine learning and programmers to grasp the basic tenets of sensor engineering). An exercise for students at UC Berkeley uses resume parsing. The task was to take "pasted-in" text from resumes (just normal ASCII text, no rich PDFs or other formats) to extract "skills" from the content. Using this knowledge, the task was to recommend one skill to a person, which, if acquired, may lead to a recommended job (the skill set the person already possess is one skill away from the recommended job). The goal was to provide guidance for veterans entering civilian life (https://www.shift.org/). Topic modeling was accomplished using genism (https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/ and also https://github.com/RaRe-Technologies/gensim). See https://github.com/jameslamb/skills and view http://bit.ly/JAMES-UC-BERKELEY

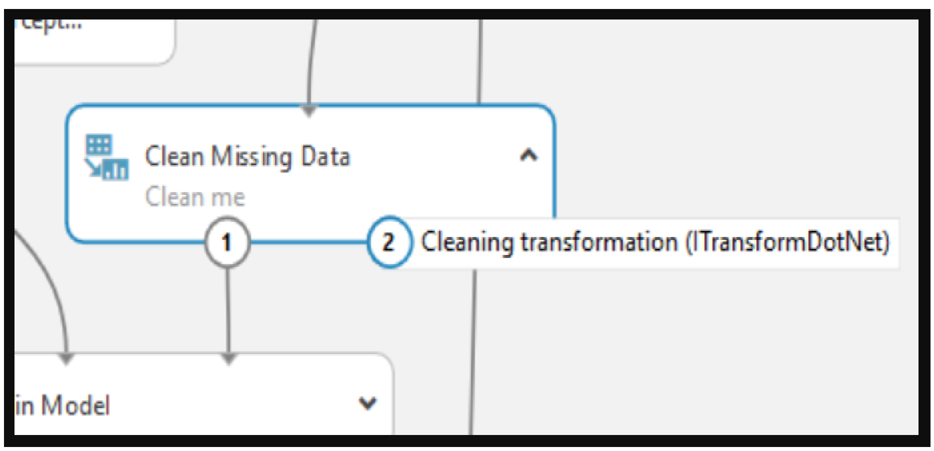
[III] Create 'feature' library / downloadable 'feature tool' to extract sensor details (>1,000 labs)
Target: 2019 ● http://www.jmaxkanter.com/static/papers/DSAA_DSM_2015.pdf
Task: Accomplish the same as [II] but for >1,000 labs.
Comment: Using repository SENSEE 1.0, select features that can accommodate variety of possibilities (expected sensor types from >1,000 labs). Feature selection and feature engineering must serve at least two purposes, the software must be able to operate as a search engine and extract the feature from publications/papers related to sensors (search engine, ML, limited to the library of features). In SENSEE 2.0, feature-guided ingestion of sensor-specific data (pH, salts, temperature, analyte) where the brand of the sensor may be different (different manufacturers) but the pH data from all/any pH sensor, can be pooled into a database using the selected feature (populate the data, that is the pH data, for selected feature pH). The aim of this PoC (at this time) is to create SENSEE 1.0 as a repository for different types of sensors by aggregating ***sensor descriptions*** from labs, worldwide (see examples in xl sheet, link on page 3). When completed, combined SENSEE 1.0 and SENSEE 2.0 is expected to fuel ***artificial reasoning tools*** (**ART**).

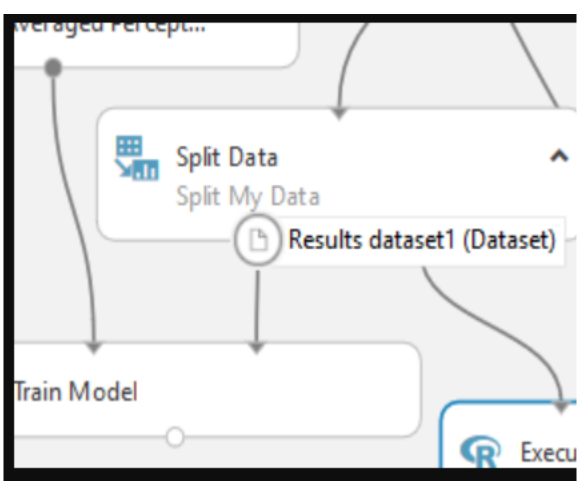
[IV] Automated feature tool linked to **sen**sor **se**arch **e**ngine SENSEE 1.0 (scale to 10,000 labs?)
Target: 2020 ● https://blog.featurelabs.com/deep-feature-synthesis/
Task: "Drag & Drop" tool to gather sensor type (***later,*** sensor data in SENSEE 2.0). Distributed tools: feature automation ● https://people.eecs.berkeley.edu/~dawnsong/papers/icdm-2016.pdf
Comment: Build feature engineering "engines" to process unstructured data and automate tools to arrange it "meaningfully" to serve queries. Idiot-proof "for dummies" interface is provided to users who may rearrange "modules" (data description, knowledge extraction) using ***drag & drop*** tools, perhaps similar to Lego Mindstorm (www.lego.com/en-us/mindstorms/learn-to-program).

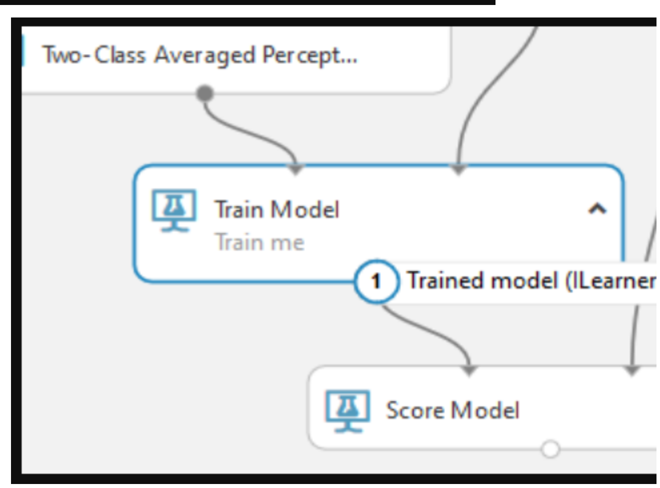
DRAG AND DROP USER INTERFACES - AGNOSTIC ABOUT USER'S KNOWLEDGE OF PROGRAMMING

NEXT GEN LEGO MINDSTORM - EXAMPLE FROM MICROSOFT AZURE MACHINE LEARNING STUDIO









Source: https://github.com/hning86/azuremlps

10 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

In this approach, users are **not** required to understand programming, logic and underlying processes. If the masses are not inhibited from using the tool, it will accelerate the diffusion of the tool. Democratization of access through lego-esque, modular "drag and drop" user friendly interfaces, will catalyze adoption of the tool, not only in the agro-ecosystem (the discussion in http://bit.ly/SIGNALS-SIGNALS) but in any domain, for example, healthcare, manufacturing (think digital twins https://arxiv.org/abs/1610.06467), finance, oil & gas, logistics and transport.

Effective and efficient auto-generation and selection of features by modeling information about features (Dawn Song - https://people.eecs.berkeley.edu/~dawnsong/) is a milestone development, we wish to embrace. Using machine learning to ***describe features of features*** and form better expectations of which features might be worth generating and testing, is an incisive advance. ExploreKit feature automation approach is a positive evolution from the brute-force, opaque, model-based approaches to data transformation, which are still a part of machine learning (ML), for example, back propagation (https://www.nature.com/articles/323533a0) and random forests (https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf).

In FeatureLabs (https://idss.mit.edu/staff/kalyan-veeramachaneni/), feature engineering is performed relative to richer descriptions of input data and successfully applied for commercial purposes by corporations (Einstein AutoML https://www.salesforce.com/video/1776007/ and https://github.com/salesforce/TransmogrifAI/tree/master/features/src/main/scala/com/salesforce/op/features; also see – improving AutoML transparency – https://arxiv.org/pdf/1902.05009.pdf).

One aim of this project is to partner with bonafide experts to create open feature automation tools. The principles, as indicated above, will be applicable to a broad spectrum of applications.

[V] Gift SmartPath SENSEE tool (USDA/NIFA/NSF) - national/global sensor data repository
Target: 2020
Task: Monitor and Model ➔ Detect and Predict ➔Diagnose and Explain ➔ Decide and Act (actions generate outcomes, which are monitored, and the data-informed process is repeated).
Comment: Open tool will democratize access to data and can be adapted for other domains.

Future Scope (FS)

[FS-VI] Development of SERVICE – "PEAS Platform for Agro-Ecosystem" for users. Farmers and growers (food) seeking solutions from logic tools and help from **ART**. (Target FY 2021)

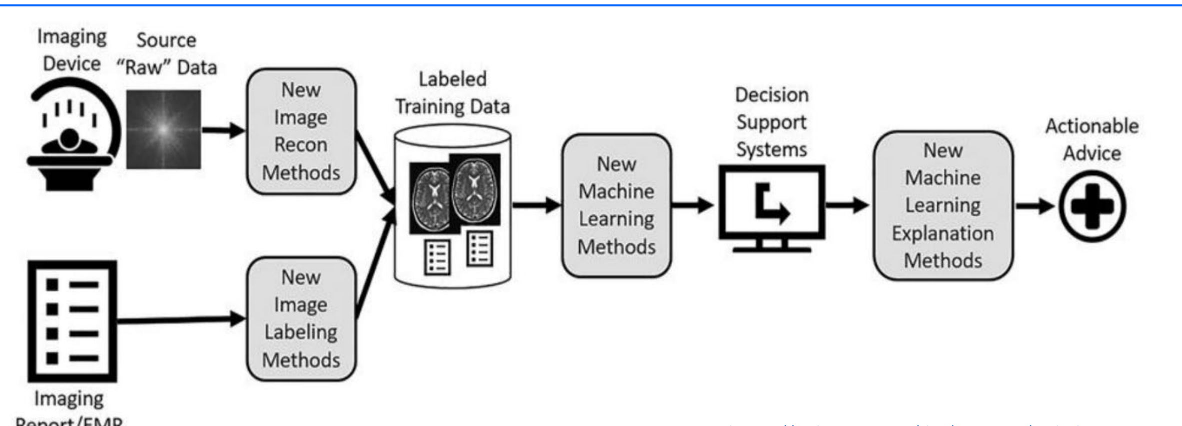
[FS-VII] Pay-A-Penny-Per-Use (**PAPPU**) data-informed decision as a service (**DIDA'S**) is a far-fetched fusion between data, science, systems and social business. Convergence of [V] and [VI] creates value from uber connectivity between sensors and may provide "meaningful" outcomes. The "meaning" will evolve if we develop tools (II, III, IV) to get the data into a form where it is anomaly-free and structured for use by logic tools in **ART**. Later, use of math-stat functions and ML. Turning statistical functions (or mathematical formulae) into algorithms, is not trivial. The rate limiting step is human resource. Teams must possess the confluence of skills to understand system science (sensor engineering), data science (stat, math, programming, computation) and *converge* it with their depth in *domain knowledge*, to address and solve, real-world problems.

11 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

This PoC is not a "software job" and professional programmers are unlikely to succeed because very few classical programmers can generate an/any optimizing sort algorithm, from scratch. What is referred to as "data science programming" is a type of app development which uses a set of pre-defined frameworks along with specific domain knowledge (in this case statistics) to create solutions. Before R and Python, "data science" was executed on spreadsheets, followed by statistical software (SAS, SPSS, statistical package for social sciences) in the era of desktops.

Data-informed service for the agroecosystem must be hyper-mobile, with high fault tolerance, operate close to the point of action (edge) in near real-time, accommodate engineering elements which are diverse, operate seamlessly agnostic of data standards or structure and must service consumer/user demands which may change often or fluctuate rapidly. Dynamic composability of data and synthesis of information, relevant to the context, is the desired outcome, even for **ART**.

[FS-VIII] Entrepreneurial Innovation – Users may pay for this service (see PAPPU DIDA'S) if we create a visual tool for non-expert end-users to grasp the curated information in SENSEE and how it may integrate with sensor data (**ART**) and machines (digital twins). To democratize access, this PoC advocates intuitive/cognitive maps (re-think Lego Mindstorms and topology optimization software, for example, see www.ansys.com/) which offers "drag and drop" icons (tactile, haptics) to orient users and catalyze the connectivity and complexity with lucidity, clarity and brevity. **Educating** and enabling users to make sense of the organization of unstructured knowledge will immensely demystify "blackboxes" and aid in the diffusion of **ART** tools, leading to reasonable adoption. By providing a mechanism to represent existing information, knowledge graphs in logic tools describe and enable access to other information sets. Diffusion of **ART** may lead to a better educated crowd and improve crowd-sourced (farmers, growers) **architecture to access knowledge.** It is a prelude to the development of an open platform for convergence of data and information, in a meaningful context, to move beyond logic tools in **ART** to data-informed **DIDA'S** and then to ***knowledge-informed*** decision as a service (**KIDS**). In 1980's decision science, DIDAS (https://link.springer.com/chapter/10.1007/978-3-662-21637-8_2) was a control theory concept. The use of DIDA'S in this document is also about decisions and, in principle, it may resonate with "DIDAS Family" (for automatic control). The sense of **DIDA'S** in this PoC is a step after **ART** and before **KIDS**. Beyond knowledge, the extraction of 'experience' may enrich the outcome from **KIDS**, but it will be difficult and must include agent-based selection (**ABS**).



https://pubs.rsna.org/doi/10.1148/radiol.2019190613

In the short term, to deploy **ART**, we expect to create SENSEE 1.0 and improve its ability to deal with a broad range of questions, before sourcing sensor-specific data for SENSEE 2.0 PoC.

1) Which sensor in the McLamore lab has the highest sensitivity?
2) Which sensor in the McLamore lab has the lowest LOD?
3) Which sensor in the McLamore lab has the highest selectivity?
4) Which sensor in the McLamore lab has the fastest response time?
5) Which sensor in the McLamore lab has the highest durability?
6) What is the most durable glass capillary sensor?
7) What sensors can be fabricated on conductive paper?
8) What sensors can be made with nanocellulose?
9) What sensors can be made with cabbage extract/anthocyanin?
10) What sensors are used for hydroponics research?
11) What sensors are used for irrigation water research?
12) What sensors are used for cell culture research?
13) What sensors are used for lake water research?
14) What sensors are used for wastewater research?
15) What sensors are used for plant roots research?
16) What sensors are used for coastal monitoring/seawater research?
17) What sensors are used for tissue culture research?
18) What sensors are used for stem cell development research?
19) What sensors are used for differentiated stem cells/neurons research?
20) What sensors are used for wound dressings research?
21) What sensors are used for osteoblast/osteoclast research?
22) What sensors are used for INS1 cell research?
23) What sensors are used for blood research?
24) What sensors are used for human tears research?
25) What sensors are used for mouse pancreas research?
26) What sensors are used for honeybee wax research?
27) What sensors are used for honeybee honey research?
28) What sensors are used for saliva research?
29) What sensors are used for food product research?
30) What sensors are used for food packaging research?
31) What sensors are used for juice research?
32) What sensors are used for soup/broth research?
33) What sensors are used for ice cream research?
34) What sensors are used for drinking water research?
35) How many sensors measure H+/hydronium ion/hydrogen?
36) How many sensors measure NH4+/ammonium ion?
37) How many sensors measure NO/ nitrogenous radical/nitrous oxide?
38) How many sensors measure H2O2/O3 oxygen radical/hydrogen peroxide?
39) How many sensors measure DO/dissolved oxygen?

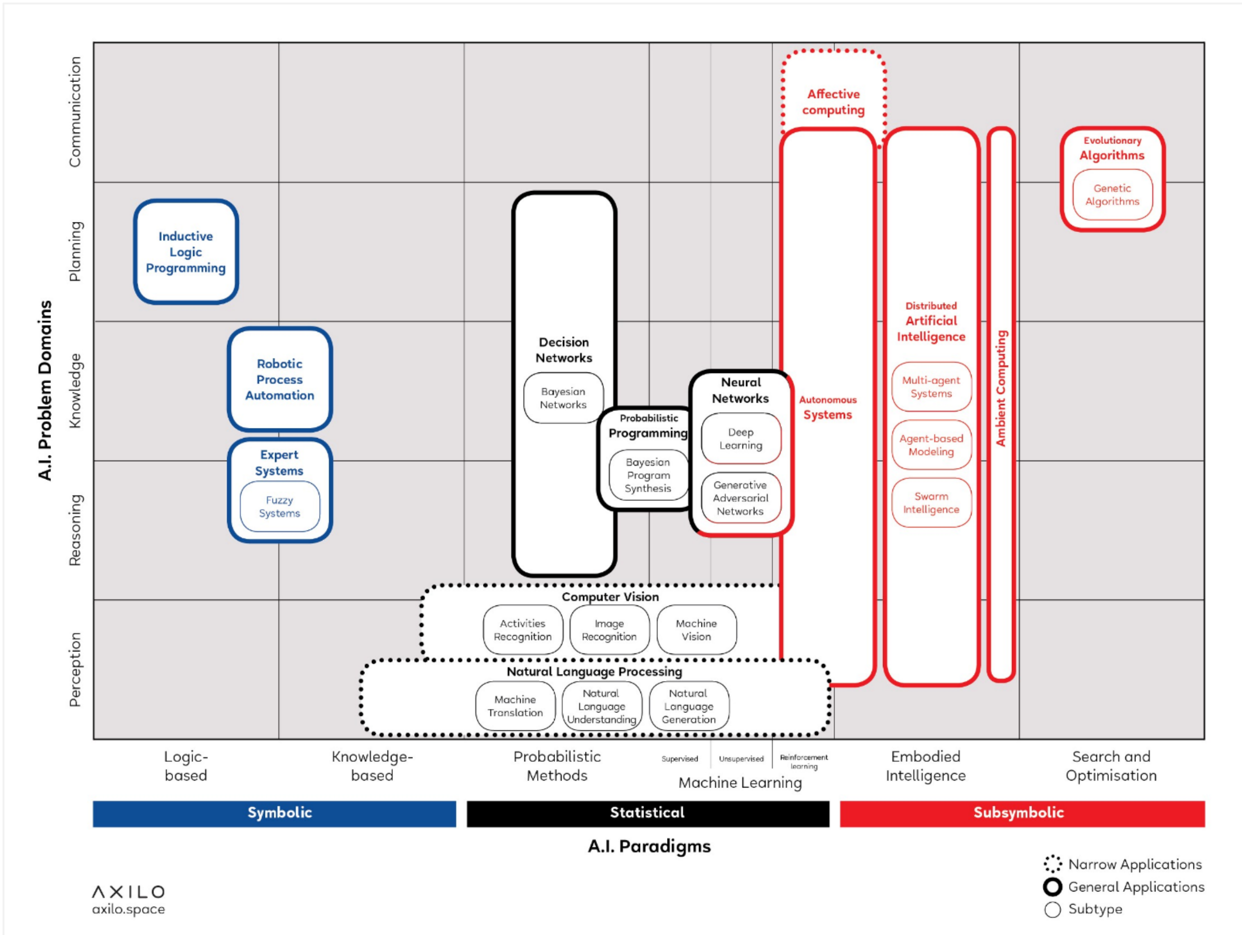
40) How many sensors measure K+/potassium ion?
41) How many sensors measure Ca2+/calcium ion?
42) How many sensors measure CH3COO-/acetate?
43) How many sensors measure NO2-/nitrite?
44) How many sensors measure NO3-/nitrate?
45) How many sensors measure Ag+/silver?
46) How many sensors measure histamine?
47) How many sensors measure glutamate?
48) How many sensors measure catecholamine?
49) How many sensors measure indole acetic acid?
50) How many sensors measure malate?
51) How many sensors measure glucose?
52) How many sensors measure ionic mercury?
53) How many sensors measure methyl mercury?
54) How many sensors measure paraoxon?
55) How many sensors measure ATP/adenosine triphosphate?
56) How many sensors measure MBF1/multi bridging factor 1?
57) How many sensors measure interferon gamma?
58) How many sensors measure superoxide dismutase?
59) How many sensors measure E. coli?
60) How many sensors measure E. coli O157:H7?
61) How many sensors measure Salmonella?
62) How many sensors measure Listeria monocytogenes?
63) How many sensors measure Campylobacter?
64) How many sensors use carbon nanotubes?
65) How many sensors use graphene?
66) How many sensors use graphene oxide/ GOx?
67) How many sensors use graphite?
68) How many sensors use glassy carbon?
69) How many sensors use liquid ionophore membrane?
70) How many sensors use solid state ionophore membrane?
71) How many sensors use nanoplatinum?
72) How many sensors use nanoceria?
73) How many sensors use nano titanium dioxide/nTiO2?
74) How many sensors use nano zinc dioxide/nZnO2?
75) How many sensors use platinum porphyrin dye?
76) How many sensors use fractal materials?
77) How many sensors use nano palladium/nPd?
78) How many sensors use diamine oxidase?
79) How many sensors use aptamer?

80) How many sensors use antibody?
81) How many sensors use lectin?
82) How many sensors use phage?
83) How many sensors use alkanethiol?
84) How many sensors use nano copper/nCu?
85) How many sensors use copper oxide/Cu2O?
86) How many sensors use phosphotriesterase?
87) How many sensors use chitosan/CHI?
88) How many sensors use PNIPAAm/ poly(N-isopropylacrylamide)?
89) How many sensors use hydrogel?
90) How many sensors measure CIP2A/ Cell Proliferation Regulating Inhibitor of Protein Phosphatase 2A?
91) How many sensors measure internalin A/ InLA?
92) How many sensors use concanavalin A / ConA?
93) How many sensors use mannose-binding lectin ?
94) How many sensors use C-type lectin?
95) How many sensors use specific intercellular adhesion molecule-3-grabbing nonintegrin / SIGN-R1?
96) How many sensors use wheat germ agglutinin N-type lectin?
97) How many sensors use lectin for N-acetyl-D-glucosamine (NAG)?
98) How many sensors use F-type lectin?
99) How many sensors use fucose binding lectin / FUC?
100) What sensors can be fabricated with glass capillary?
101) What sensors can be fabricated with graphene paper?
102) What sensors can be fabricated with laser scribed graphene/laser inscribed graphene?
103) What sensors can be fabricated with a platinum/iridium electrode?
104) What sensors can be fabricated with a 96 well microtiter plate?
105) What sensors can be fabricated with Pt/Ir microelectrode wire?
106) What sensors can be fabricated with a gold IDE?
107) What sensors can be fabricated with DropSense IDE?
108) What sensors can be fabricated with Zensor SPE?
109) What sensors have the lowest LOD **AND** the highest selectivity?
110) What sensors have the highest LOD **AND** the lowest selectivity?
111) What sensors have the lowest LOD **AND** the lowest selectivity?
112) What is the cheapest **AND** most reliable sensor for measuring mercury in water?
113) Which sensors can I use to detect pathogens in juice?
114) What is the smallest amount of glucose that a sensor can detect in tears?
115) Are paraoxon sensors commercially available in Canada?
116) How long does it take to detect salmonella with a biosensor?

117) How stable are LSG sensors?
118) What chemical components of hydroponic media can be measured with sensors?
119) Which sensors can be used to detect toxins in breast milk?
120) What is the simplest **AND** most cost-effective sensor platform for detecting glucose in buffer?
121) What sensors are available for testing pollutants in lake water?
122) What is the most common material used for fabricating hydrogen sensors?
123) What is the most versatile glucose sensor?
124) What's the most popular material platform for building ammonium sensors?
125) What is the lowest detection limit for ATP sensors?
126) How specific are calcium sensors?
127) Do listeria sensors have a high chance of producing false positive results?
128) Are wastewater sensors reusable?
129) What is the saturation concentration for glutamate biosensors?
130) Which kinds of sensors are selective towards E. coli O151:H7?
131) What's the widest operating range for a H2O2 sensor?
132) Which sensor has the lowest durability?
133) Which sensor has the highest durability?
134) What is the lowest LOD among all sensors?
135) How many different kinds of platforms have been adopted in sensor testing?
136) What is the best LOD for H2O2 testing?
137) What type of H2O2 sensor has the best selectivity?
138) Which platform should we use when we test the NH_4^+ regardless of price?
139) Does the liquid ionophore always have better performance than the solid ionophore?
140) Which type of sensor has the longest response time?
141) What is the fastest sensor in small molecule testing?
142) What is the most commonly used sample in the test?
143) Which sensor has the largest range in Glucose test?
144) Which platform is most commonly used in Mclamore's lab?
145) Is it possible to test H2O2 in ocean water?
146) What recognition-transduction scheme do we use to detect glucose?
147) Is it possible for a sensor to have a response time lower than 0.1 sec?
148) Which platform should we use when we want to detect the NH4+ in two seconds?
149) Could we use the LSG to detect H+?
150) Can we make a bacteria sensor whose response time is <500 seconds?
151) What is the response time for liquid K+ ionophore in detecting K+?
152) What is the most popular platform for hydrogen peroxide biosensors?
153) Which targets can be determined using diamine oxidase sensors?
154) In which samples potassium ions can be detected?

155) What molecules can be detected in breast milk using biosensors?
156) What is the difference in sensitivity between glucose biosensors based on graphene or platinum foil?
157) What is the most sensitive biosensor based on carbon nanotubes?
158) How many biosensors have been proposed for glucose determination?
159) Anthocyanin is used as a target for which biosensor?
160) Which biosensors can be used for hydroponic medium?
161) In which samples, glutamate was determined using biosensors?
162) Which biosensors were proposed for catecholamine determination?
163) What is the lowest limit of detection for graphene-based biosensors?
164) What is the maximal range for nitrate biosensors?
165) What platforms can be used for ammonium detection?
166) Most durable recognition-transduction scheme for interferon gamma biosensors?
167) Best limit of detection achieved with phosphotriesterase-based biosensors?
168) How many biosensors were described for ATP determination?
169) What platforms were proposed for ATP-sensitive biosensors?
170) What is the average LOD of K+ sensors?
171) Which platform could be used for selective glutamate analysis?
172) What is largest analyte/molecule for which there is a sensor in the database?
173) Is there any cost associated with any type of sensor?
174) How many labs are making sensors to detect lead in water?
175) Are there sensors to detect air-borne viruses in the air?

17 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS



DEMYSTIFYING THE ANALYTICS BLACKBOX – A PREREQUISITE FOR ADOPTION?
https://arxiv.org/pdf/1902.05009.pdf

Snake oil sales of "intelligent" tools in the name of "AI" is an anathema to those who respect, appreciate and understand, albeit in part, the immense contribution of scientists and engineers who delve into details in quest of robust, evidence-based, numerically-supported, i/o, even for the basic form of artificial reasoning. Even then, it is not a general "one-shoe-fits-all" app that can be peddled willy-nilly. If users are better equipped to ask probing and precise questions, the tools and systems (eg in this cartoon) can serve the users, perhaps with greater accuracy and precision, before the information perishes. The most common question "what happened" (descriptive analytics), may lead to "why did it happen" (diagnostic analytics). The collection of logic tools (ART) we have discussed in this document and the cartoon (above) must work in confluence to respond to the most likely follow-up question "what is going to happen, next" (predictive analytics) and then the obvious: "what is your recommendation, what should I do" (prescriptive analytics). In PAPPU DIDA'S concept of data-informed decision as a service, prescriptive analysis may suffice for human-in-the-loop systems where the actuation is human-controlled. With increasing scale of concurrent levels of operation and improving control of automation (think about 0-5 levels of autonomy, think OODA concept by John Boyd and PEAS paradigm in Agent based systems), it is not impossible that users may eventually trust and enable systems with case-specific permissions to "take action, execute" (automated analytics). Auto-actuation, in the SARS to SARA paradigm, is discussed in the essay with the title SARS♠AG.

PAPPU DIDA’S – BEFORE WE ASPIRE FOR KIDS

Irrespective of the strength of ideas, in this and other essays, the path to adoption is fraught with challenges. Therefore, ideas are often useless and without value unless we tackle the hardest questions, first, which defines the pragmatic aspect: will anybody pay to use this idea in reality? Ultimately it is the economics of technology which defines and controls the diffusion of ideas.

The complicated answer has several moving parts and none may be fully correct or incorrect. One canonical response is that adoption will be determined by the cost versus value paradigm. In a global economy where products are receding in the background and services (including those services which are based on products, eg, washing machines) are gaining momentum, the idea of “PEAS Platform for the Agro-Ecosystem” begs to ask whether one expects that users will buy sensors, of different types, in bulk. How will the non-expert users collect, analyze and integrate the data from the sensors. to help them in their real task, the task of food production?

The physical product (sensor) must deliver value (data, decision) which will inform responses and lead to actual work (actuation) to improve ag systems and help to increase food production.

One option is an age-old, time-tested, solution where lowered cost to the user (opex) is a function of the frequency of use and generally free from sunk or capital costs (capex). In the last century, this model was epitomized by POTS, the plain old telephone system, where the user paid only the “charge per call” which was reasonably affordable even when the per capita income was low.

Pay a penny per use (PAPPU) re-invents POTS with the qualifier that the user pays a penny (US) for each use (perhaps unwise to restrict it to one penny). The “use” may not be a thing, object or tangible product but rather a “process” which we refer to as data-informed decision as a service (DIDA’S). The “penny per use” idea may draw scorn from certain segments of investors and corporate leaders because the idea does not support the “next quarter” earnings (greed) report.

A version of PAPPU (pay a penny per **unit**) could stretch to fit “99 cents hamburger” model evident in PayPal’s 2018 revenue ($16 billion from 12 billion transactions, $1.25 / transaction). The “unit” view of PAPPU may be applicable in transport, energy, water (as **units** delivered).

SENSEE, ART, DIDA’S and other data-informed decision-support on an open-platform calls for synergistic systems integration. The value is realized at the “end” when systems data may be synthesized to provide meaningful use in the context of the problem. It delivers information at the point of use, at the edge. Is this of actionable value for the user? Consumers may pay only for the desired outcome. Transaction cost economics is perhaps key to this *modus operandi.*

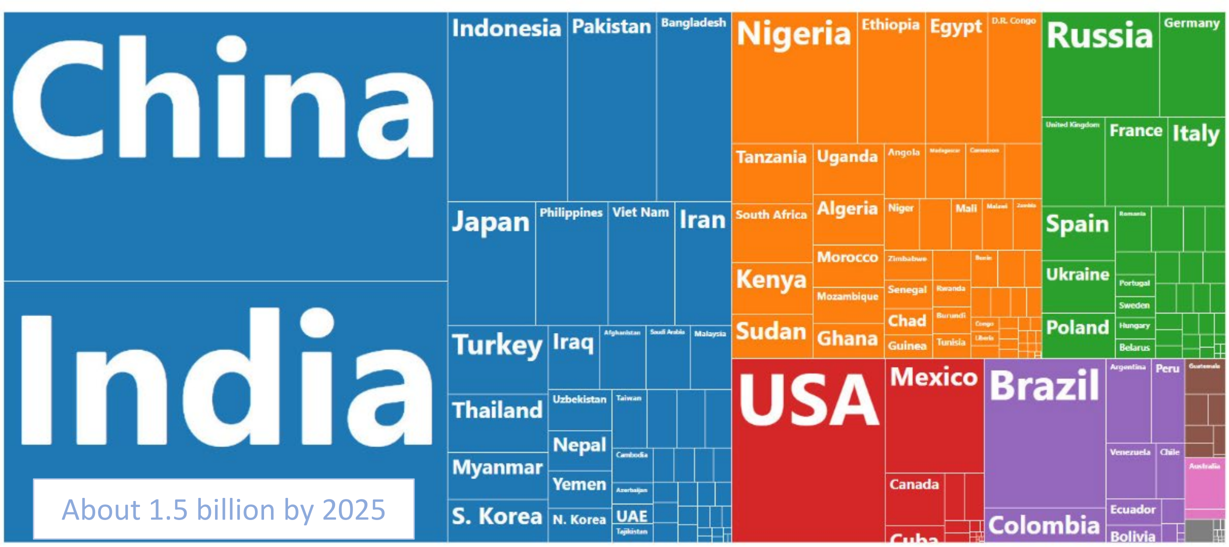
If the outcomes are dependent on a plethora of sequences in the operational process, then each process is a “profit center” and may generate a penny in revenue each time the user “touches” the system to extract information (or knowledge). If the economy can bear the economics of PAPPU then systems diffusion and adoption will continue to grow (decades) based on the economy (until saturation, when demand plateaus irrespective of cost). The number of sensors, and other data, are likely to intersect with vast number of decisions (ART, DIDA’S, KIDS). The actual *transactional volume* of payments, from ‘micro’ or ‘nano’ payments, are potentially gargantuan.

Documenting that the system was “touched” and billing/collecting that one penny is a technical challenge which requires tracking events (think IPv6, as an “indicator” for ***system*** activity). The task of segmenting that one penny revenue, between several service providers, may be a massive challenge in “weighted” decomposition/recomposition of events, to distribute earnings based on the degree of contribution of the provider who executed that step/event (for example, sensor manufacturer, systems integrator, platform provider, software vendor, analytics, mobile fintech).

Since no new “physics” is necessary to delineate these processes, it is safe to state that these can be accomplished without any invention but with forward thinking imagination and innovation. It is a déjà vu scenario from the “Store of the Future” (2000-2001, RFID track and trace) which sputtered and asphyxiated in the face of systems integration challenges, only to be resurrected by Amazon, which, finally, implemented the retail concept in Amazon’s GO (September 2018).

Increasingly, PAPPU (DIDA’S) will be the monetization mantra for the ART-IoT generation and the future where equality, equity and égalité may re-claim its rightful place in society striving for ethical profitability. It may take 20-30 years to overcome the resistance from despots, investors and corporate behemoths, but eventually the infectious spread of this concept may succeed in sowing a critical-mass of practitioners. The concomitant growth of infrastructure (for example, affordable access to low latency, reduced jitter, high bandwidth wireless telecommunications, 5G, trusted mobile banking) may be necessary to pave the road for PAPPU in ART and DIDA’S.

The ability to escape the dead weight of old technology (eg Africa, Asia) may accelerate the implementation of ***pay a penny per unit*** (PAPPU) as an integral part of the socio-economic fabric of a product-less, service-based economy, which may exclude the tiny population residing in OECD nations and/or the red and green zones in the cartoon show below.



PAPPU may evolve as a preferred business model for the global economy by lowering the barrier to entry into markets where people are surviving on about $2 per day. The impact may be especially profound on healthcare and the agroecosystem for production of food.

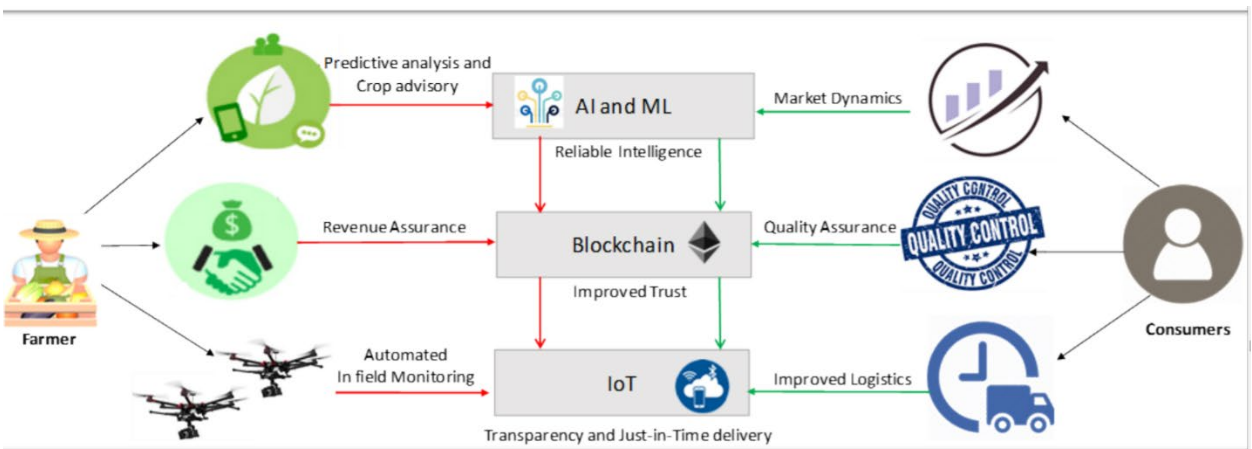
20 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

Because PAPPU is inextricably linked to mobile payments, distributed banking and digital finance, the pundits of social media will jump on this discussion to claim PAPPU is incomplete unless "blockchain" is integrated in the process. Blockchain hype-mongers are worse than snake oil sales and the adage or aphorism "hammer in search of a nail" seems too respectful in view of the torrent of garbage that is spewed in the name of blockchain. However, trust in transaction is undeniably central and an age-old concept (https://www.jstor.org/stable/20752121).

Therefore, it is important for PAPPU to provide tools to ensure safety of the payment system and other steps where verification guarantees are related to the service or product (for example, food safety). But, informed organizations may not, blindly, consider blockchain security for PAPPU.

Whether and how and in what form the concepts in blockchain may be helpful, remains to be seen. It is not entirely useless and such "solutions on steroids" deserves a place in society to counter the unethical practices that rapidly multiply in financial operations. However, such specific examples of use, and value of blockchain, may not be ***generalized*** as a solution for ***all*** levels of transactions. It is deceitful and malicious for blockchain proponents to tarnish all verticals and industries using the broad brush of finesse that is rampant in the financial industry.

Blockchain is erupting into an euphemism for avarice, for the sector of people involved in the process of marketing tools for blockchain. It is an anathema for >80% of the world trying to survive beyond the gluttonous grip of tools and technologies of dubious value. Blockchain is certainly not a panacea. There may be a few other low-cost ways to achieve safety, security, identification and authorization (for example, https://dspace.mit.edu/handle/1721.1/102893).



Chacun voit midi à sa porte – hammer in search of a nail: peddling the "blockchain" at the "center of the world view" of operations. It is not necessary for individuals in trains, planes and automobiles to wear an armor-suit. The safety belt is sufficient, although it may not be enough, in certain instances. The latter is the risk that emanates from the rewards due to progress, which society has, and will continue to, shoulder. Rather than feeding people, the burden of blockchain will starve the hungry, where food is most needed, by increasing cost of operations. Imposing rules and regulations will secure profit for the blockchain industry, deliver little for food safety and deprive nations from food. (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8598784)

FAAQ - FREQUENTLY ASKED ANGRY QUESTIONS

■ Is this really a PoC ?

■ What is the purpose of repository (pg 11) in step V ?

Explanation of the lack of purpose: unclear from SENSEE PoC alone

Steps I and II (see page 5) are indeed proofs of concept to show we can create a simple mobile dialog box to ask questions (examples on page 3) about types of sensors available from a small group of labs (10-100) who are creating sensors. One example from the McLamore Lab is in the xl sheet available from http://bit.ly/SENSOR-LIBRARY-ERIC-MCLAMORE

The important distinction, with respect to this discussion, is that, this is ***not about sensor data alone***. We discuss ***types of sensors*** (SENSEE 1.0), ***sensor-specific data*** (SENSEE 2.0) and phases of decision support (ART, DIDA'S, KIDS – PEAS Platform for the Agro-Ecosystem).

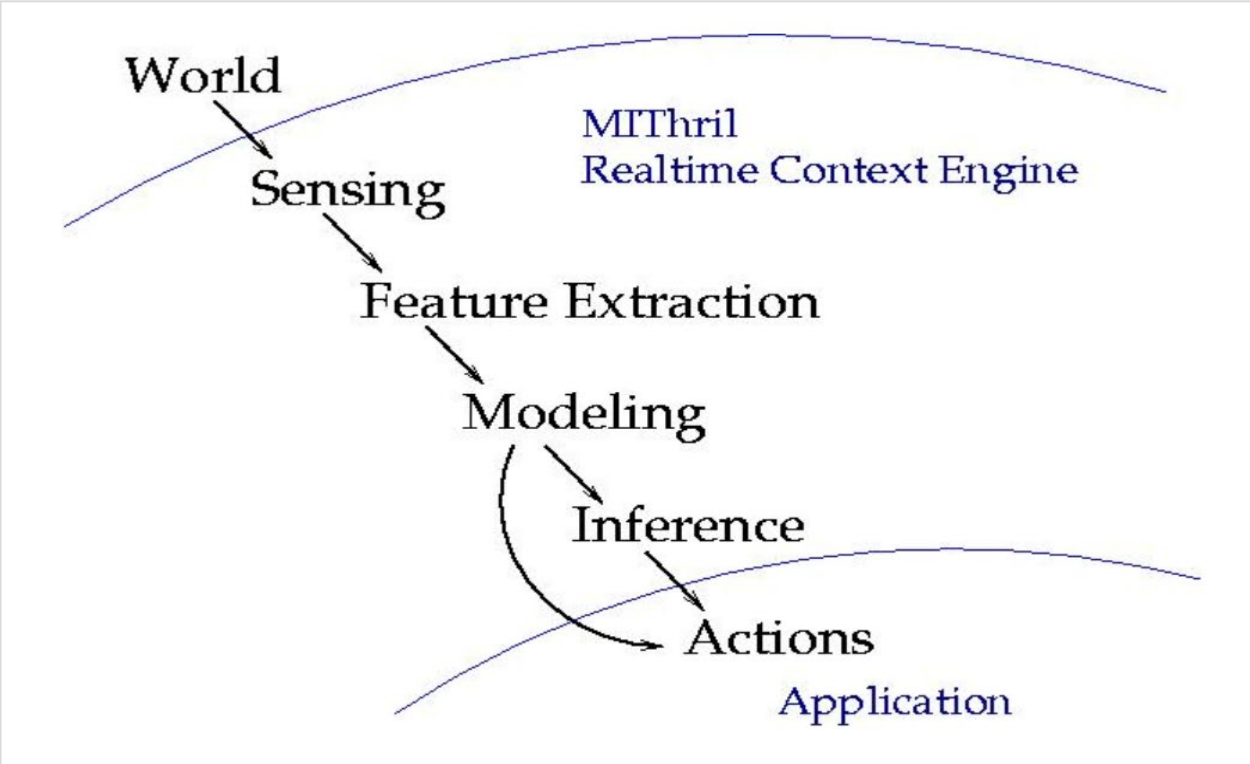
SENSEE 1.0 is the PoC which refers to characteristics, properties and features of sensors, as engineered by academic or industry labs, (tabulated in xl sheet). When we mention "users" typing in a query in the dialog box, the "user" is **NOT** a farmer, grower, consumer. SENSEE 1.0 users are experts and sensor engineers or students exploring the repository (step V). Who has made what sensors? What is the limit of detection? What platforms (impedance spectroscopy, photonics, SERS) were used to capture the signal? Critics may point out that Google has done that job and Google will retrieve millions of references to papers and documents with any key word. In defense of our aspirational step V (the repository), we wish to point out that the queries (examples on page 3) using the SENSEE 1.9 dialog box (step I) may reveal specific information, from a ***curated catalog*** of information, related to sensor science and engineering.

> ***Necessary Digression:*** In the context of the "big picture" it may be worthwhile to understand that this "curated catalog" refers to the xl sheet with sensor specifications (example provided from McLamore Lab). In step II we will have, hopefully, extracted similar information provided by another few labs. This collection of information will be in databases (SQL, noSQL, GraphQL, SPARQL, TSDB) which will serve to retrieve responses to questions (example on page 3). Hence, this "curated catalog" in this discussion may lead to the repository (see step V on page 11). The reason for this preface to the "digression" is to point out how this "curated catalog" may morph into a "semantic data catalog" as explained in **Figure 12** in the **"SITS"** section of the essay series "SIGNALS" which may be downloaded from http://bit.ly/SIGNALS-SIGNALS. In steps I through III we may use elasticsearch and NLP to parse words in a query and match with keywords (in search engine) to retrieve the most likely answers. This *syntax-driven* process is error-prone (depends on how well the human developer has coded/trained the engine). To boost performance, semantics may remove, albeit partially, some errors due to syntax. The latter varies with expertise, mother language and social environment. Hence, the intended move to semantic databases may be facilitated by transforming this "curated catalog" (this discussion) to the future ***semantic data catalog***.

One may conclude this is a repository for experts. At this time, this ***is*** true. This is the beginning of an ambitious attempt toward ***synergistic integration of platforms*** (SIP) which will converge data (**ART**), information (**DIDA'S**), knowledge (**KIDS**) and, eventually, *experience*, to suggest solutions. Users may be experts from industries, farmers, meat packers, distributors, food inspectors, grocery stores, cold chain logistics providers. It may be **any user** who may benefit from data-informed decision as a service (DIDA'S) and may even pay for the service (PAPPU). The future of ***experience as a service*** may be your personal mobile agent, which you lease to a buyer, or air-drop (www.imore.com/airdrop) to a local client, interested to profit from your experience.

Broad spectrum of users (above) may need different types of information, which may be in different databases. Hence, the choice of knowledge graph connectivity to synthesize and deliver a meaningful response, by selecting data and information using distributed architecture to access a multitude of resources (see figure 11 on page 28 in the "SITS" section of "SIGNALS" http://bit.ly/SIGNALS-SIGNALS). Extraction of data and information from the "world" may benefit from **context engine** architecture (cartoon below). The latter may be one way to create knowledge bases without reliance on ontologies, using publish/subscribe (ingest from CSV, xl, relational databases, JSON/XML feeds), perhaps in a manner analogous to CMS (content management software for data). Ontologies may become key to future knowledge extraction.

This endeavor is **ONE of the resources** (SENSEE 1.0) we aim to develop to address that *future* **SIP platform**. We expect that aggregation of **contextual data** and ***curated*** information may further improve the performance of this basic service (ART) through connectivity with other distributed resources (see **"SARS♠AG"** in SIGNALS http://bit.ly/SIGNALS-SIGNALS). To achieve that goal, the open source platform must support data interoperability (for example, DDS) between local and global databases/platforms, enable dynamic composability to pick and choose (drag & drop) data/information from diverse sources, always explore user-friendly tools for synergistic integration with domains of data, information and ***crowd-sourced*** knowledge, which may enable user ***experiences*** from the past to inform the future. Also, we expect ***actual "world" sensor data*** (eg temperature sensor) to be aggregated, ***agnostic*** of the make and model of the sensor (SENSEE 2.0). Sensor *data*, and extracted *information*, may be *more useful* for pragmatic, and profitable applications, in the near-term, that we may deliver through **ART**.



The 'world' in context of applications. www.media.mit.edu/wearables/mithril/context/index.html

23 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

Science and engineering have enabled an embarrassing wealth of sensors but without an organizational repository (aspirational step V on page 11) the value of these sensors may remain under-utilized. The proposal to create a World Sensor Organization (**WSO**) to address these issues, remains unexplored (Commentary [C] discusses WSO, in the PDF, **"IoT is a Metaphor"** which may be download from https://dspace.mit.edu/handle/1721.1/111021).

To catalyze science to serve society, in a parallel endeavor (Eric McLamore, personal communication), experts interfacing with the edge, that is, with end-users (farmers, growers, meat packers, aquaculture, retail grocery suppliers), are attempting to harvest questions which end-users may ask or *should* know, in order to better use data to inform and transform their practices (address contamination, understand regulation, use of technology). The measure of success is the outcome (***PEAS Platform for the Agro-Ecosystem***), in terms of food production at an affordable cost, of a better quality, as well as quantity, using ethical tools and practices.

To clarify, this PoC may be divided into an actual proof of concept phase (steps I and II) and a R&D approach in steps III through V (which is no longer just a PoC but a more thoughtful path to step V). The sum of PoC plus R&D is an essential (***but only one***) part of the SIP platform concept discussed in SARS♦AG (see essay **"SARS♦AG"** in http://bit.ly/SIGNALS-SIGNALS).

SARS♦AG combines the tools and sensors discussed in "SITS" (see essay **"SITS"** in SIGNALS http://bit.ly/SIGNALS-SIGNALS) with questions that **end-users** may want to ask. This ***combination*** makes it possible to bridge the wealth of advances in sensor research with the need for tools and technologies, on the ground, at a pragmatic level. Data-informed end-users may meaningfully converge this knowledge, with their experiences, in order to improve the outcomes (food production, food distribution, food safety, prevent food wastage, profit margins).

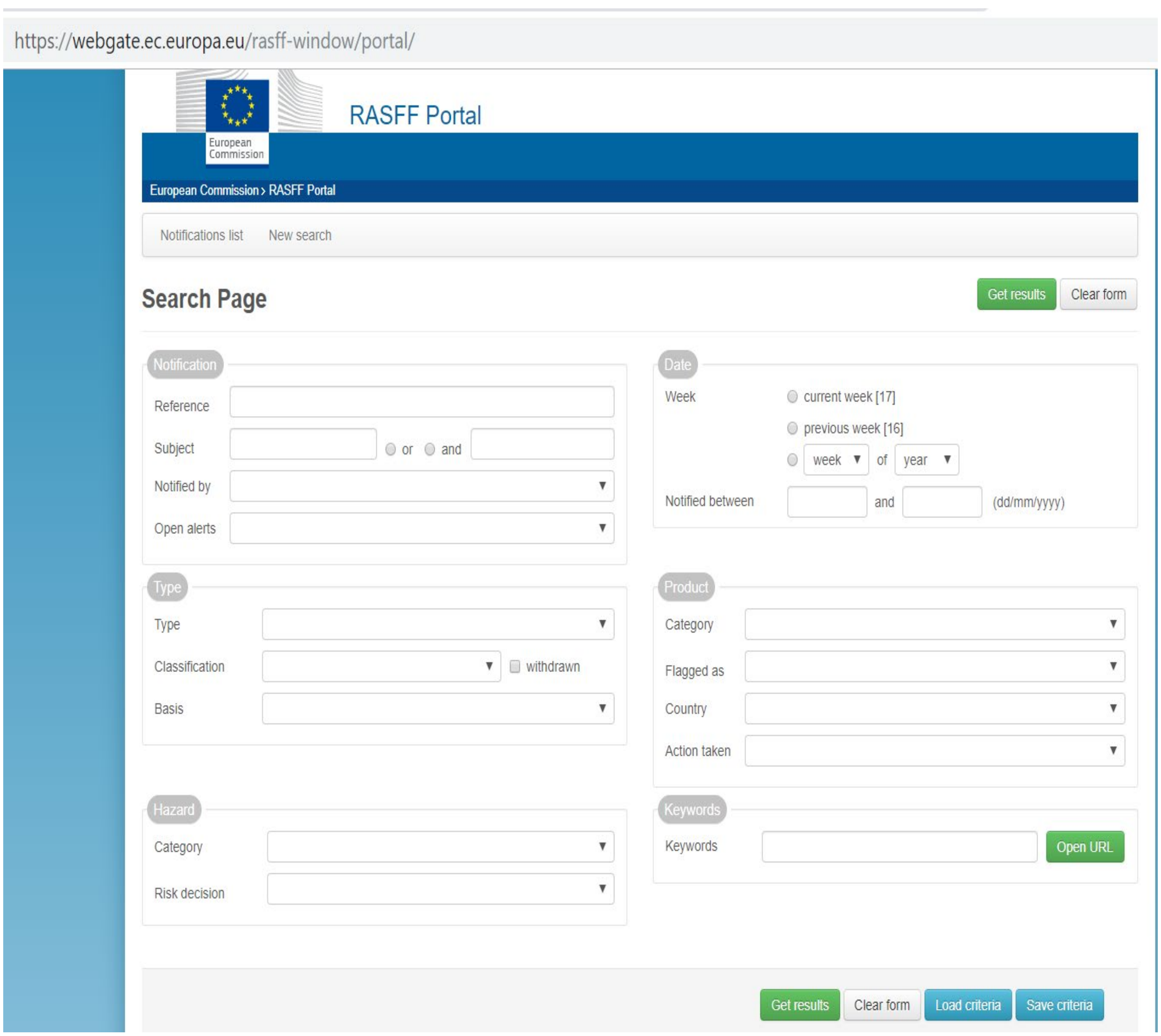
Questions, whose answers may help end-users, are the sign-posts for the development of ***PEAS Platform for the Agro-Ecosystem***. Sensor repositories (SENSEE 1.0, SENSEE 2.0) must meaningfully connect with questions from the field-workers. Some of these questions may have nothing to do with sensor science and engineering. Thus, a tremendous amount of analysis must be invested in understanding, classifying and identifying the nature of the sources we need to connect, in order to answer some of the questions from end-users. It may be clear to the reader why multiple sources of data and connected information (knowledge graphs) may be essential.

The RASFF portal (next page) may be an example which can be adapted, in principle, to guide end-users to ask questions and organize the input in "buckets" or holders. In this approach, the input data may be amenable to classification or clustering algorithms, to help sort out the nature of the questions. If we allow "question collection" to an open format (write down top 10 questions) using an open dialog box, where anybody can ask anything, in any form, using syntax devoid of context, then extracting the key ideas from this unstructured mess (without standard keywords) may be frightfully exhaustive, if not impossible.

The RASFF approach could use keywords and harvested frequency of words or terms from this question-gathering exercise. Using tools like PCA (principal component analysis), it may be relatively easier to identify the topics covering 80% (Pareto principle) of the questions.

Using these topics, as a guide, we can begin to build / connect with contextual domains of data (for example, micro-climate from federated nano-satellite weather channels), information (example, price of bio-diesel) and knowledge (example, crowd-sourced experience of end-users, elsewhere). When synthesized, it may help us to respond, in near real-time, appropriately, to the end-user, delivering ***actionable information***, perhaps 80% of the time, with respect to desired level of relevance, precision, accuracy and value, to reach a certain quality of service (QoS).

Theoretical discussions about questions, data and platforms, using power-point filled with boxes, with arrows and artificial acronyms, is easy. Providing meaningful value to the end-user is not easy. We shall strive to combine data and logic informed approaches in the context of case-specific problem-based artificial reasoning tools (**ART**). The outcome and quality of service (QoS) remains to be determined. The aspiration is to approach DIDA'S after critically evaluating the successes and failures from ART implementations, in real world scenarios with actual clients. The journey to KIDS is still amorphous, as outlined in ***PEAS Platform for the Agro-Ecosystem***.



25 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

Challenges in Knowledge Extraction & Application: *KIDS is a journey, not a destination*

The assumption in ***"actionable information"*** is the strength of ***credible content*** which induces humans in the loop to perform a process or execute a specific action. The ***trust*** in this suggestion is dependent on the depth of the connectivity between system of systems and the ability of the tool (ART) to collect, synthesize and propose a meaningful outcome. Hence, the process of delivering ***value*** for the user in terms of "actionable information" is ***not*** an instant step. It may be best described in terms of bio-mimicry. For example, if you ask a 5-year old about "errand" planning (grocery store, library, co-op, laundry), the answer may be correct or incorrect because the 5-year old may not know the locations, what you need at each location, store hours and if the traffic on the road may change while you are between errands. If you ask the same question to a 15-year old, she uses Waze and store hours of operation to customize a Google map with a sequence/pattern you may wish to follow, based on data and information (data-informed decision as a service). The 15-year old has "learned" how to plan and manage time, fit the process to parameters of family's needs, and intuitively, understands semantics.

This PoC, SENSEE, ART, DIDA'S, KIDS, hence, are sign posts on the road ahead. We continue to learn, improve accuracy, precision and credibility, to increasingly gain the trust of the user. We continue to explore tools to address long-term "learning" and apply the results.

In the real world, tools often lead to questions about standards because a tool is not an one-off product. Standardization is viewed as an unifying process (for example, IPv6), which enables creation of tools agnostic of where (location) it is used or manufactured, as long as it is in compliance with standards. However, dynamic systems involved in decision making may be hard to standardize, primarily because of geo-political and socio-economic factors with respect to the decisions and the impact of those decisions.

Ecosystems are in a perpetual quest to develop and deploy advances in standards, which can be driven by adoption (for example, Android and Windows operating system). But, standards of decision making are far from homogeneity. The diffusion of any standard operation procedure in terms of decision making depends on the strength of measurement science (including data), interoperability of information between systems and software tools which can ***combine data, analytics and information with knowledge and experience*** (aspiration in ART, DIDA'S, KIDS).

This PoC "plan" is the ***beginning*** of analytics, which, theoretically embraces experience. ***PEAS Platform for the Agro-Ecosystem*** attempts to bring together the value of aggregating and querying sensor descriptions. Conventional wisdom suggests that sensor data and analytics, as "information" is more useful for end-users. The description of sensor types (PoC for SENSEE 1.0) is a step before jumping into sensor data.

It may help to remind the readers that this ***planning document*** is relevant to one idea from the suggestions in the series of essays (SITS, SIP-SAR, SARS♠AG and PEAS). How many types of sensors are available for any given task? What are the characteristics of the sensors in terms of signal detection, sensitivity and other key attributes? This may not interest the end-user in a farm, but it is critical to the ***design of field deployment*** and those in academia and industry.

This PoC creates a foundation with SENSEE 1.0 that does not contain sensor data but contains sensor descriptions (http://bit.ly/SENSOR-LIBRARY-ERIC-MCLAMORE). In its first version, we aim to collect sensor descriptions from about 1,000 labs using a partly automated document management system to populate the database (SENSEE 1.0). Using a simple mobile web service type app, (please request the IP address of the web service if you wish to test the usability of the app), experts may query SENSEE 1.0 to ask direct/specific questions (examples: at the beginning of this document, page 3) or use Boolean operators (for example, how many sensors use laser scribed graphene *and* plasmon resonance spectroscopy for signal transduction).

It is the aim of step II of this PoC to demonstrate [1] the ability of SENSEE 1.0 to contain a *critical mass* of sensor descriptions in the form of ***curated*** value fields and [2] the ability to answer a variety of questions using the simple app (user interface, dialog box) developed in step I. The ability to answer questions, group values using keywords and other combinations, depend upon the natural language processing (NLP) tools to be developed. Initially, the questions may be limited to those which may include keywords the rudimentary string parser is able to handle. This is a *learning* process which we anticipate will improve over time (if we invest) and presents opportunities to explore creative ideas (https://openreview.net/pdf?id=rJl-b3RcF7).

SENSEE 1.0 may provide an upload/ingestion tool for sensor labs to upload their sensor descriptions (xl file) and use interactive tools to clean/curate the fields if the ingested "data" is in the incorrect field (for example – incorrect column heading – limit of detection score inserted in the column with the heading "molecular weight in Daltons/kiloDaltons" of molecule/analyte). The utility of SENSEE 1.0 may be limited unless we have a critical mass of sensor categories and attributes in the database. Experts and students may find this tool to offer specific answers compared to PubMed or Google search using key terms. It is hard to imagine farmers, growers and meat packers, who may be interested in sensor characteristics, for example, detection of ammonia using microfluidics versus laser scribed graphene (LSG) sensors.

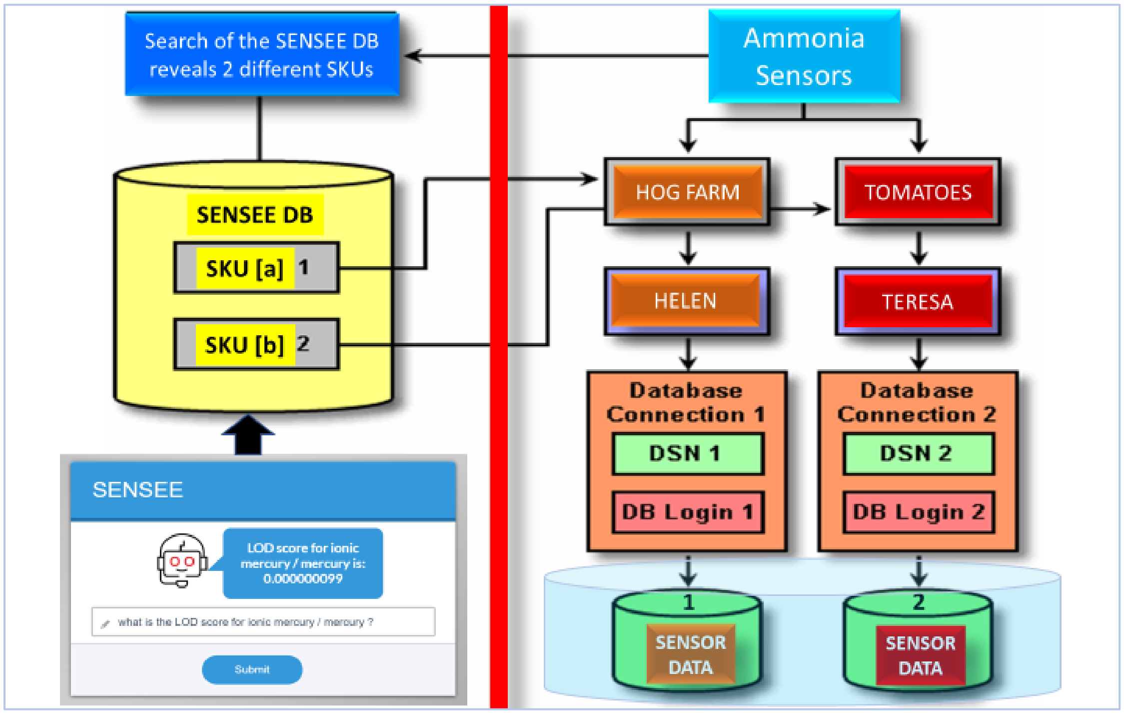
SENSEE 1.0 will continue to strengthen the search function (NLP), automating data ingestion tools, developing curation and data clean-up codes (step III). The research quest is in step IV, where the suggestion is to develop feature automation tools. ***Through all these steps we are still discussing sensor descriptions (types). We have not yet discussed sensor data.***

Each sensor, for example, ammonia sensor using [a] glass capillary or [b] LSG (as platforms) are independent sensors (***different SKUs*** or stock keeping units, different items). If sensor [a] or [b] is manufactured, then it can be used to detect/quantify ammonia gas. If sensor[a] is manufactured in a small batch (100 widgets) but sensor[b] is manufactured in a large batch (10,000 widgets), then SKU[a] may have the serial numbers 1-100 and SKU[b] may have the serial numbers 1-10,000 (hypothetical).

This digression is critical to grasp the distinction between SKU and serialization. It is important for the discussion about data from sensors (SENSEE 2.0) with respect to tracking and tracing (id) of sensors, when they are in the field (often tagged with RFID tags for identification purposes). Data acquisition from sensors in use (in the field) must be specific for sensors which are related to a specific case (real world client) which is a part of a problem that we are trying to solve for the client (test bed, customer). Sensor data is the topic for the next PoC, referred to as SENSEE 2.0 in ***PEAS Platform for the Agro-Ecosystem.***

SENSOR DATA – SENSEE 2.0

At the hand of hog farmer Helen, sensor SKU[a] serial numbers 1-5, is now capable of generating ***sensor data (measuring ammonia)***. At the hand of tomato grower Teresa, sensor SKU[b] serial numbers 10-25, is now capable of generating ***sensor data (measuring ammonia).***



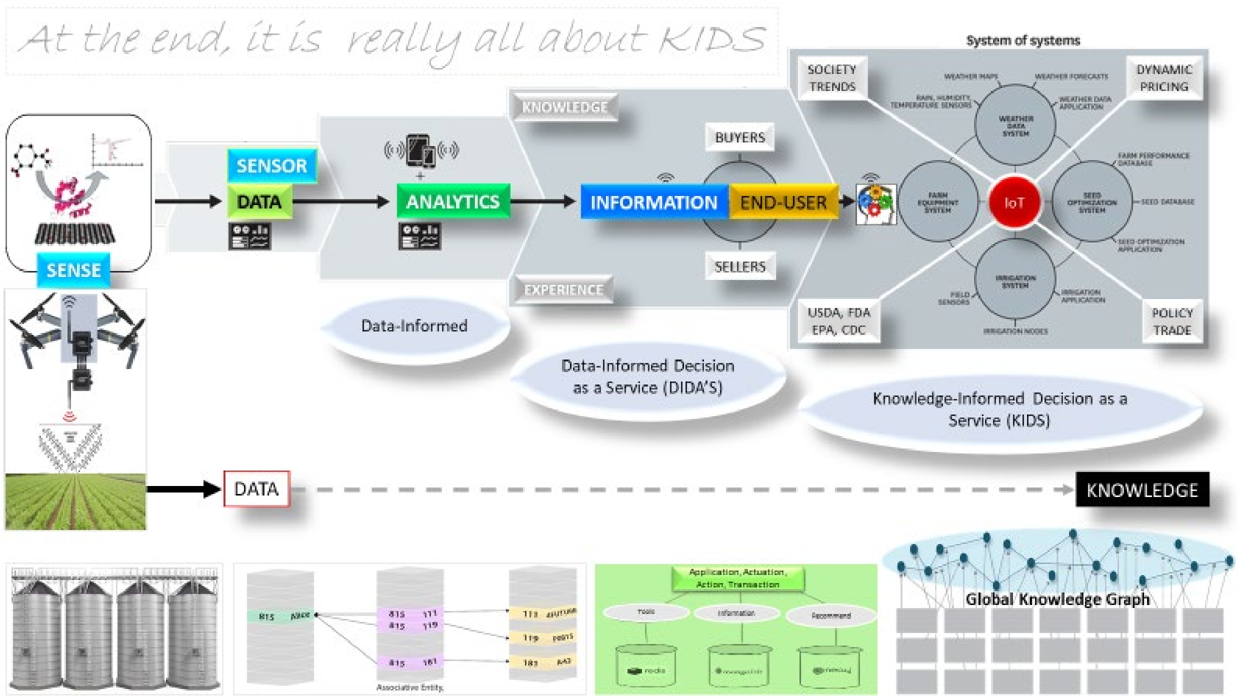
This PoC is about the LEFT side (SENSEE 1.0) where sensor descriptions are in SENSEE DB. Use of the sensors will generate sensor data – sensor data acquisition (right) is for SENSEE 2.0

This PoC does not address the sensor data database. That is the next step (SENSEE 2.0) where data is sensor-specific, case-specific, problem-based, has a purpose (end-user directed). It is used in conjunction with logic tools and data analytics in order to evolve as information expected out of the umbrella of artificial reasoning tools (ART). The latter is a step toward curating information and knowledge with respect to DIDA'S (data-informed decision as a service). ART may "crawl" for a while till it gains strength in its logic spine to walk, albeit slowly, to reach DIDA'S and hope for KIDS (knowledge-informed decision as a service).

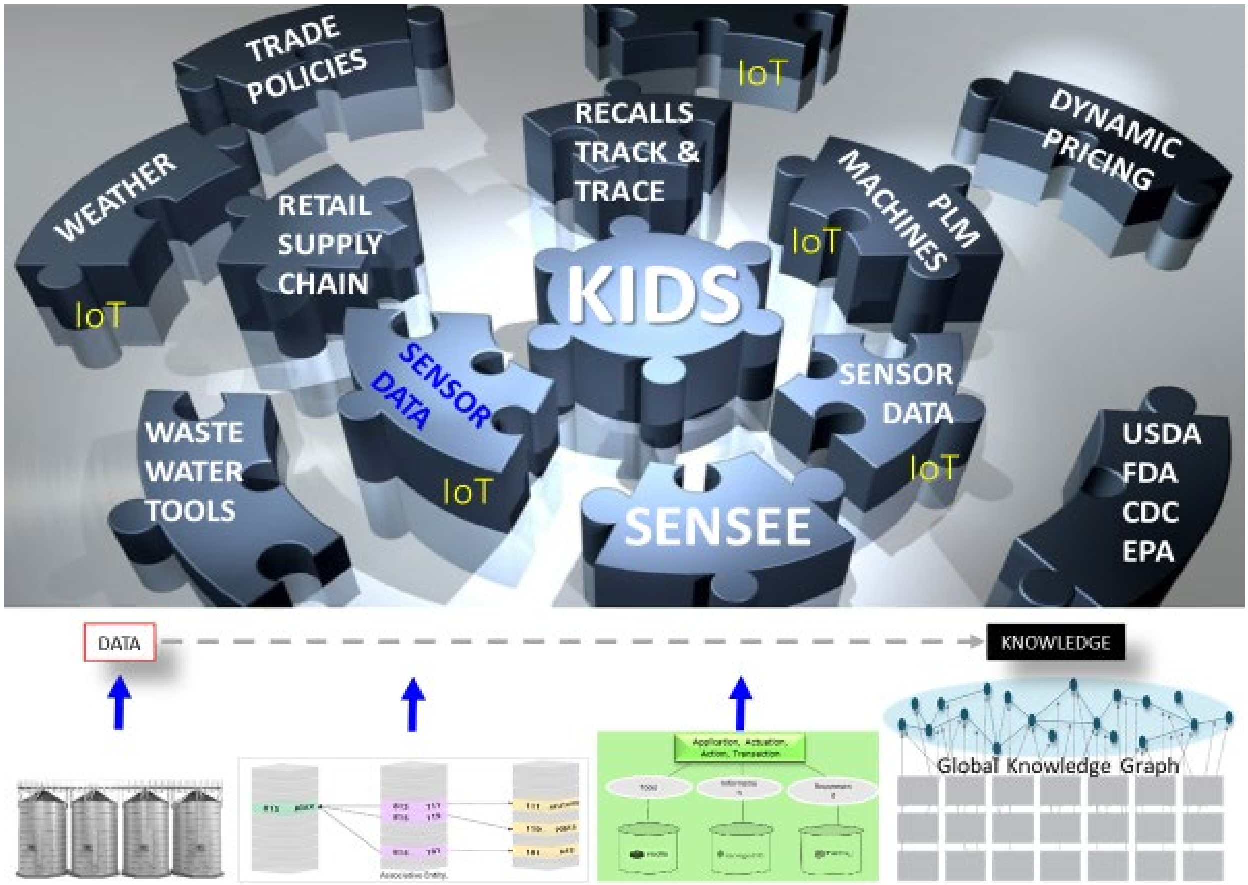
The range of questions in the next phase will include questions about sensor data. For example, the lab or manufacturer of ammonia sensors may query SENSEE 2.0, which sensor is more stable for higher concentrations of ammonia? Helen may ask when are the hogs producing the maximum volume of ammonia? Teresa may ask is there a difference between the ammonia concentrations during dawn versus dusk? **ALL** these questions are different but relying on the data from specific sensors (why data persistence is central to ***different views*** of data analysis and analytics). To provide value to end-users, ***SENSEE 2.0 and SENSEE 1.0 can be independent but connect when graphs "discover" them in the process of identifying resources for a case.***

DATA-INFORMED 2 KNOWLEDGE-INFORMED: KIDS ARE FAR OFF IN THE FUTURE

In the context of delivering real value to end-users, "actionable information" must move from **ART** (buzz word) to outcome. Future tools must be able to extract actionable, computable, contextual domain knowledge from informal sources of data (for example, text-based data). It is ***not an easy task*** in reality to augment user's ability to perform analyses using common models, dynamic composability of modular components, sensor-specific data, environmental data (big data is a bad word) and near real-time data from the edge. In order to capture "experience" even ontological frameworks may be useless. Even if domain experts can capture "experience" in a text-based format, it is doubtful if such text may "conform" to ontological frameworks. The latter is generally useful for text-based data (https://schema.org/).



This cartoon summarizes the journey from data to knowledge (didn't dare to include experience). In our phase one approach (PoC SENSEE 1.0), we focus on sensor characteristics. In phase two (not within the scope of this PoC) sensor data (bottom, left, silos) fuels data analytics and relationships between data (SENSEE 2.0 in cartoon of relations/associations between databases). Information becomes valuable to users if ART can proceed to synthesize and generate the data-informed decision as a service (DIDA'S) model (bottom cartoon, in pale green). As we approach knowledge integration phase, connectivity of local data and information with global system of systems adds value for data-informed policy decisions, understanding local dynamics and pricing in the context of market economics and trade practices. Creating GKG (global knowledge graph, and in future, labeled property graphs, LPG) is an evolutionary task, as we continue to stitch resources that can inform and provide knowledge to end-users. These "end-users" are no longer only farmers, growers, academic or manufacturers, it could include global organizations (FAO, WTO, UNDP) as well as policy forums and politicians (for example, farm bill, world customs organization, public health, other institutions, such as, FDA, CDC, WHO, ADB, EBRD, WTO).



KIDS is an open plan platform concept. Platforms are comprised of multiple applications and integrated solutions with embedded tools and databases that function as complete, seamless environments. Product innovation platforms are intended to support users collaborating across domains and businesses in the ***PEAS Platform for the Agro-Ecosystem***. These capabilities are increasingly needed throughout the entire extended enterprise in almost every vertical, agnostic of the type of application or function or users, including farmers, meat packers, produce growers, retail stores, customers, suppliers, and business partners. Developing open platform tools and technologies are not limited to any one domain because these modular tools can be applied, used and re-configured for re-use, almost anywhere, for example: error correction, graph and search engine algorithms, natural language processing (NLP), automated feature engineering, drag and drop tools, analytics, workflows, and services, such as KIDS, where "open" means 'plug & play' user friendly human-computer interactions and interoperability between system of systems. The monetization of KIDS is key for entrepreneurial innovators and investors seeking ROI. Users will ask how KIDS relevant to ***my*** farm or ***my*** manufacturing operation (think emergence of digital twins, eg, to mimic shop floor for machine tools) or ***my*** healthcare. Dynamic composition of the tools in the KIDS tool kit will be essential for the high degree of differentiation that must be achieved on-demand and served in real-time. To serve individual users or groups, raw data acquisition must be specific for the user's domain and the analytical tools (algorithms) must be contextually relevant. Cybersecurity will demand that user data is sufficiently protected. KIDS will offer different views and instances to different users. Only software can deliver granular services at a feasible transaction cost. For global adoption of ART, DIDA'S, KIDS, and PEAS, embracing the *pay a penny per unit* (PAPPU) pricing model, may be a potent and vital catalyst.

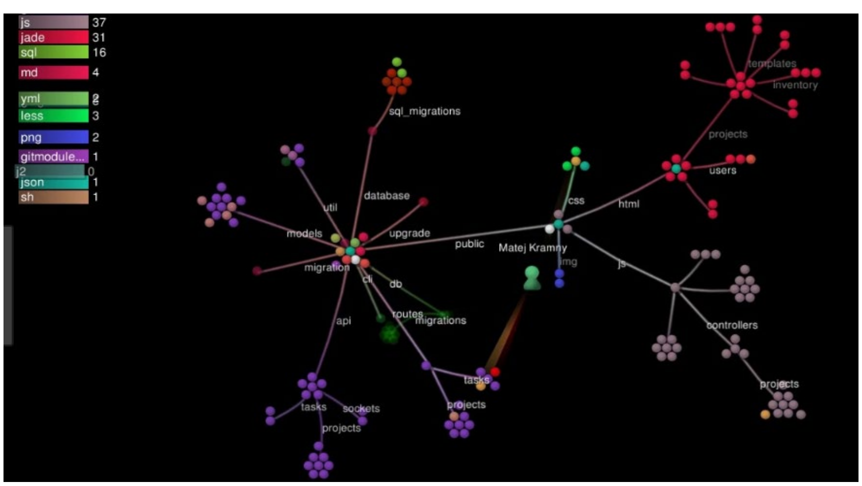
30 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

Human expertise is an **embedded** component of procedures, processes and decisions in any system. However, integration of human knowledge and human-computer interactions in operational analysis, are difficult to execute (why "smart" systems are still dumb). In certain systems, for example, the agro-ecosystem, human-generated knowledge is quintessential, yet it may be captured in text-based documents, which may be inconsistent and contain domain-specific vernacular. Even more challenging are the facts and observations that humans may grasp but *unable to articulate* or capture in writing or in notes. Thus, we have lost that intuitive factoid because of our inability to sufficiently capture what we are thinking.

The adage "hard to express in words" is a normal neurological state. It may be safe to conclude that there is *no intelligence in artificial intelligence* (https://arxiv.org/abs/1610.07862). AR (artificial reasoning) may be the best outcome from machine learning, no matter how "deep" one claims it to be. Hence, a name change from AI to AR is long overdue. The suggestion of **ART** in this document is more appropriate rather than perpetuating the lies and myth of AI.

Integrating ***knowledge and experience*** can better inform (DIDA'S) decisions, rather than relying only on ART. But, operational decisions at the point of use may find it difficult to synthesize the data (for example, streaming sensor data) with knowledge systems, in near real-time, to inform the user. The end-user in the field or farm or manufacturing shop floor, is more interested in the ***integrated information*** rather than data streams on a slick mobile dashboard.

Relationships between knowledge domains may boil down to ontologies. In other words, knowledge extraction must include design and development of taxonomies and metadata strategies for content management. The "fit" of these strategies to text-based data and other forms of unstructured sources of experiences remains to be explored with respect to existing (and/or evolving) vocabulary/taxonomy/knowledge organization system (KOS)/ontology software. The older thesaurus standards (ANSI/NISO Z39.19 or ISO 25964), newer ontology standards (OWL, RDF), and the SKOS (Simple Knowledge Organization System) model for "controlled vocabulary" may be relevant in this context (OWL/RDF is discussed later in this document). Most organizations may need more than one kind of controlled vocabulary. Hence, combined taxonomy, thesaurus, ontology and knowledge graph structures (discussed later) are emerging (for example, graph database-based Synaptica Graphite, cartoon shown below).



Most systems are starved of information, but we have an abundance of data, albeit uncurated data, often replete with noise and/or a poor ratio of signal to noise. Relationships are key to extracting experience and knowledge (previous cartoon) in order to inform and integrate data and analytics. Smartlogic Semaphore (https://github.com/ansible-semaphore/semaphore) provides a good view of these connections but in reality these are rarely "available" for rapid deployment (for example, in the agro-ecosystem). The cartoon is an example of what we ***think*** might be helpful for knowledge supported decision systems (KIDS). Other related software tools include PoolParty, TopBraid Enterprise Data Governance's Vocabulary Manager, Mondeca Intelligent Topic Manager and VocBench, to name a few specific suggestions, from an extensive list discussed in this book: http://www.hedden-information.com/accidental-taxonomist/.

Systems of the future must address this chasm between technical output (for which systems integrators want to charge money) versus the user value (the outcome for which the user is willing to pay). In any vertical, data plays a key role as a business driver. In the knowledge economy, the data analytics business will remain in the doldrums unless tools and technologies can deliver meaningful knowledge extraction mechanisms to support the context of applications.

Most corporations are eager to stop at **ART** rather than invest in DIDA'S and make sense of data in order to synthesize the ***knowledge support*** that end-users in the field *don't even know* that they are missing. This is a systemic problem, not limited to agro-ecosystem. The diabolical claims made by semantic web experts, machine learning and artificial intelligence marketing arms, makes one fearful to suggest that this problem of information extraction and knowledge-informed suggestions, needs, and may benefit from, ontological frameworks, semantics, ML tools and perhaps, artificial neural networks (ANN, CNN, RNN), at some later stage, in KIDS.

We need one or more networked platforms (mobile, edge/mist/fog/cloud, federated learning, distributed sub-domains, high fault tolerance, seamless interoperability between nodes, data distribution services, ontologies) where we "drag and drop" entities to combine, select, push-pull and dynamically hybridize, multiple tools, in order to create application-specific, domain expert-curated methodologies, for classifying, clustering and quantifying data, information and knowledge. Extraction from highly unstructured data calls for advanced Natural Language Processing algorithms (http://bit.ly/Interested-in-NLP) which must work with graph-theoretic methods (see figure 11 on page 28 of "SIGNALS" http://bit.ly/SIGNALS-SIGNALS) and ontology tools, for example, Synaptica Graphite, which offers directed-graph visualizer.

Optimization of value for the user is a continuous process of labelling and analyzing diverse sources of data, before the information perishes. If text-based documents are sources of knowledge and experience, then we must find new ways to harvest that contribution in our attempt to synthesize data (external data, crowd-sourced data), information and knowledge with experience, to aid the gradual transition from artificial reasoning tools (ART) to data-informed decision as a service (DIDA'S) to knowledge-informed decision as a service (KIDS). Continuous improvement will contribute to the domain-specific enrichment (ontologies due to pecan farmers vs tomato growers, agro-ecosystem vs manufacturing). Usability and functional preferences may lead to *de facto* standards and support "organic" growth of open-source toolkits to fine-tune knowledge extraction from distributed domains and improve the value of decision support, in the context of the journey from SENSEE to ART to DIDA'S to KIDS.

BEYOND KIDS ➔ EXTRACTION OF EXPERIENCE

Decision making is largely driven by human expertise. Automated decision-making works really well during sales presentations, using power-point. Human experts contain a wealth of tacit information that is intuitive, informally captured and explicitly under-utilized due to our inability to capture, catalog and re-use experience. For example, an experienced physician needs to see only the color of the sclera to “know” if the patient has contracted jaundice. Capturing the ontological framework of this knowledge and creating a computational equivalent of this type of “expertise” may be the Holy Grail for the future of advanced decision support. The forthcoming exodus of experienced physicians (due to Brexit?) will leave the NHS (British National Health Service) denuded of a repertoire of critical knowledge which may be irreplaceable. To the best of our knowledge, there are no mechanisms or tools in UK or anywhere else in the world, to capture such knowledge and re-use the experience to support new, or less experienced, employees or re-train other physicians who may have gaps of knowledge in the areas that were covered by the physicians who may be leaving UK.

Even if this tacit knowledge is implicitly or explicitly represented in text-based documents, these documents are not amenable to analysis using the existing tools of knowledge representation, for example, SKOS (Simple Knowledge Organization System) model for “controlled vocabulary”. These documents are likely to contain jargon, abbreviations, and domain-specific cryptic remarks, which may be difficult, if not impossible, to analyze with any “controlled vocabulary” commercial solutions. Text-based documents often yield important contextual ***pattern*** of information that is based on recurring experiences, which data alone cannot provide (for example, pathology report of blood cell count or streaming sensor data). ***Contextual information*** is useful when these patterns occur again to inform and guide decision support.

The research question is whether we can create domain-specific methods, guidelines, and toolkits to study and analyze formal and informal, text-based documents to extract patterns and other supporting information to aid future operational decisions? It will ***not*** be a one shoe fits all “AI” solution, rather a portfolio of domain-directed methodologies for transforming documents into a computable format, to augment our ability to integrate their value in future analyses. A group of experts may jump to standardize methods to capture this knowledge. The latter may be an acceptable and even useful approach, but just one part of a multi-part dynamic solution set which may consist of overlapping open source toolkits, and domain specific guidelines, to map unstructured patterns to symptoms, indicators and other detectable parameters (for example, smell of acetone in breath indicates excess of ketone bodies in blood, likely due to diabetes). Another group of experts may clamor to establish ***best practice*** guidelines for analyzing text.

The arm-chair academic and the business strategy version of this discussion will evolve with every reiteration but how closely will it reflect the complex needs of the practitioners? If we were to compile a "to-do" list for the future and compare it with actual examples of questions and problems from end-users in the field, today, we may begin to observe the gaps between the technologies we think we are combining, to answer questions, versus the non-linearity of the real issues. There are no easy solutions except for constant human involvement in decisions. Mining ***user experience*** does not fit the boundaries intrinsic in a scientific tool or pre-set vocabularies.

Experience is not structured to "fit" a tool or tools we may develop (as discussed above). User experience will evolve and *how it is recorded* will evolve, too. Hence the tools we thought could be useful for mining experiences from yesterday may prove to be impotent tomorrow. For example, thesaurus management software (also used for taxonomies), such as Synaptica KMS, and other products, no longer exist. If we improve the tools, combine the functionalities and aggressively pursue concurrent evolution of tools, then, it may provide a second-rate approach to harvest user experiences.

The next task is to ***synthesize*** extracted knowledge with technical information and data, to augment the user experience, in near real-time, at the point of use. It is a ***very difficult*** task.

Mind the Gap – *between tools and technologies versus real-world issues and problems*

A Sense of the Future: To-Do List for Tools and Technologies *(adapted from www.nist.gov)*

■ Develop open source ontologies (schema.org) and tools for curating, cleaning, labeling, feature selection and feature engineering for text-based logs, databases, data, and information.
■ Develop/standardize/innovate NLP techniques for descriptive data from int/ext ecosystems.
■ Integrate extracted data with analytics, workflows, feedback loops.
■ Create tools for knowledge access and visualization of components. Show real-time data (eg streaming sensor data) combining with past patterns to inform decision support / suggestions.
■ Develop metrics for verifying and validating methods and calibration at the granular level of sensors and actuator. This data should aid in diagnostics and prognostics of the system.
■ Data from real world test beds must be accessible by local and global partners (see Figure 16 on page 33 in "SIGNALS" http://bit.ly/SIGNALS-SIGNALS) for distributed learning tools.
■ Focus on ***value delivered*** to the consumer (value is dynamic in context, *never* in equilibrium).

A Sense of the Future: Problems and Issues from Fields and Farms *(source: expert end-users)*

● Average rainfall in a drought prone village decreased from 1000 mm to 500 mm, over the last decade. Can we reduce the annual volatility of water availability for irrigation and drinking, for humans and cattle? Technologies at hand include [1] Improving Ground Water Level (IGWL) [2] Energy-Water nexus technology developed by Datamatrix Infotech and [3] tools of IoT. Key problems to address: sensing and modeling. Do we sufficiently understand the science involved in the physical processes? Without the grasp of the basic tenets, the case-specific outcomes, even if they are successful for the problem at hand, may not be generalized or reproducible, elsewhere.

34 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

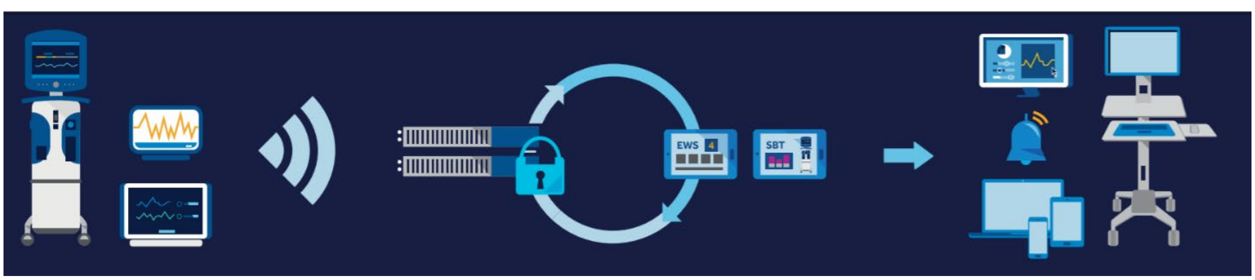
● Using an improved rice growing technique (http://tiny.cc/SAGUNA), 3000 farmers found that in addition to increase in yield, soil carbon has also increased, soil moisture is retained for a longer duration and number of earthworms have increased. Can we study and capture the interactions and dynamics between micro-climates, microbes, minerals, nutrients and signals from the soil?

● Flood forecasting and management model using 0.20m x 020m resolution LiDAR data to build 3D terrain model. Citizens and authorities may use this information to make decisions but micro-variations in rain intensity complicates forecasting. The ripple effect involves irrigation, water management, soil moisture and erosion – the sum of which affects crops and production of food. How do we converge metrics and measurements with knowledge and experience to focus on micro-environments in order to provide operational guidance to those who are in these zones?

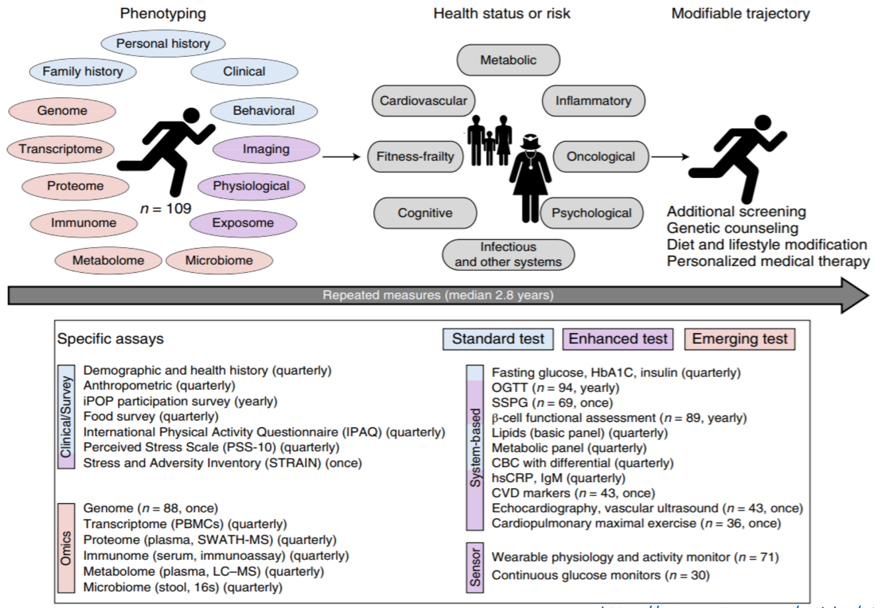
Sense of the Future: Meaningful Support**:** *Knowledge Graphs, Knowledge Supported Decisions*

Data fusion may not be information. Data-informed processes are more than **ART**. Evolution from information to knowledge may be a far more ***difficult*** process because the "reason" why the information may become knowledge must be represented. as a part of the general problem-solving logic. This concept from the 1950's is at the heart of AI and integrating "reasoning" remains an unsolved problem. AI was originally referred to as artificial reasoning and shares certain principles with cybernetics (http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf). Hence, AI to AR.

The current form of data analytics and logic tools may offer a layer of ART before a true data-informed decision making (DIDA'S) system may claim success. The educated customer may benefit from DIDA'S but the less informed clients may find ART adequate for specific cases of "low hanging fruits" which may require lower level of skills, available in ART. Our elusive quest for ***knowledge-supported*** decision making is aspirational. It may ***evolve from DIDA'S to knowledge-informed decision as a service (KIDS), if there is customer demand.***



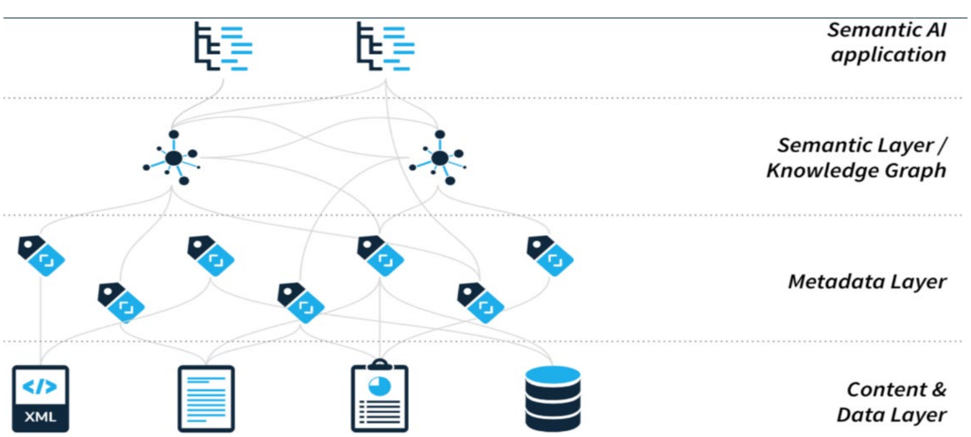
Application of sensor data and data from devices are not limited to any one vertical. It is a systems approach which transforms data to information for users at the edge. The cartoon (above) is an application of ART to KIDS in health/healthcare where real-time information can save lives and money. The ability to connect various troves of data are critical to the point of use, which indicates a future dominated by knowledge graphs to generate the knowledge. It is not unique to healthcare. It is applicable to agro-ecosystem, energy, finance. For example, Goldman Sachs is creating social graphs which integrates email (who emailed whom), telecommunication (who called whom), trading (who traded what) and linked financial data (who sent money to whom). It has 100 million edges and 2 billion nodes (http://bit.ly/GKG-KIDS).



https://www.nature.com/articles/s41591-019-0414-6

Fig. 1 | Study design and data collection. Overview of the in-depth longitudinal phenotyping used to determine health risk and status. Data types were categorized as: standard (blue), enhanced (purple) and emerging (red) tests. PBMCs, peripheral blood mononuclear cells; HbA1C, glycated hemoglobin; OGTT, oral glucose tolerance test; SSPG, steady-state plasma glucose; CBC, complete blood count; hsCRP, high sensitivity C-reactive protein; CVD,

KIDS in Health and Healthcare: From molecular profiling to behavioral changes, which may improve a healthy lifestyle, is a process which involves, for each individual, multi-year deep longitudinal studies, before actionable health discoveries may provide relevant data. If this data is sufficiently analyzed and synthesized, it may provide some information for precision health, of that one person (individual, patient). The challenge in scaling this process is the availability of qualified human resources to deconstruct and reconstruct the data in the process of extracting actionable information. Knowledge about the patient must be stored for use in the future. Barring the hype, this is a potential area for use of machine learning (ML) algorithms that can select the data from distributed databases and using knowledge networks (knowledge graph algorithms) find and weigh the relevant connections and correlations. By selecting weights as an index or metrics, ML algorithm engines may be trained to issue a set of recommendations and indicate the probability of confidence associated with each (suggestion, diagnosis, prognosis). This approach is universally applicable to any domain (data, knowledge graph, decision support) including non-human entities (digital twins). The implementation of knowledge-informed decision as a service (KIDS) starts with data. Bulk of the data from humans and machines originates from sensors.



KIDS WITH ABS ➔ CONNECTING IN CONTEXT: DOTS, DATA AND INFORMATION

Connectivity, in context, may be a basic instinct for all life forms. This generic statement appears less trivial if we consider that plants are designed to seek out sunshine and that ability, is in part, due to fractal patterns in the organization of leaves and the phototropic plant[1] hormone auxin, which induces photomorphogenesis. For animals, the quest for food[2] and flight from predators, are examples of connectivity, in context.

Almost all decisions, in humans, connects various data, information and knowledge stores, in our brain. The pathways remain poorly understood, despite an avalanche of foolish claims and blasphemous stupidity[3] of pompous statements from individuals dyed with hubris.

> "We want to create a brain in a box."
>
> IBM's Dharmendra Modha

An elementary form of bio-mimicry of decision systems may be at the heart of KIDS. We must be able to connect, in context, data from different domains, selected to suit the user's query, to synthesize information and knowledge, which will offer value to the user, if delivered, in time.

The connectivity we aspire to extract from global knowledge graphs (GKG) for use in ART, DIDA'S, KIDS, may find analogies from the annals of telecommunications. Networking is the bread and butter of connectivity for the telecom industry. Since GSM was introduced in 1991, the industry has pushed incessantly for increasing bandwidth, speed (data rate) and higher power for fairly expensive devices (iPhone). Circa 2015, telecoms were forced to accommodate, adapt and re-invent its practices, with the diffusion of IoT. Connectivity between vast number of devices, sending data pulses (sensor) or short bursts of data on-demand (user query) may survive on low bandwidth, low data rates, low power for IoT-type connectivity between devices, many of which are low cost devices. In other words, the opposite of the conventional wisdom espoused by the practitioners in the telecom industry of the 1990's.

IoT drove a fork in the telecom industry and non-traditional players invested in low power wide area networks (LPWAN). To counter LPWAN penetration as the key IoT backbone, frantic traditionalists created a partnership (3GPP) and agreed, in haste, on the NB-IoT standard, a mix of NB-LTE and cellular IoT (2016). Thus, emerged agile hybrid networks of traditional cellular, non-cellular, non-traditional mesh and other protocols that can take-over or hand-off any signal (eg WiFi, Bluetooth, WiMax), anywhere, anytime, from any device or any object.

[1] https://www.untamedscience.com/biology/plants/plant-growth-hormones/
[2] http://library.mit.edu/item/002405318
[3] https://www.foxbusiness.com/features/after-watson-ibm-looks-to-build-brain-in-a-box

The agility, with which traditional telecom players could surmount the barriers (due to dead weight of old technology) and embrace new tools, is a lesson for the decision sciences. For the latter, churning "data-driven" into "data-informed" still falls short of customer demand. The knowledge-informed decision is far more valuable. Transition from data-informed to knowledge-informed calls for incisive changes in synthesis of data and information, for KIDS. The caveat in taking this analogy from the telecom industry, too far, is the vastly convoluted pace of creating "standards" in the data and information domain. The operational failure of the semantic web and sluggish progress in creating ontologies are indicative of the challenges facing KIDS.

An example from the telecom industry which may resonate with proponents of KIDS is the case of a "connected" car or future of semi-autonomous or autonomous vehicles. The car needs instructions in real-time with near-zero latency. It must receive software upgrades, bug fixes, send reports from sensors, enable instructions to modulate actuators and maintain constant dialog with control centers using cloud, fog or mist computing. The user and the connected car ***doesn't care*** if data arbitrage is being conducted over a fixed connection, WiFi, WiMax, DSRC (dedicated short-range communications[4]), C-V2X (cellular vehicle to x), SDN (software defined networking[5]) or NFV (network functions virtualization[6]). The vehicle needs ***connecting***, with a certain quality of service (QoS). End-users may not care ***how*** the connectivity is implemented, as long as the ***network-agnostic networks*** can work seamlessly, in harmony, using whatever media is available (copper, fiber, wireless, LTE) to deliver the contracted quality of service, every time.

For users seeking assistance from KIDS, the user does not care which data domains the knowledge graph[7] must connect and whether it is a RDF graph or a labeled property graph (LPG). The ***quality of the outcome*** is the relevant determinant of value from ART, KIDS, for the end-user. The dynamic pricing index (service fee) for KIDS may be linked to the QoS metric-on-delivery. The domains of data, data analytics and information databases, are static "nodes" or resources from the perspective of the end-user. The query from the end-user is the ***trigger*** to instantiate an ***user-centric*** selection of the nodes. In other words, ***the query from the user will drive the connectivity*** between data swamps, analytics and information nodes, necessary to answer ***that specific question*** from the user. The analogy in the networking world is referred to as application driven networking (ADN) or application centric infrastructure (ACI) and are variations of the concept commonly referred to as service-oriented architecture (SOA). The ***service*** call ***shapes the events*** which will ***follow*** in order to respond to the specific request.

User-centric dynamic composability to create *ad hoc* knowledge graphs will benefit from integrating Agents[8] in the design of KIDS. One role of the Agent will be to parse the query and determine which nodes must be connected for KIDS to attempt to answer the query with a decent QoS metric. To provide a trivial analogy, imagine a busy intersection in Mumbai or Mombasa. The traffic lights aren't working due to brown out and motorists are confused by the DETOUR sign. A traffic policewoman is at the round-about, motorists are driving to the circle and policewoman directs the driver, depending on the driver's question. In KIDS, a software Agent, in the role of an "analyst" may execute the function of the policewoman, when it detects a query from an user.

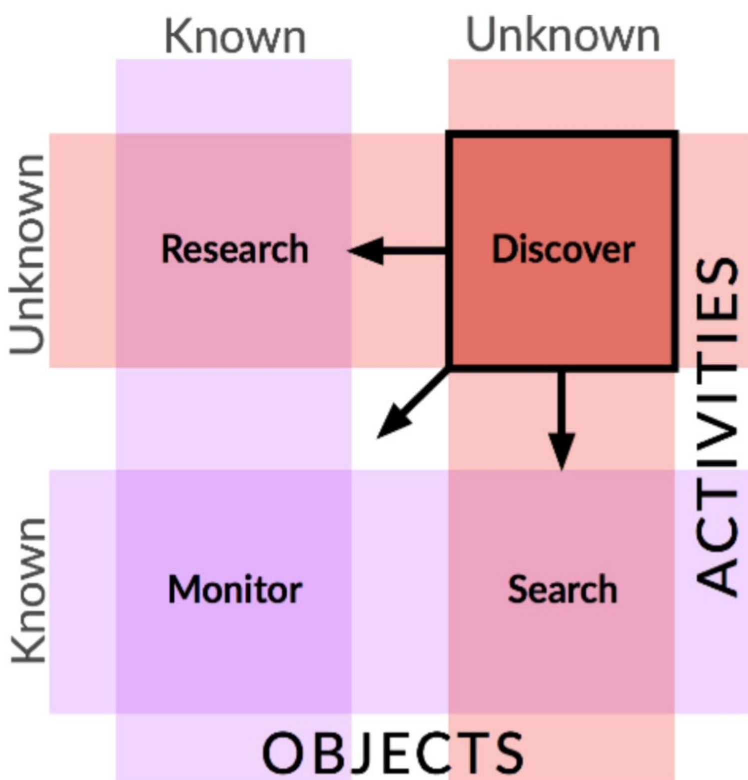
---

4 https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety
5 http://publications.uni.lu/bitstream/10993/20596/1/survey_on_SDN.pdf
6 http://www.ttcenter.ir/ArticleFiles/ENARTICLE/3431.pdf
7 https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html and https://youtu.be/mmQl6VGvX-c
8 http://ermolayev.com/psi-public/SOTA-TR-PSI-2-2004.pdf

What happens when the query falls in the "unknown unknown" category[9] shown in the cartoon? Although this scenario is more prevalent in the cybersecurity domain, the principle of the solution is similar to the discussion, at hand. The Agent must work as an *analyst* in handling the queries to "discover" the "unknown unknown" concepts in the user's query.

Discovery is a critical part of the knowledge graph future because the query-driven process must have a mechanism to find out ***what*** to connect, to create the knowledge graph. Data domains ***relevant*** and ***relative*** to the ***context*** of the query must be identified and ***connected***. The principle of **R2C2** (relevance, relation, context, connect) may be key to connect correct nodes of the knowledge graph. The graph, thus constructed, and the graph network which will ensue, must be capable of extracting the relationships, correlations or convergence, the queries are seeking. This graph (linked RDF triples, relationships) may be stored in a knowledge graph database and the abstraction may be recycled or the actual instance may be re-used.

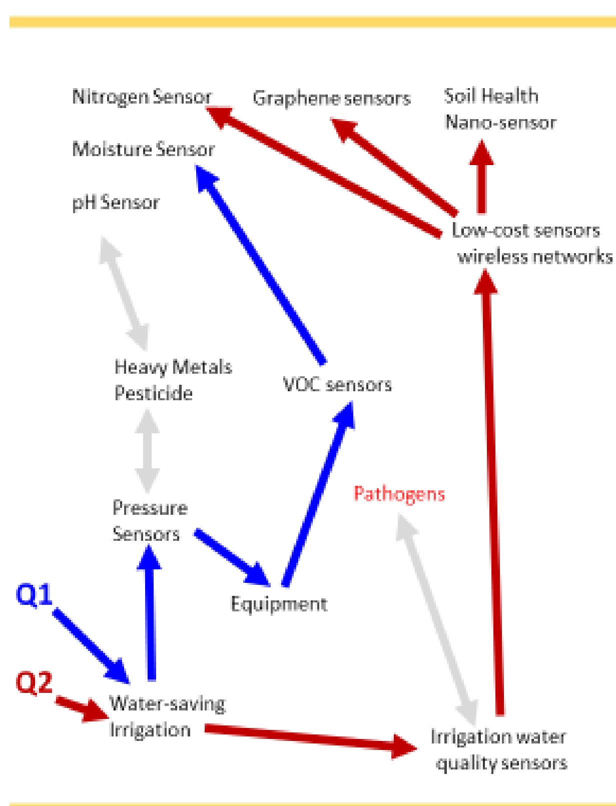
It will be remiss not to mention that one of the most egregious errors in the IoT hype is the idea that billions and trillions of devices and objects will be connected due to IoT. Even if there are trillions of things, the ability to connect is dependent on the ability of one object, with the correct tools to connect, to know and to ***discover***, that there is another object, within its reach, which is safe, compatible and configured to access, and connect. Although the central role of knowledge access[10] and ***discovery*** is an established principle, it is seldom emphasized for IoT.



[9] https://agile-defense.com/wp-content/uploads/DarkLight-Game-Changing-AI-for-Cyber-Security-Brochure.pdf
[10] http://oxygen.csail.mit.edu/KnowledgeAccess.html

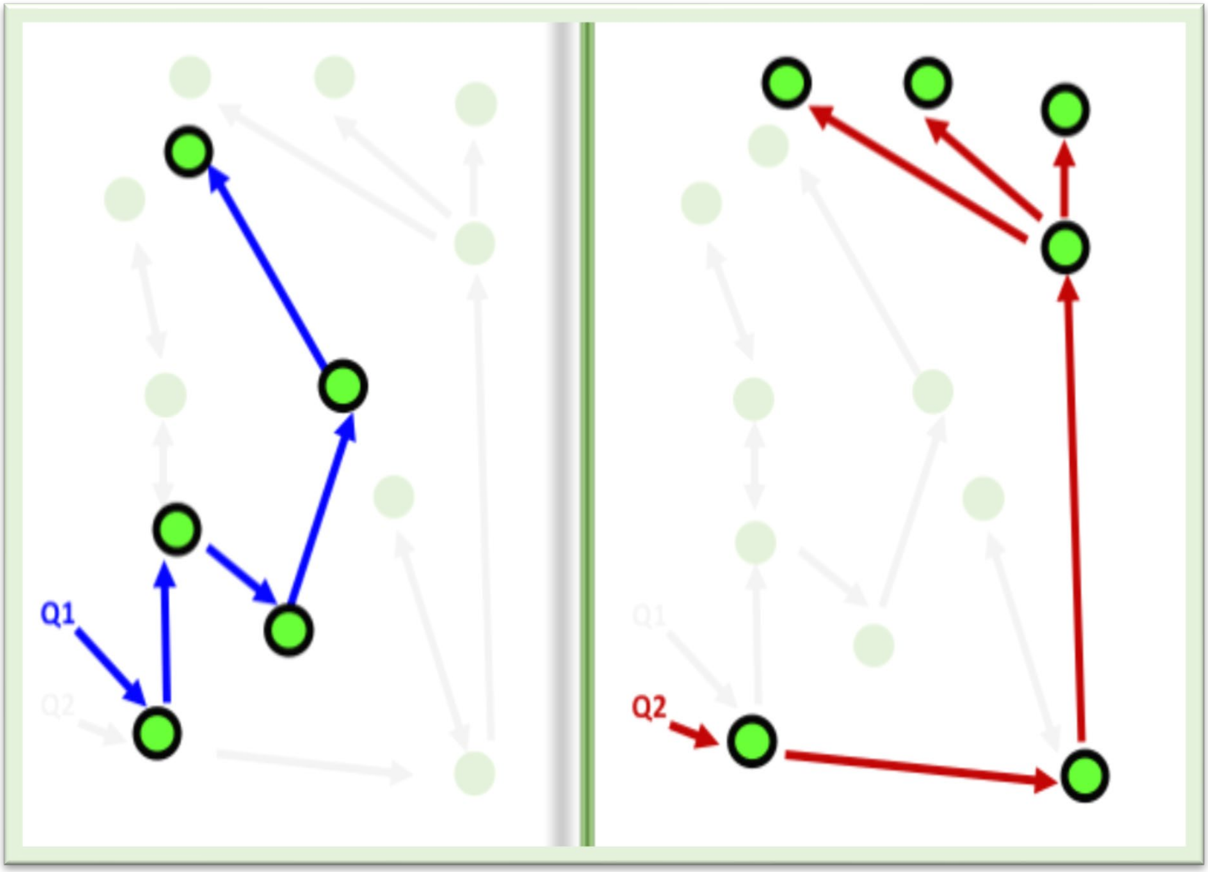
User-centric dynamic composability may also find some parallel concepts in network functions virtualization (NFV), a tool from the networking domain. The key idea of NFV is to ***replace*** dedicated network appliances, that is, ***hardware***, such as routers and firewalls, with ***software*** running on commercial off-the-shelf (COTS) servers. The aim of NFV is to transform the way communication service providers (CSP) architect networks and deliver network services. In the KIDS paradigm, the CSP equivalent may be domains of data, analytics and information stores or *service providers*. In the KIDS model, examples of *service providers* (in terms of data, analytics and information) may be weather data, provided as a publish/subscribe tool from the Weather Channel, prices of commodities (food) from Bloomberg Business, pollution data in terms of particulate matter (PM2.5) in the air from feeds maintained by the EPA. In NFV, network function software is *dynamically instantiated* in various locations in the network *as needed*, without requiring the installation of new equipment. The parallel for KIDS may be the Agent which catalyzes the *dynamic composability of domains, on demand* (triggered by user's query). This *ad hoc* composition is necessary to respond to the query, in the context of the query. The Agent does not install new components. The Agent simply selects the domains which may contain contextual and related resources, which are salient create the graph, to answer the query.

Development of embedded analysts, Agents-based ***selection*** (ABS) in the design of software architecture[11] is an old concept, which is a grand idea still hiding under a bushel. KIDS with embedded ABS may be necessary to navigate available resources and stitch the correct sequence of domains to synthesize knowledge-informed decision as a service, on-demand.



KIDS cartoon illustrates various domains of data and information available to a system. Two incoming queries, both on irrigation, are asking quite different questions. Q1 appears to be interested in saving water. Knowledge graph Q1 connects the nodes which provides information about soil moisture, which may optimize water distribution by the irrigation system. Q2 is also interested in saving water but not before understanding the quality of the water and condition of the soil. Graphs for Q2 connect different domains. KIDS with embedded ABS may be the tool (hypothetical suggestion) necessary to ***understand*** the query (syntax, semantics, ontological schema, unstructured vernacular) and then direct the path the knowledge graph must choose to respond to Q1 and Q2 with sufficiently high QoS metric (value). The latter will allow KIDS to charge users a small fee (exceeding a contractual QoS metric triggers the pay a penny per unit scheme).

[11] https://dl.acm.org/citation.cfm?id=122367



KIDS cartoon illustrates the connectivity between different domains by overlaying a knowledge graph on top of available resources (resource agnostic) in a system, for example, agro-ecosystem. The abstraction demonstrates very different (number of nodes, edges) knowledge graphs, due to queries from end-users (Q1 and Q2 are both related to irrigation water, in this fictional scenario). Dynamic composition of these *ad hoc* knowledge graphs are query-driven, user case specific. A distant analogy from the telecom domain is the creation of virtual private networks (VPN) by building a virtual network overlay on top of multiprotocol label switching (MPLS) network components. The "brains of the network" are managed[12] by software defined networking (SDN) controller platforms, which contains a collection of "pluggable" modules to perform different network tasks. For KIDS, an equivalent "brains" platform may contain modular ABS (Agent-based selection) analysts, to direct the formation of knowledge graphs, by extracting ***connectivity*** structures (paths or graphs between entities, see colored circles in the cartoon) relevant to the ***context*** of the query. By training these tools (eg graph neural networks[13]) to parse the questions, various clusters of ***meta structures*** may be created to facilitate ***knowledge discovery*** tasks to locate[14] ***where*** is the data or information, related and relevant to the query. These algorithms[15] may add value to the embedded ABS analysts and in turn enhance the performance of KIDS.

[12] https://www.sdxcentral.com/networking/sdn/definitions/sdn-controllers/
[13] https://repository.hkbu.edu.hk/cgi/viewcontent.cgi?article=1000&context=vprd_ja
[14] https://blog.cdemi.io/beginners-guide-to-understanding-bgp/
[15] https://iswc2017.semanticweb.org/wp-content/uploads/papers/MainProceedings/272.pdf

Extracting useful knowledge using the graph theoretic approach must be anchored to deliver ***contextual meaning*** of data. Hence, to connect nodes using the R2C2 principle, one must take into consideration the semantic profile of what is being connected. The data intensity of system of systems may be comparable to data intensive science projects, for example, the LHC (Large Hadron Collider) and ASKAP (Australian Square Kilometre Array Pathfinder), which generate petabytes of data, each year. If such vast volumes of data are stored in "data swamps" then we may lose its value unless curating contextual data from data swamps meets a miracle.

To make data relevant and meaningful for end-users, the applications must [1] select and coordinate data and information, [2] provide synergistic data integration between data domains (data from specific sensors or equipment, crowd sourced data), [3] enable visualization (plots, suggestions, recommendations) and/or [4] take action (actuate, execute). All of this is expected to happen in a seamless manner, in near real-time, without the need for users to understand any of the underlying representations and structure of the data.

In this vein, semantic[16] tools provide categorization capabilities and may facilitate machine-encoded definitions of vocabularies (which could be different based on vernacular), concepts and terms. In addition, semantics may explain the interrelationships among them (different vocabularies residing in different documents or repositories). The challenge is (and may always will be) balancing expressivity (of semantic representation) with the complexity of defining terms (used by experts, scientists, engineers) and implementing an end-user-friendly resulting system. This balance is *application-dependent*, for example, in terms of ease of use between tomato growers and nurses. The degree to which the implementation must be user-friendly depends on the intrinsic technical competency of the users. A single solution may not fit all, even *within* groups, for example, pediatric nurse practitioners vs geriatric care nurses.

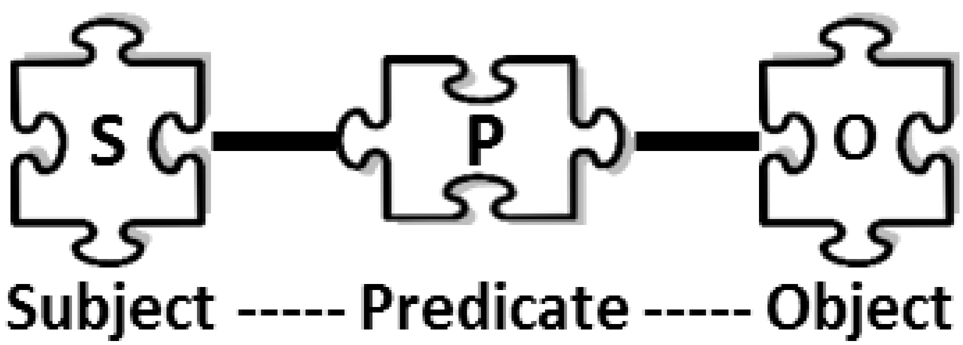
The success of semantics in this respect will be governed by a very different form of human relationship. **Leadership** in this collaborative approach will determine how the fields may progress in the future (for example, agro-ecosystem vs automotive vs health vs finance). Success of semantic structures necessary for data driven software processes will depend on peer relationships, where [a] domain experts or scientists in specific fields will form co-dependent liaisons with [b] computer scientists, as well as software architects/engineers and [c] data providers, data system administrators and so-called data scientists. Fields which are traditionally "farther" away from computer science and software engineering, for example, agriculture, chemistry, economics, must strive harder to bridge this chasm by co-locating computer science departments with agriculture, chemistry and economics, perhaps in the same building or quadrangle or emulate instances[17] where agriculture is a part of a media laboratory. Without global and cultural cross-fertilization, ontological schemas and semantic catalogs of the future may be anemic, half-baked, sloppy and second grade (yet, masquerading as good enough).

Semantic tools may be a part of data-driven, evidence-driven, reasoning solutions, logic tools (ART). Statistical and mathematical modeling-based ML may be part of logic tools, too. These are different but complementary. The distinctions may get fuzzy when we combine high volume data, analytics, distributed information and several knowledge domains, as in KIDS.

[16] https://pdfs.semanticscholar.org/9256/c883b1ecaea08abc46179e2927302523a66d.pdf
[17] https://www.media.mit.edu/groups/open-agriculture-openag/overview/

- Deductive reasoning, syllogism & categorisation
  (Aristotele, 384 BC – 322 BC)
- Formal logic & calculus rationicator (reasoning, symbol
  (G.W. Leibniz 1646 - 1716)
- „Begriffschrift“, technically: predicate logic
  (Gottlob Frege, 1848 – 1925)
- Frames for representing stereotyped situations
  (Marvin Minsky, 1974)
- Rules & expert systems
- Ontologies
  (Leibniz, Kant, Gruber 1994)
- Description Logics
  (Baader & Hollunder, 1991 et al.)
- Semantic Web
  (Berners-Lee, Hendler, Lassila, 2001)
  & Linked Data
  & Knowledge Graphs

The promise of semantic tools and technology may be rooted in its ability to capture the semantics of the data with the data itself. It can capture meta-description of different kind of objects, attributes, associations, and activity into a conceptual model, which can be populated with instances of actual data. Described using OWL/RDF[18] *syntax*, the conceptual model "ontology" represents the data itself in a single, consistent manner that is independent of how it is physically stored. With exceptions, ontologies formally describe taxonomies and classification networks, defining the structure of knowledge for various domains: nouns representing classes of objects and verbs representing ***relationships*** between the objects. Ontologies can represent information coming from heterogeneous data sources, hence, it can deal with structured, semi-structured, and unstructured data. The latter is particularly valuable for diverse end user groups.



When data is mapped against an OWL/RDF ontology, instances of the data are expressed based upon the idea of making statements about resources in the form of **subject–predicate–object** expressions. These expressions, also referred to as S-V-O (subject, verb, and object) are known as *triples* in RDF terminology. The 'Subject' denotes the object, and the predicate (verb) denotes a single semantic trait or aspect of the object that can be a literal value or expressed as a relationship between the subject and another object that is the target of the relationship.

[18] https://www.w3.org/OWL/

For example, "soil pH 8" in RDF triple is **subject** denoting “soil” and **predicate** denoting “pH” and an **object** denoting “8” which is the OWL/RDF take on using the object as the subject from the entity–attribute–value model within object-oriented design: entity (soil), attribute (pH) and value (8). The object (soil) can have another attribute (contains) that points to another object (phosphate). The object (phosphate) might have an attribute (produces) another object (acidity). Yet again, the object (soil) might have an attribute (contains) another object (microbes).

This is why RDF triples, despite their shortcomings, enables the formation to link a series of relationships between two or more objects. A graph, in this context, is a linked set of RDF triples. OWL/RDF-based data model may suit certain kinds of knowledge representation better than relational models because it can fuse data from multiple relationship tables about the same object. It is the foundation on which directed graphs are built. A collection of RDF statements intrinsically represents a directed multigraph.

A **knowledge graph** (mentioned often in this document) is a knowledge base that is made machine readable with the help of logically consistent, linked graphs that, taken together, constitute an interrelated group of facts. RDF triple represents human knowledge in standard, machine readable form by linking a subject, predicate/verb and object. RDF representation can be visually displayed as the nodes (subjects and objects) and edges (verbs/predicates) of graphs.

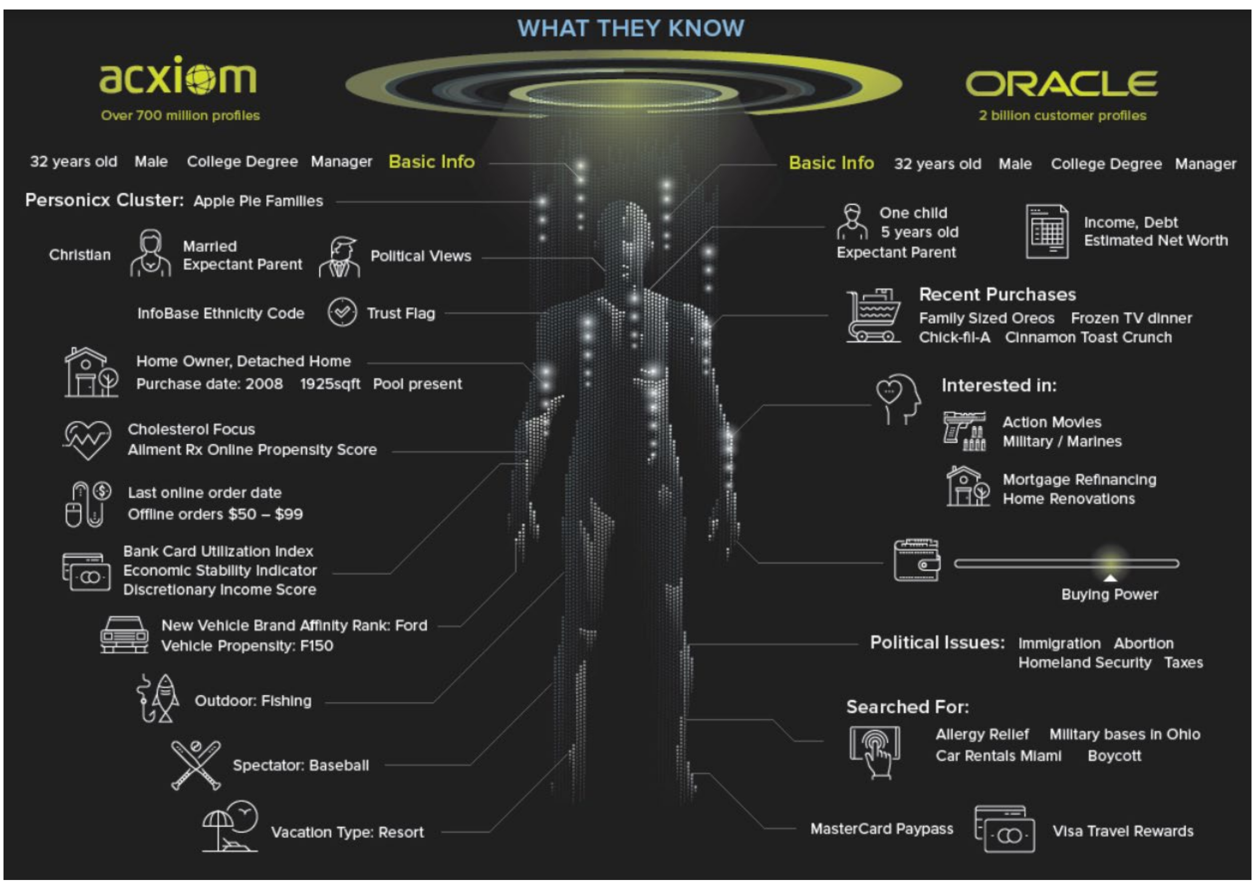
The idea of ***artificial reasoning in logic tools*** relies on machine readable statements of facts. The expectation is that “triples” linked in a logically consistent way, via knowledge graph, will possess ***reasoning*** ability. When logically consistent factual triples are added to the graph, ***machines can infer new links or connections***. Hence, new connections can be discovered by humans due to the ***reasoning power of machines*** (**artificial reasoning tools**). Machines can then gain ***access to the relevant data*** in the ***context*** of these linked triples (knowledge graphs) as part of an ***information service*** (discussed earlier in the section KIDS WITH ABS) provided by **ART.**

For example, **ART** (artificial reasoning tools) may uncover the ***relationship*** between water, pH and Pb. The ***machine reasons*** that if pH of water in distribution pipes is less than pH 7 then the probability increases for metals, such as, lead (Pb), to leach out of the material (alloy) of the pipes (acid leaching) and increase the concentration of Pb ions in domestic drinking water supply (Pb is a neurotoxin). Having uncovered the relationship, the function of the knowledge graph (in ART) is to contribute to a solution, preferably quantitative. ART must ***discover*** and ***locate data*** to synthesize the solution and display the outcome on the end-user’s mobile device. ART must discover data for each parameter, analyze and aggregate to create logical fusion. This demands data interoperability and choice of open APIs between systems, for example, the water distribution ***map*** (GIS), water quality ***in*** the distribution system (county public works database), chemistry knowledge for rate of leaching of Pb vs water pH (extract and merge the standard data with the actual pH of water, in this case). The ***predictive analytics tool*** may wish to forecast the (cumulative) increase in the concentration of Pb, with each passing day of inactivity. ART must display the useful version of this outcome and recommend mitigation strategy to reduce the morbidity of neurotoxicity due to Pb ions leaching from pipes into drinking water. This is the expectation from combined SENSEE 1.0 and 2.0 project, in terms of solutions for real world problems. We start with ***logic tools*** to deliver **ART** rather than “boil the ocean” with ML and other tools, which takes years, to reach the “knowledge-informed” quality of service (QoS).

44 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

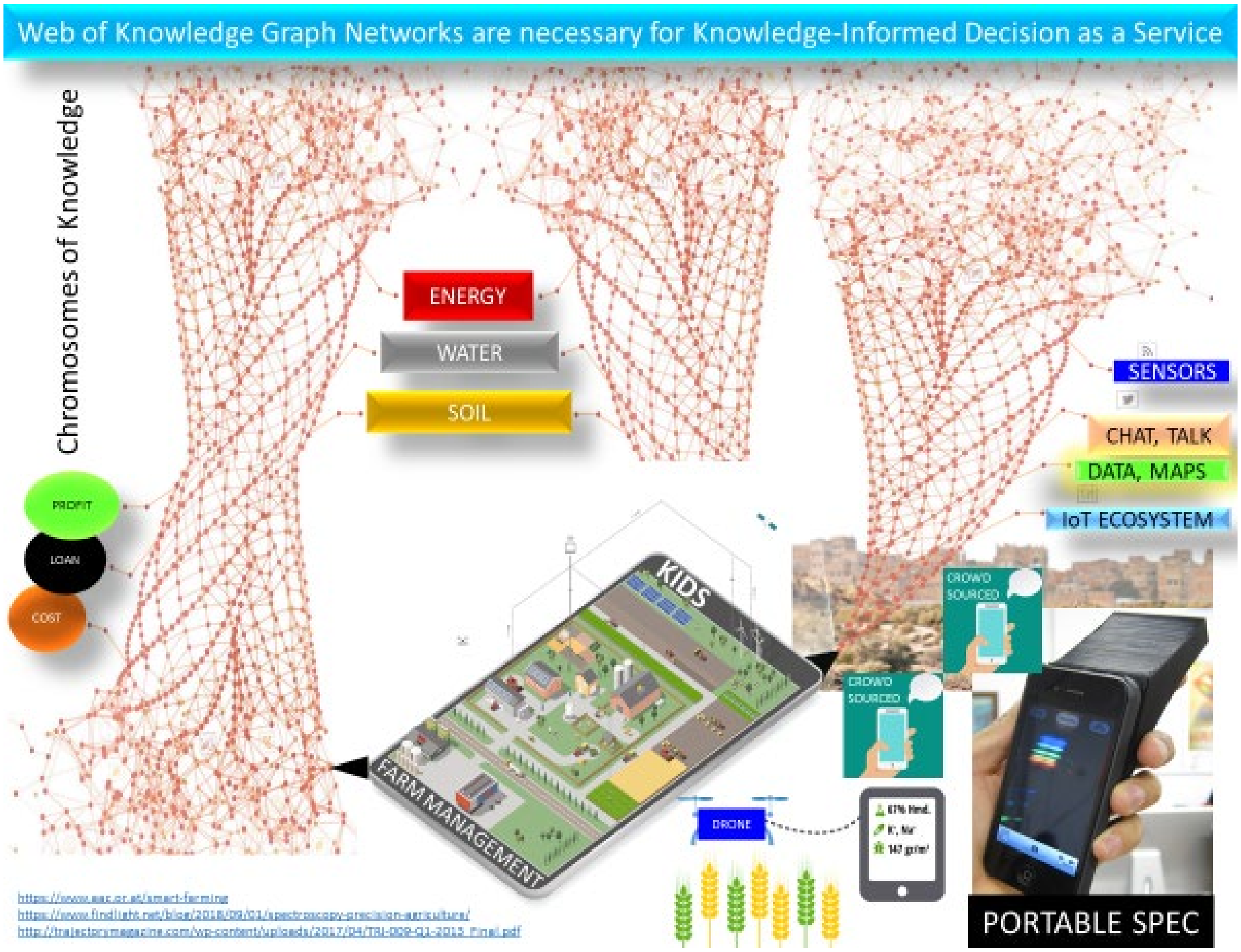
Neither OWL/RDF standards nor graph networks or knowledge graph databases, are a panacea. They may not represent everything and all advantages are temporary. Application of graph theory will not obliterate the role of other architectures and databases. The balance of tools vs interoperability between systems, are central to "understanding and forging relationships" between relevant systems, through contextual combination of tools and confluence of ideas.

Knowledge combination/integration beyond (heterogenous) rules and ontologies are not only difficult[19] but calls for ***new thinking***. ***The semantics of knowledge bases other than rules*** (for example, descriptions of temporal processes like workflows in ART which could logically decide using logic tools when the irrigation system must turn on/off water pumps, or protocols in spatio-temporal logic) ***must be integrated***. We need a higher plane of logic framework in which knowledge modules, with different native semantics, can be overlaid with meaningful semantics, preferably agnostic of linguistic bias, ideally as a "plug and play" operation, graph-friendly "drag and drop" operation for non-expert end-users, who may wish to decompose and/or re-compose the choice of logic and logic tools, based on experience or input from other expert humans in the loop. Chaperoning convergence between distributed knowledge domain(s), operational rules, data, information, and systems science, is a daunting and challenging goal (see cartoon below).



https://www.visualcapitalist.com/personal-data-ecosystem/

[19] http://www.kr.tuwien.ac.at/staff/tkren/pub/2008/rowschool2008.pdf



CAN KIDS UNDERSTAND THE QUESTIONS FROM REAL-WORLD END-USERS?

The list of sensor description related questions that the PoC attempted to answer in Step I was sourced from experts and the queries (list of questions in this document) also originated from experts. Descriptions of sensors from experts or descriptions extracted from web document searches (for example doi:10.4172/2329-6798.1000111) may be in sharp contrast to queries from real-world applications where questions are from users in agro-ecosystem, retail or healthcare.

Unstructured questions from users must be sufficiently understood by KIDS, if we are aiming to provide value for real-world applications where the user may be paying a fee for the service. To make SENSEE useful to end-users, to a limited extent, we have to start with the questions from end users (***PEAS Platform for the Agro-Ecosystem***). The end- user may want to know what types of sensors are available to detect mercury, who are the manufacturers, which brands are highly rated, what is the price, what is the maintenance fee and software licensing cost. These questions may not be answered by SENSEE 1.0 but by **ART**, in future. The query-triggered search must be able to understand the domains that the search engine must connect in order to extract the data and information relevant to the question. The latter is beyond the scope of SENSEE 1.0 and 2.0 but expected to be a building block for artificial reasoning tools (**ART**).

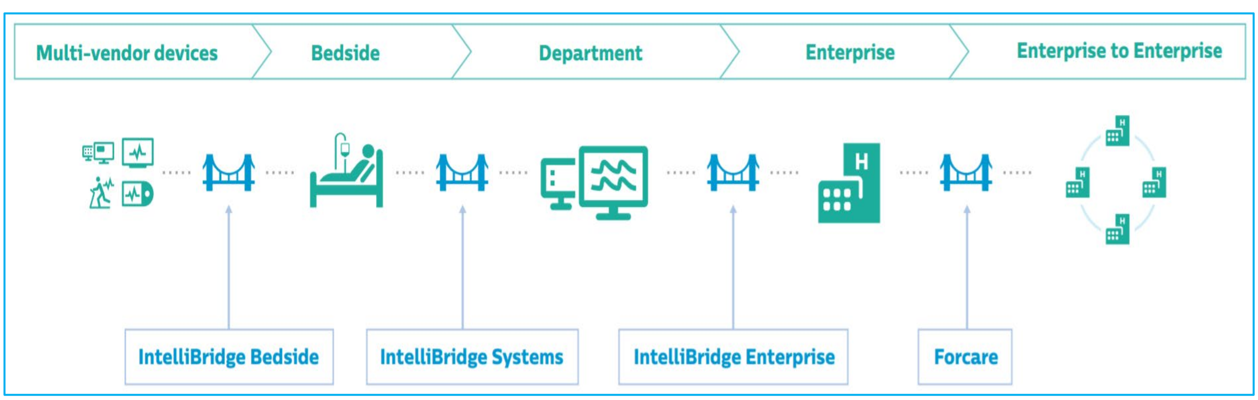
In many instances, user may ask questions about systems and technologies which may not involve sensors. For example, end-users on a farm may have a question about the amount of moisture in the soil vs the volume of water that must be dispensed by the irrigation pump system. Multiple domains must be integrated to address the diverse range of questions expected from end-users in any ecosystem.

The ability of ART, DIDA'S, KIDS, to understand the question and ***relationships in the question*** are critical to the success of PEAS platform. ABS analyst may be critical to evaluate and understand which direction to pursue and which domains to connect, based on the question.

When we move from SENSEE 1.0 (sensor descriptions in sensor search engine) to sensor data (SENSEE 2.0), the source of the data will be ***sensors in use by end-users*** (sensors deployed in farms, stores, shop floors, transportation). The end-users will have to agree to upload sensor data streams to the ***open PEAS platform***. The incentive for the user is the expectation that ART (KIDS) will make sense of the data and provide end-users with actionable information. In future, perhaps, offer knowledge, to improve decision systems or aid human users in decision making.

Participation of manufacturers, who are possessive about data and dissemination from their sensors and equipment, are potential sources of conflict. It is well nigh impossible for any one manufacturer to provide the range of sensors and equipment necessary for all operations. The manufacturer-specific dashboard may always remain a data portal, short on information and devoid of knowledge. Users, however, can change the *status quo*. User-adoption of ART (KIDS) will depend on the critical mass of data and information connectivity, as well as the ability to understand questions from users and answer them with a very high QoS (quality of service).

Aggregation platforms in the agro-ecosystem may share some analogies with the lack of device data interoperability in healthcare (https://mdpnp.mgh.harvard.edu/projects/ice-standard/). Deaths due to lack of interoperability is calling for change (https://www.himss.org/file/1325897) in the healthcare system to aggregate data in the context of the patient and transform the data to information, relevant to the patient and the point of care medical professional, as well as the extended enterprise. In other words, the cartoon below may represent **KIDS in healthcare.**



Aggregation of tools on a platform is an old (https://dspace.mit.edu/handle/1721.1/56251) idea which may find its origins in the "bazaars" of ancient Mohenjo-Darro and Mesopotamia, the "clusters" in town centers and the modern "malls" which are almost universal. Radio, TV and movie halls aggregated music, shows and movies. Digital aggregation pioneered by Amazon, eBay and Napster is evident in the streaming platform ROKU (https://blog.roku.com/oxygen). In healthcare (ICE, clinical environment, www.mdpnp.org/MD_PnP_Program___OpenICE.html) or in the agro-ecosystem, or most other system of systems, data aggregation offers value. KIDS is in good company and not an enigma for end-users, if they have the patience to start with ART.

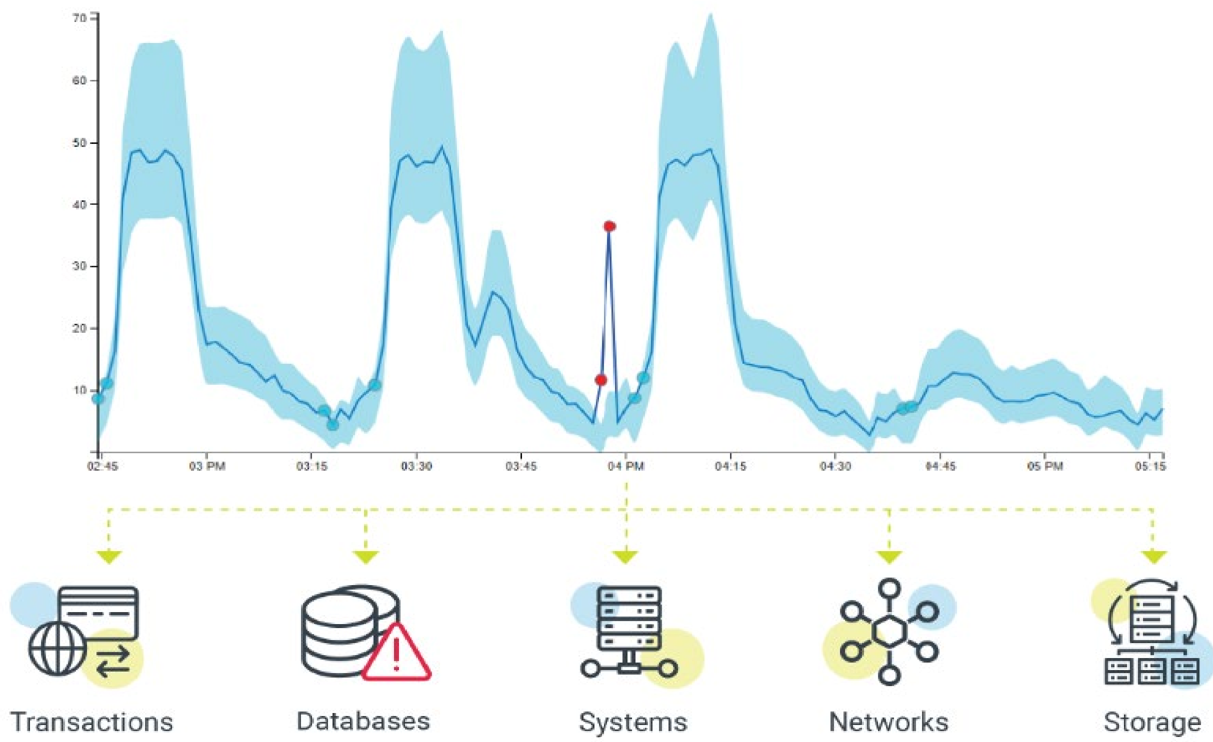
KIDS can catalyze data fusion by aggregating contextually relevant data from different systems, and provide analytical support, for decision making, in near real-time, on a mobile app through a smartphone or tablet, anytime, anywhere. Bringing the ***algorithms to the data*** at the edge (point of use) is possible by running computation at the edge (http://eyeriss.mit.edu/) and using sparse, trainable artificial neural networks (https://arxiv.org/abs/1803.03635) to help humans make better decisions, at the edge, before the value of the information perishes. By partially automating the system, the actionable information can also actuate sensors or systems (paradigm shift from SARS to SARA – see SARS♠AG here http://bit.ly/SIGNALS-SIGNALS).

ART is expected to deliver low-risk automation to solve specific problems, for example, based on the situation and feedback from the outcome (control theory feedback optimization loop), turn on/off irrigation water pumps, selectively, by distribution zones, using a GIS map.

The value of ART (logic tools and ART are not unique approaches) and the monetization potential from knowledge-informed analytics, is linked to performance of KIDS. Imagine how agent-based artificial reasoning (ABAR) bots, may continuously seek non-obvious exceptions, non-obvious correlations and non-obvious errors. Creating machine learning algorithms and deep learning tools only to search for anomalies (red dots in the cartoon, next page, example of root cause analysis) is an under-utilization of the benefits from tools of artificial reasoning (AR).

ABAR may be trained to find positive, as well as negative, correlations. Training is still in an enigmatic *black-box* domain. With greater clarity, perhaps, training can harvest crowd-sourced nuggets of knowledge. For example, an apocryphal anecdote from an ALCOA plant describes the breakdown of a chemical processing step, just days after the retirement of an experienced plant operator. When the operator was invited back to help identify the problem, it turned out that the operator used to spit in the smelted ore chamber. After his retirement, no one was spitting during that processing step. The surfactant from the spit was key. Surfactants catalyze chemical purification processes for aluminium. Hence, the value from crowd-sourced information, knowledge, experience and wisdom.

Creativity and innovation will be necessary to capture these occasional unstructured events. The next task is to integrate them with ongoing ML/DL *training tools* "educating" ABAR. How can crowd sourced wisdom train ABAR? For mass adoption, the process must have an ETL type tool (mobile capture and upload, drag and drop) for non-experts (plant managers, transportation planners, *any* end-user).

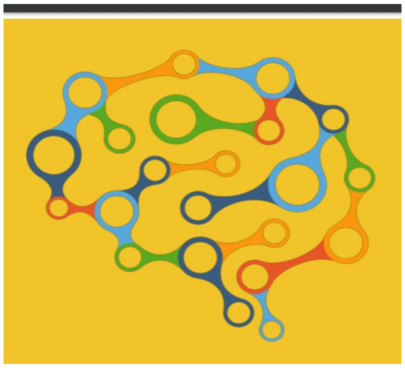


But, how will we know that a nugget of wisdom just hatched during a conversation? That is a very difficult problem. If the "nugget" is captured, how can we add it to training algorithms? One poor analogy is cooking. First, fry onions and garlic, then add spices, followed by the meat, add more aromatic spices. Another analogy is a human attached to an IV drip (intra-venous), which can drip saline, blood, morphine, antibiotics, cyanide. The key is the flexibility and modularity of adding things to a process. Do we need a "funnel" or app or API for delivery?

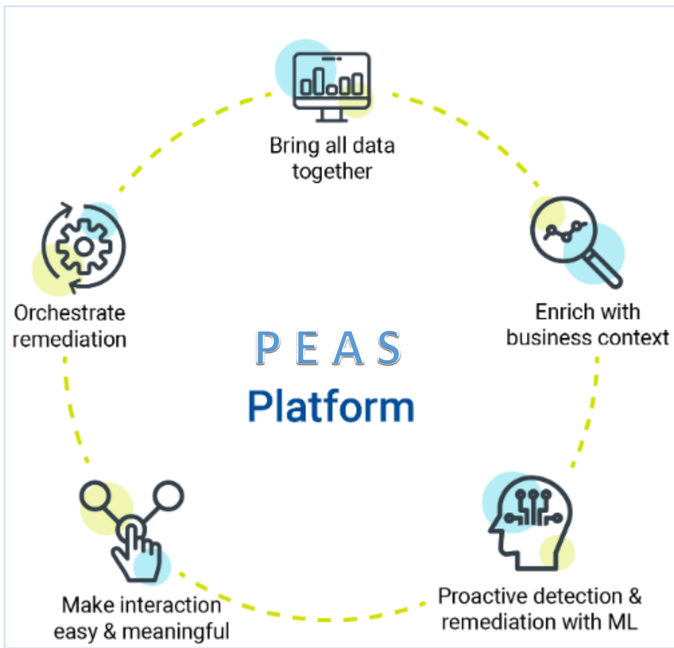
The outcome of training ABAR is to create an army of AKBAR (**a**gents with **k**nowledge-**b**ased **a**rtificial **r**easoning). In "IoT is a Metaphor" the attempt is to transform data/information to knowledge, and find new ways how organizations and enterprises may create value for users, through knowledge-informed decision as a service. KIDS with agent-based selection (ABS) will evolve to include AKBAR (agents with knowledge-based artificial reasoning), where mobile agents can travel *between* networks and cross-pollinate domains with information. Spread of mis-information, hence, raises its ugly head and cybersecurity considerations become central. However, it is still tempting to speculate, as mentioned elsewhere in this document, how we may "air-drop experience" from those who have it, to those who may wish to use it, for a fee.

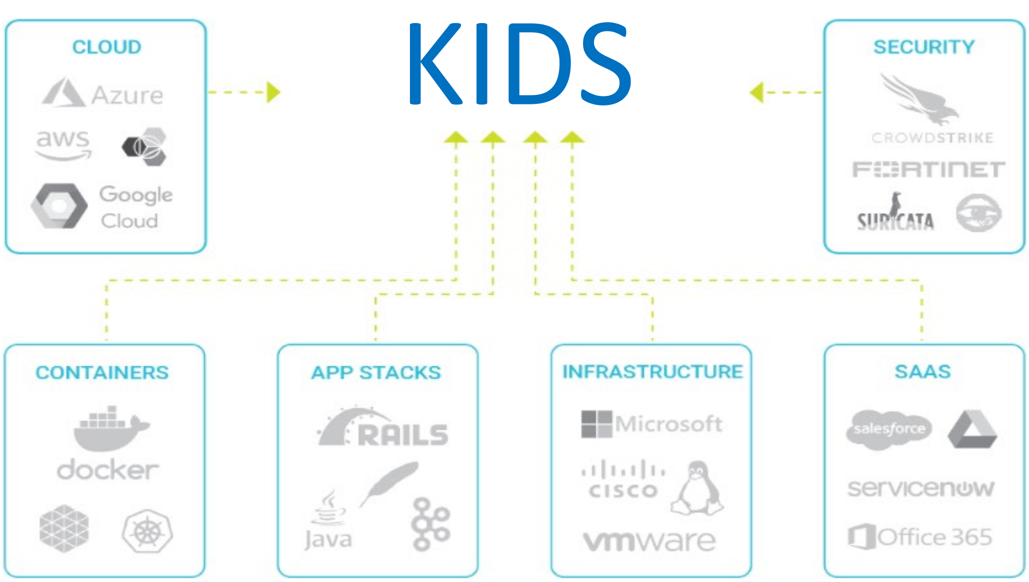
The future demands we ask different questions, new questions, relevant and contextual questions, obvious and non-obvious questions, incisive and analytical questions. Hopefully, at least some of these next generation questions will also contain a few *correct questions*, to spur new thinking, create tools that are still cryptic among the unknown unknowns, and help us to visualize the possibilities, with *new eyes*.

49 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS



The cartoon on the left and everything else discussed here is non-linear. Optimization routines, linear programming and static databases are rigid and less useful. All the rage about Hadoop is almost dead (HDFS, without transactions, search, indexing or caching, failed to solve real data problems) even though it was a NoSQL distributed data technology. Lack of semantics (context) turned Data Lakes into Data Swamps. To get out of the "swamp" we need NLP (not LP) and new eyes. We wish to use knowledge graphs (KG) and KG algorithms as a new path in the journey from data to knowledge (KIDS).





Towards building the next generation database query engine

Yesterday's DB engines are incapable of solving today's problems. David Mack.

In the 1970's the relational database was born (www.seas.upenn.edu/~zives/03f/cis550/codd.pdf) and remains a staple in the industry. But, enterprise companies are beginning to explore machine learning, because databases are inadequate for the company's informational needs. Relational databases have been wildly successful, forming an essential piece of almost any application. With this success has brought a rich deluge of data into database systems. Relational databases are great at supporting the developer defined symbolic relationships in the data (e.g. purchase belongs to user), but have barely any support for the noisy, sparse, probabilistic relationships that arise within the data itself (e.g. users with higher disposable income tend to make more purchases). This limitation is reflected in query languages (e.g. SQL) themselves. They are famously unfriendly for non-technical business users, so much so that entire teams of data analysts, BI experts and data scientists are drafted to help non-technical employees access their data. A very simple query such as "get the second highest salary" translates into:

SELECT DISTINCT Salary FROM Employee e1 WHERE 2=Select COUNT(DISTINCT Salary) FROM Employee e2 WHERE e1.salary<=e2.salary

A new generation of database query engine is taking a different approach. At its core, it's very different from current database query engines:

It accepts natural language (e.g. English) instead of SQL for queries

It represents data as a mixture of sparse features instead of as items from fixed categories

In the real world, nothing fits into neat boxes. Words have many meanings. Sentences can be ambiguous. Concepts and thoughts are related to others, in many different nuanced ways. Fall-leaves, tobacco and leather seem to go together, but why exactly? Our data representation supports and embraces this deep interconnectedness. We achieve this by representing data as mixtures of sparse features (i.e. many dimensional vectors). These representations are created using learned embeddings and learned transformation functions. This allows the query engine to better use the nuance of the query's words to find relevant data. It allows it to aggregate and filter data based on learned sub-categories, of which membership is not binary.

It learns multi-step deep algorithms from examples

Many times in life, we can specify the inputs and outputs, however working out how to get between them is hard (for example, try writing a series of rules to tell if a photo is of a hotdog). ML can work out the middle part in the right circumstances. Classical algorithms, the ones readily implemented in traditional database query engines, are very rigid. Each step must be a clear-cut decision with easily specified inputs. In a learned algorithm, each step can incorporate many weak signals to work out what to do next. Furthermore, it can do many different sub-steps in parallel, weaving a much more complex solution than could be written by an engineer. This is like comparing how many people cook in the kitchen vs a recipe book: we measure ingredients by eye, combine ingredients by feel and cook it until it smells and looks good. We improvise. None of which is captured in a recipe.

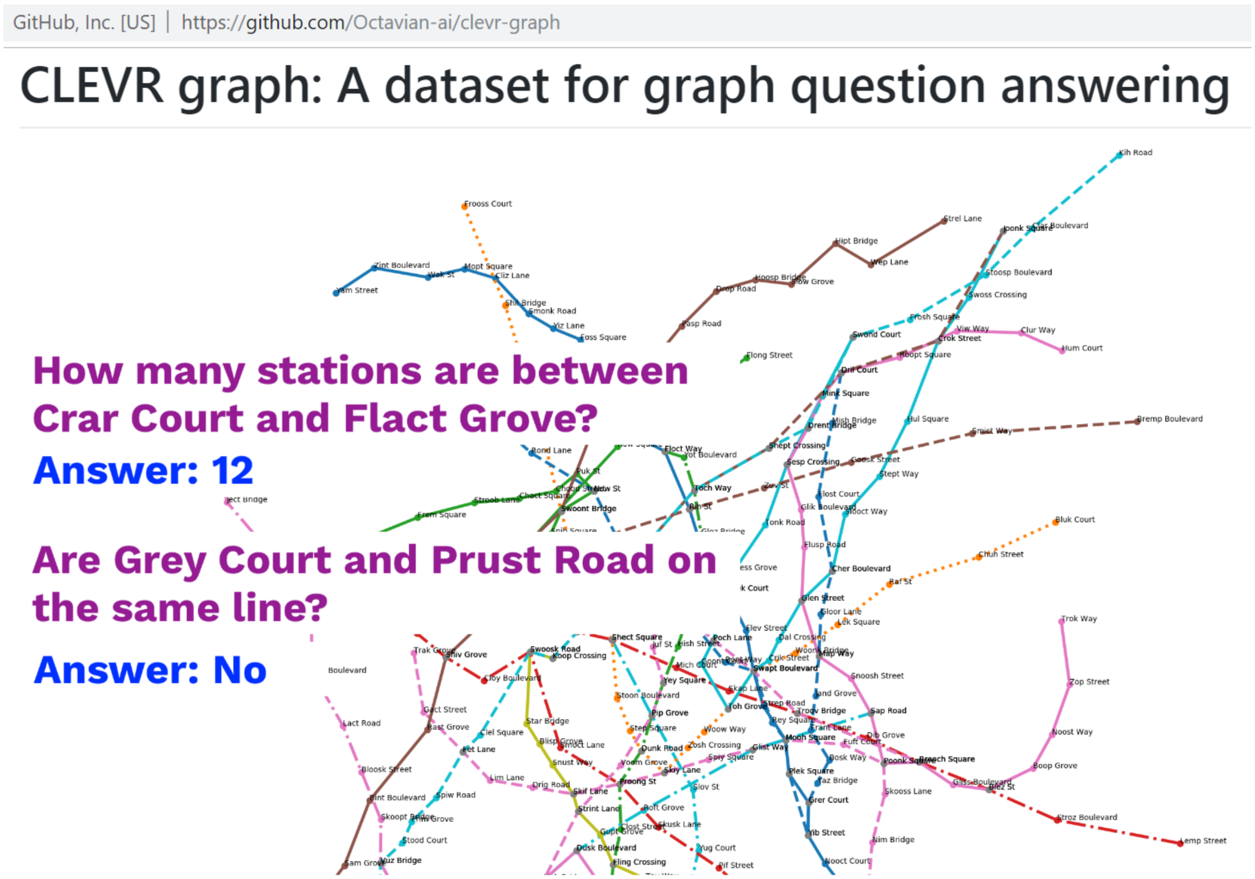
How to build it

Such a radical departure from how current query engines work requires a similar departure in the underlying technology. We're using a neural network as the core of the query engine. We present the database information as tables of data and adjacency matrices (e.g. an array of a connections) to the neural network, and let it process the data and query to produce a result. The network processes the query through an RNN and learned word embedding. This provides both an array of query tokens and also an overall query vector. The data is then processed through a network reminiscent of the Transformer architecture. After applying learned embeddings to data, it is passed through series of attention systems https://arxiv.org/abs/1706.03762 and https://arxiv.org/abs/1902.10186. These allow the network to leverage task-specific sub-networks and to combine earlier calculations together to form complex aggregates. Working example here.

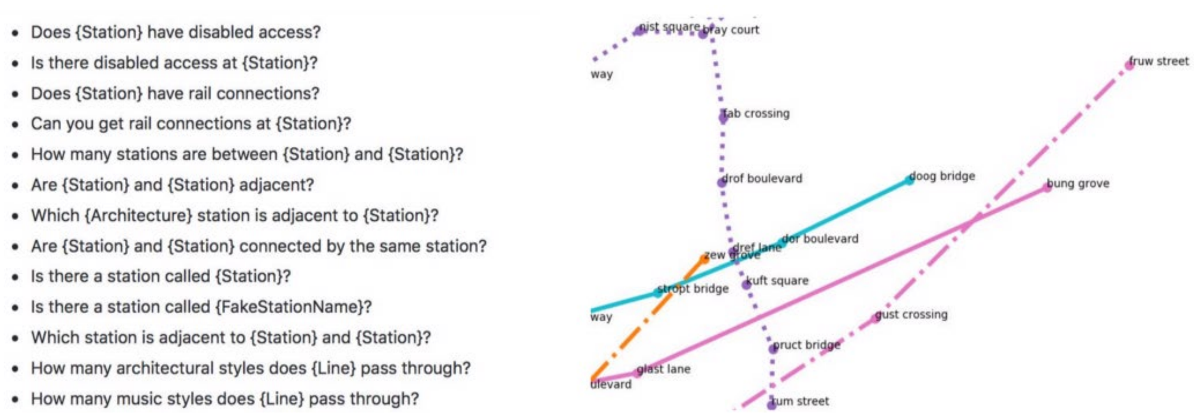
Some of the sub-networks include (for our graph-processing network):
Node property recall
Edge (i.e. relationship) recall
Using previous step's output as addressing instructions for the above
Iterative message passing
Recalling previous step's output and transforming them in a range of ways

EXPLORE KEY PAPERS in this zipped folder http://bit.ly/ML-MISC-01 and the list provided here:
https://arxiv.org/abs/1706.03762
https://arxiv.org/abs/1806.01261
https://arxiv.org/abs/1803.03067
https://arxiv.org/abs/1711.09846
https://arxiv.org/pdf/1905.12107.pdf
https://github.com/deepmind/graph_nets

[Andrew Jefferson](#) [Octavian](#) http://bit.ly/GKG-DL-TALK-2018 DB query engine [see our recent article](#).

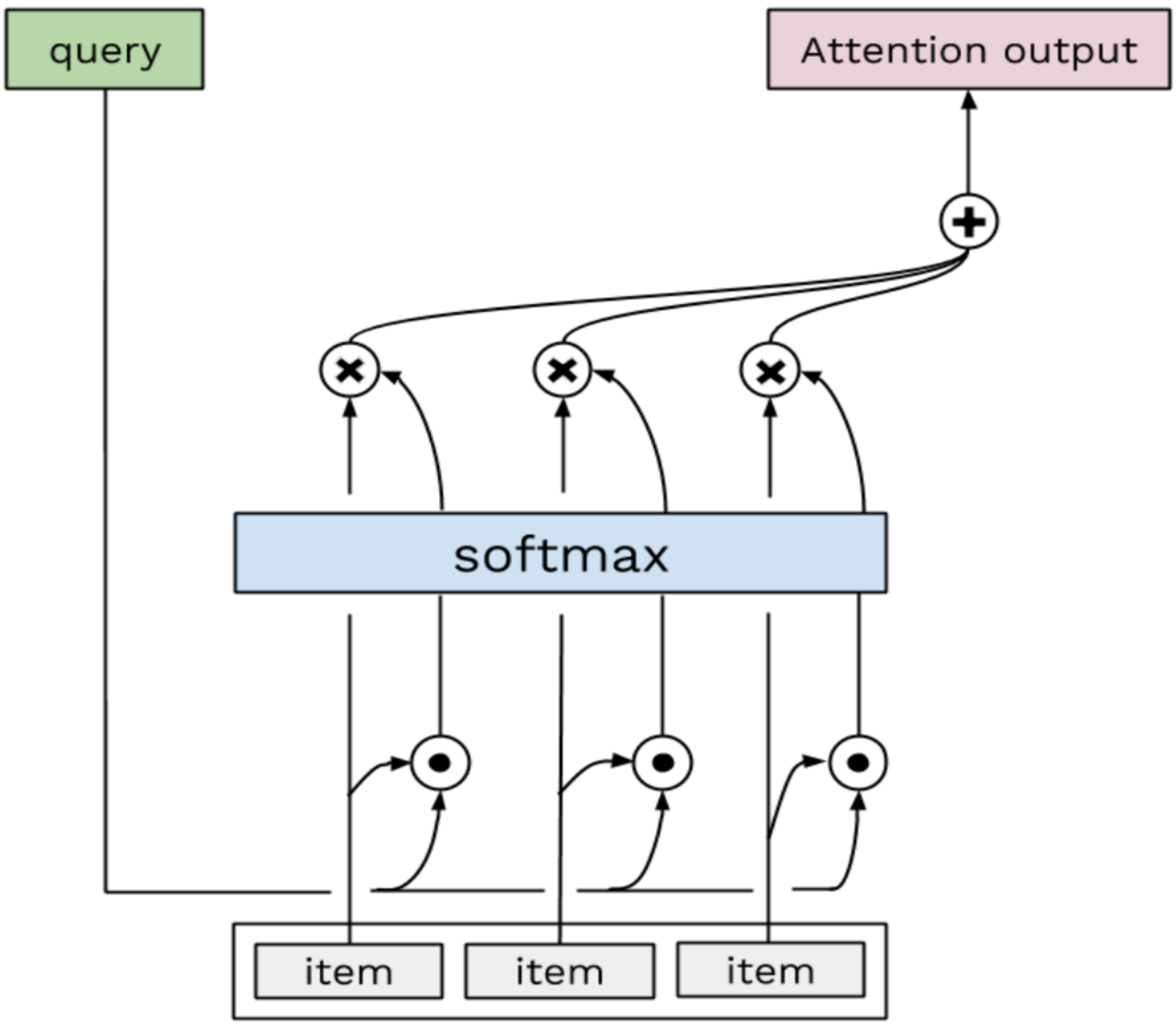


The graph data is modelled on transit networks (London tube and train network). Questions are modelled on questions typically asked by passengers (users) around mass transit (How many stops between? Where do I change?). Aim: solution to this dataset has real world applications.



Is "ATTENTION" insufficient?

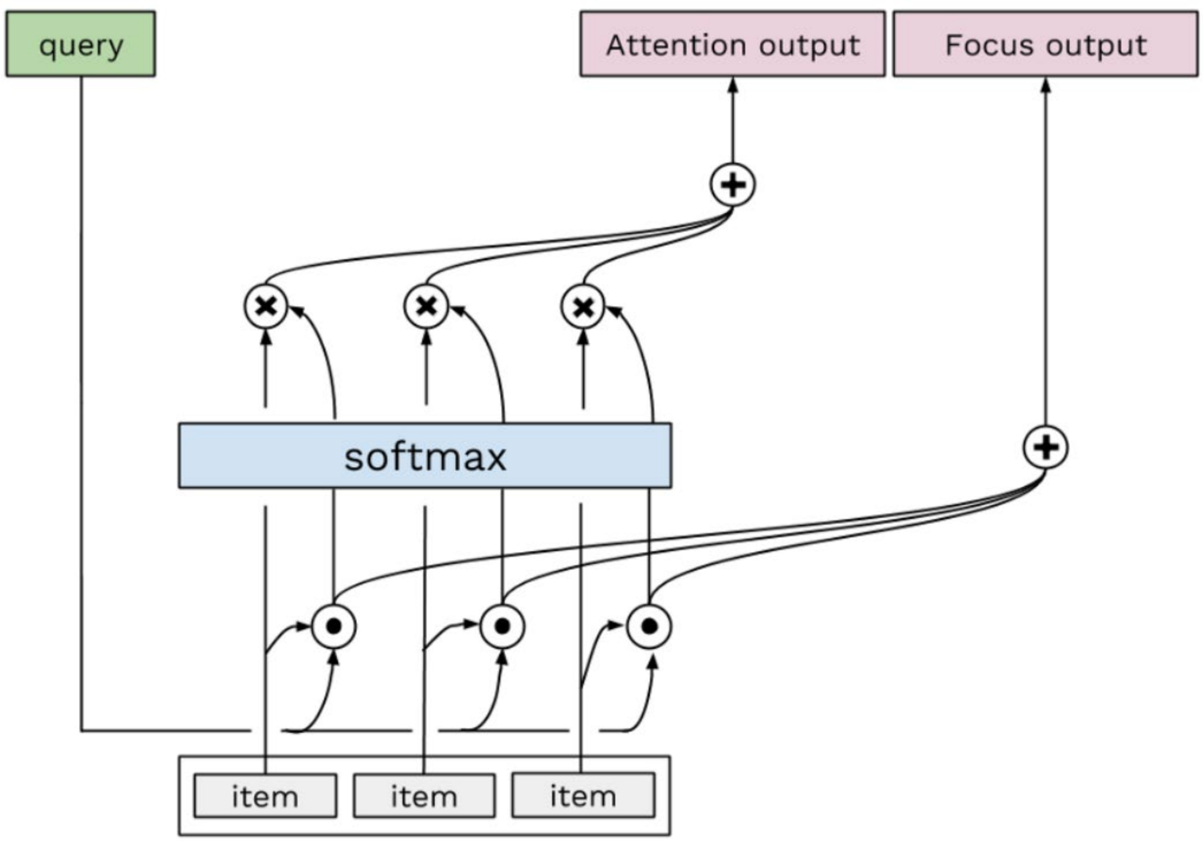
How to show if some things are not present in a list?



A popular neural-network technique for working with lists of items (e.g. translating sentences treating them as lists of words) is to apply "attention". This is a function where a learnt "query" of what the network is looking for is compared to each item in the list, and a weighted sum of the items similar to the query is output. Attention is the basis of current best-in-class translation models. The mechanism has worked particularly well because tasks can be solved by rearranging and combining list elements to form a new list (e.g. attention models have been important components in best of class translation, question-answering, reasoning models).

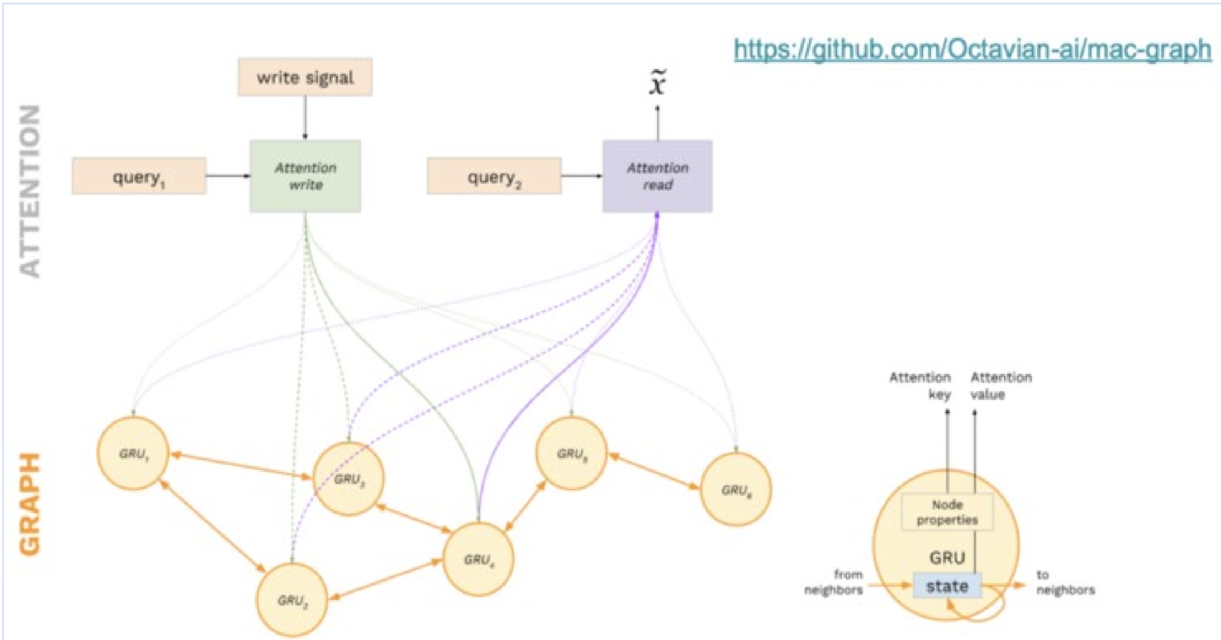
Despite attention's versatility and success, it has a deficiency that plagued our work on graph question answering: **attention does not tell us if an item is present in a list.** This first happened when we attempted to answer questions like "Is there a station called London Bridge?" and "Is Trafalgar Square station adjacent to Waterloo station?". Our tables of graph nodes and edges have all this information for attention to extract, but attention itself was failing to successfully determine item existence. This happens because attention returns a weighted sum of the list. If the query matches (e.g. scores highly) against one item in the list, the output will be almost exactly that value. If the query did not match any items, then a sum of all the items in the list is returned. Based on attention's output, the rest of the network cannot easily differentiate between those two situations.

The simple solution we propose is output a scalar aggregate of the raw item-query scores (e.g. before using softmax). This signal will be low if no items are similar to the query, and high if many items are. In practice this has been very effective (indeed, the only robust solution of the many we've tested) at solving existence questions. From now on we will refer to this signal as "focus".



Attention with focus signal, using a summation for aggregation of raw scores

https://github.com/Octavian-ai/attention-focus

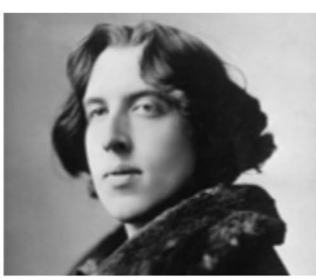


What is the cleanliness level of {Station} station?	99.9% accuracy after 10k training steps
How big is {Station}?	
What music plays at {Station}?	
What architectural style is {Station}?	
Describe {Station} station's architectural style.	
Is there disabled access at {Station}?	
Does {Station} have rail connections?	
Can you get rail connections at {Station}?	
Are {Station} and {Station} adjacent?	99% accuracy after 20k training steps
Which {Architecture} station is adjacent to {Station}?	98.8% accuracy after 30k training steps
How many stations are between {Station} and {Station}?	98% accuracy up to ~9 apart after 25k training steps
Are {Station} and {Station} connected by the same station?	98% accuracy after 20k steps
Which station is adjacent to {Station} and {Station}?	
Is there a station called {Station}?	99.9% accuracy after 30k training steps
Is there a station called {FakeStationName}?	
Are {Station} and {Station} on the same line?	Not yet tested
How many architectural styles does {Line} pass through?	Not yet tested
How many music styles does {Line} pass through?	Not yet tested
How many sizes of station does {Line} pass through?	Not yet tested
How many stations with rail connections does {Line} pass through?	Not yet tested
Which lines is {Station} on?	Not yet tested
How many lines is {Station} on?	Not yet tested
Which stations does {Line} pass through?	Not yet tested
What's the nearest station to {Station} with disabled access?	Not yet tested

Some example questions from CLEVR graph question bank. It's a synthetic (procedurally generated) dataset which consists of 10,000 fictional transit networks modelled on the London underground. For each randomly generated transit network graph we have a single question and correct answer. Each graph used to test the network is one the network has ***never seen before***. Therefore, it ***cannot memorise*** the answers to the questions but must learn how to extract the answer from new graphs.

The data (DIDA'S) to knowledge (KIDS) transition in my essays are eloquently explained by **Dan McCreary.** I am *copying* a few of his articles that are relevant to these essays. I have edited the content and removed leading suggestions (subtle advertising). Please access original versions from www.danmccreary.com/

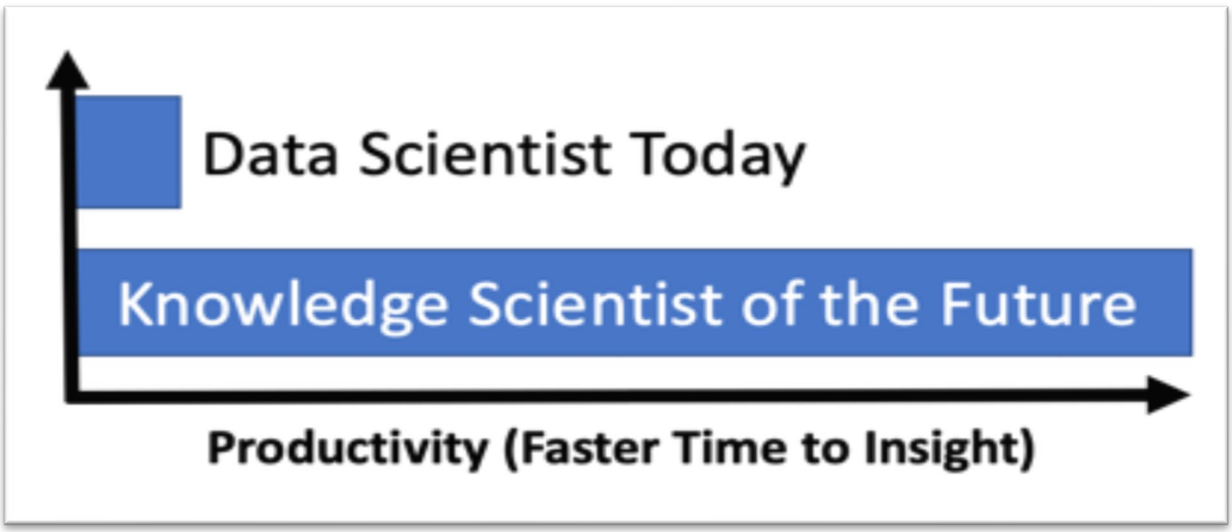
Another resource worth exploring is here http://bit.ly/Yann-LeCun but it contains a myriad of exaggerated claims and ideas of solutions "provided by powerpoint".



"Imitation is the sincerest form of flattery that mediocrity can pay to greatness."

– **Oscar Wilde**

From Data Science to Knowledge Science



Knowledge scientists may be more productive than data scientists because they may offer a new set of assumptions about the inputs to their models and store their insights in a knowledge graph. Their input features may remain **highly connected** to other relevant data as such as provenance and lineage metadata. An article pointed out: *Data scientists…spend from 50-80% of their time as mundane laborers, collecting and preparing unruly digital data, before it can be explored for useful nuggets.*

Cleaning up data involves data cleanup code. **Feature Store** is an attempt to build reusable artifacts for data scientists. Google and Uber have discussed their efforts to build tools to reuse features and standardize the feature engineering processes. My big concern is that many of these efforts are focused on building flat files of disconnected data. Once the features have been generated they can easily become disconnected from reality. They quickly start to lose their relationships to the real world.

An alternative approach is to build a set of tools for analysts to connect directly to a well-formed enterprise scale knowledge graph to get a subset of data and transform it to structures that are immediately useful for analysis. The results of this analysis can be used to enrich a knowledge graph. These Machine Learning approaches can complement the library of turn-key graph algorithms.

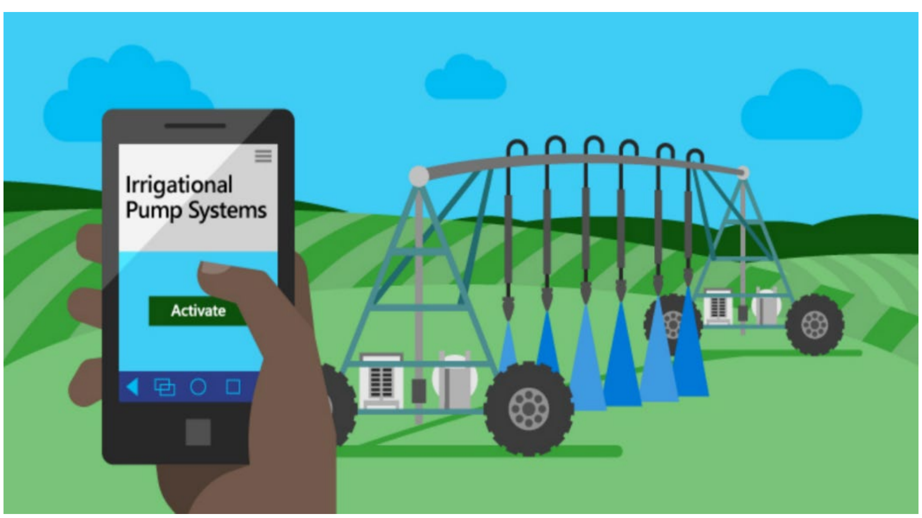
Data quality within a knowledge graph: MarkLogic is a document store where native data is stored in either JSON or XML documents. It promotes productivity through [1] document-level data quality score and [2] implicit query language-level validation of both simple and complex data structures. In MarkLogic, a built-in metadata element called the **data quality score** is usually an integer between 1 and 100 that assigned as new data enters the system. A low score (<50), indicates quality problems (missing data elements, fields out of acceptable ranges or corrupt or inconsistent data). A score of 90 may indicate that it could be used for downstream processes. Documents with a score >70 or >80 may improve search or analysis performance. To accomplish this task, a validation schema is built-in to MarkLogic. The concept of valid data is also built into W3C document query language (XQuery). Each document can be associated with a root element (within a namespace) and bound to an implicit set of rules about that document. GUI editor (oXygen XML Schema editor) allows non-programmers to create and audit data quality rules. XML Schema validation can generate a true/false Boolean value as well as a count of the number of errors in the document. Together with tools like Schematron and external data checks, each data steward can determine how to set the data quality score for various documents.

Document-level data quality scores are a natural fit with **events**, as a part of the workflow (for example, inbound call to call center, new customer purchase, subscription renewal, enrollment in a new healthcare plan, a new claim being filed, new sensor installed, new sensor data stream goes live). All of these events can be captured as complete documents, stored in streaming systems like Kafka and ingest the business event data to knowledge graphs. Data quality scores can be included in business event documents and knowledge graph (use it anywhere it is necessary for analysis).

In contrast, dumping table-by-table data from relational databases into CSV files are full of numeric codes that may not have clear meaning. Curating this low-level data from “data lakes” to deliver meaningful connected knowledge is an arduous task. Storing flattened CSV-level data and numeric codes is where features stores fall short. Once the features are extracted and stored in a data lake or object store, they become disconnected from how they were created. A new process might run on the knowledge graph that raises or lowers the score associated with a data item. However, that feature can’t easily be updated to reflect the new score. Feature scores can add latency that will prevent new data quality scores from reflecting the current status. Perhaps, relational data architects (using tables to store data) tend to under-value document models and associating data quality score with event documents.

In summary, the quality in a graph is different than quality in a document (please explore https://en.wikipedia.org/wiki/SHACL). The connectedness of a vertex in a graph will also determine quality (please explore https://www.w3.org/TR/prov-dm/). Eventually, knowledge graph products may have a metadata layer about how data was collected and transformed during the journey from the source to the knowledge graph. As RDF fades into the annals of W3C, we see a concomitant rise of the LPG (labeled property graph) ecosystem. LPG still lacks mature machine-learning integration tools to enable knowledge science.

Note: *This is a step beyond ART toward knowledge. It signals a move, in principle, from data-informed decision as a service (DIDA'S) to knowledge-informed decision as a service (KIDS). The cartoon below is a stand-alone digital proxy for irrigation pump system. The relevance of the data and information (before pump activation) must be* ***correlated*** *with soil moisture, optimum moisture saturation desired for the crop, the weather (prediction of rain within an acceptable window of time or forecast for even higher temperature which may accelerate loss of moisture from the soil), and other relevant information in the context of this action (activating irrigation pumps). ART can aggregate the information and ABS can direct pump activation (pump speed, volume of water, duration of action, coverage area, energy consumed). Analytical engines at the edge (on mobile phones) running short neural networks (https://arxiv.org/pdf/1803.03635.pdf) may assist ART, ABS, KIDS. This is an observation by the author and* ***not*** *due to Dan McCreary (http://bit.ly/GKG-KIDS).*

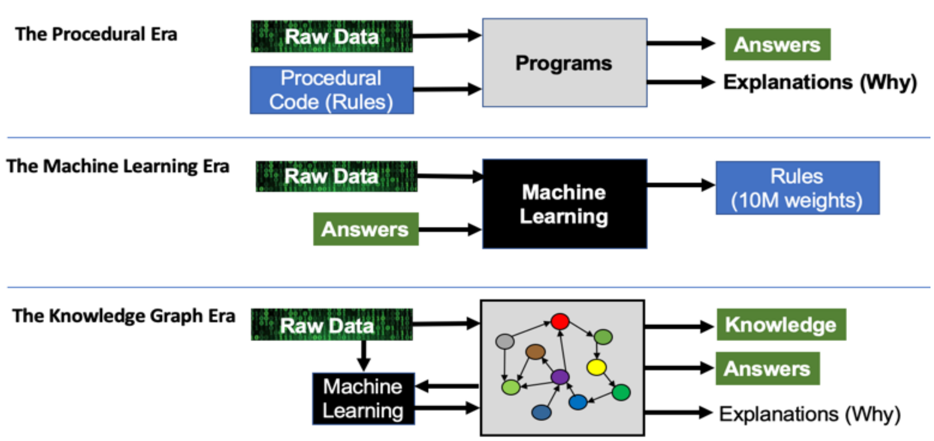


Link: https://cloudblogs.microsoft.com/industry-blog/microsoft-in-business/2019/05/15/how-polaris-energy-services-is-transforming-the-agriculture-industry-in-the-cloud/

Knowledge Graphs: The Third Era of Computing

"When did computing start?" A cuneiform tablet from circa 3,000 BC may hold the answer. Knowledge representation began when we wanted to remember things that were important to us. For example, ledger of financial transactions such as "X owes Y ten baskets of grain." It was natural to store these facts in rows and columns of a table because tables were a good "natural representation" for financial transactions. These transactions records evolved into rows of symbols which represented concepts and gave birth to written languages. These representations continued for 5,000 years. Clay tablets evolved into papyrus scrolls, then Luca Pacioli's double entry bookkeeping system, which eventually became punch cards and then flat files in COBOL, then tables in a row-store popular in relational databases and finally Enterprise Resource Planning (ERP) systems.

These tabular representations worked well when our problem had uniform data sets. By uniform we mean that each record (row) has similar attributes with similar data types. Not all problems fit well into tables. The more tables you have the more expensive the relational joins. How do we store the **analysis** of a patient chart? You might have a so-called AI agent scanning data (eg drug adherence data). What do they produce? The answer is often a list of conditions and the probability of occurrence (diabetes, asthma). Healthcare systems store these concepts in a complex hierarchy (a taxonomy) with many connections between the concepts (ontology). The AI tool may recommend next best actions. This analytical data (outcome) may not fit easily into a table. But it does fit well into a graph of connected concepts, the knowledge graph.



Page 61

The bulk of developers are still writing PHP/Java/Python over tables (not graphs). In the knowledge graph era, machine learning continuously reads raw data, combines this with existing knowledge and produces new knowledge, answers and explanations. Knowledge graphs are at the core of the third era of computing, aimed to enrich shared knowledge.

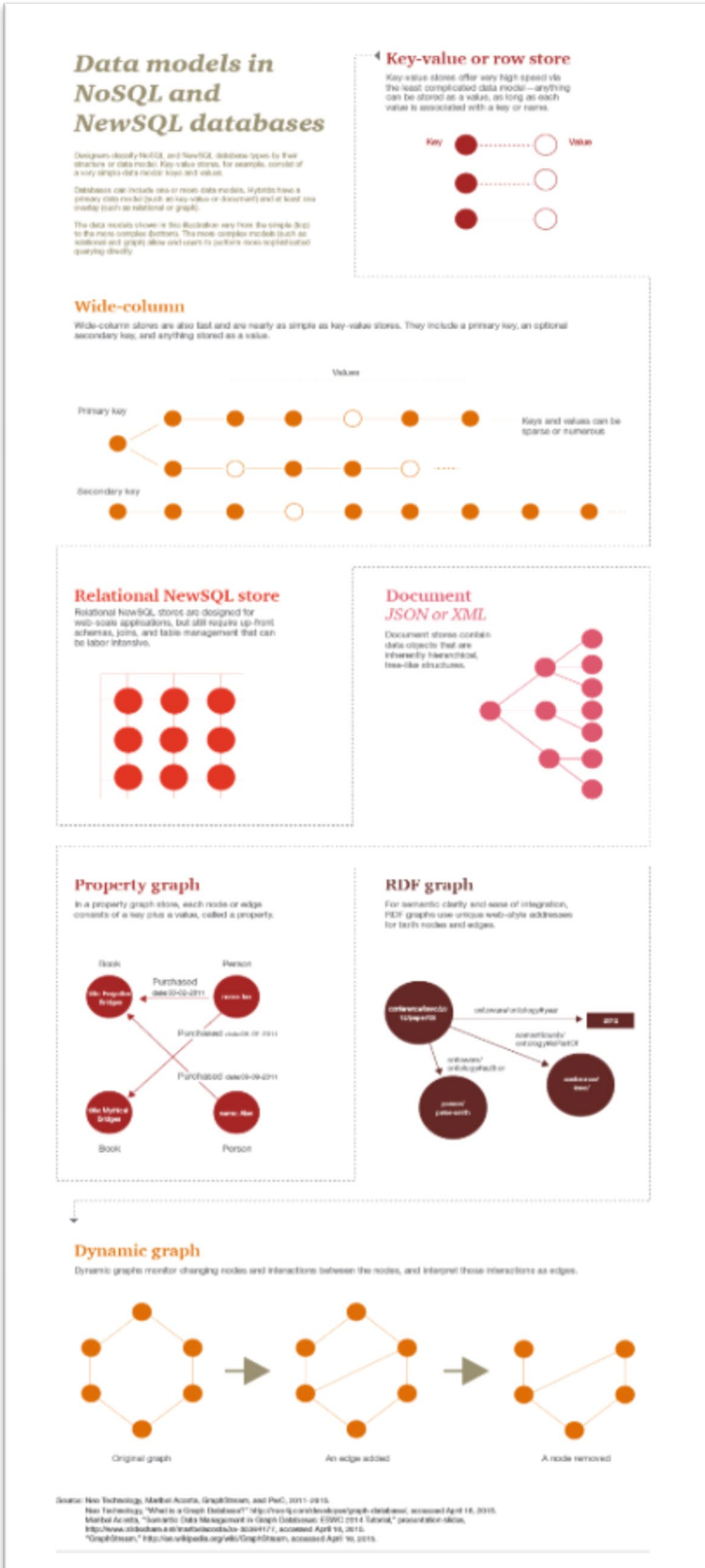
Knowledge graphs combine to produce a system that not only learns from complex data, but it also can explain its decisions. We use machine learning to harvest raw data and look for patterns in this data. Machine learning finds relevant information (people, places and things) in images, texts and sound. ML converts this to new entries in our knowledge graph along with confidence weights. The data can be checked for consistency and quality by graph algorithms. The outcome from the graph is new knowledge, answers and explanations of why we made specific decisions. Our knowledge graph becomes a repository of semantically precise vertices and relationships with confidence weights retained from the machine learning processes.

However, knowledge enrichment processes are not perfect and can easily add false assertions if new facts are not curated by subject matter experts (promotes fake news).

The justified emergence of knowledge graphs as a buzzword surrounds its ability to use very large distributed graph databases to store complex networks of concepts that can be "traversed" using a highly parallel graph query language. Knowledge graphs in Google, Facebook, LinkedIn, Amazon (product graph) and Pinterest (interest graph) have over 100 billion vertices, thus solving the scalability problems for graph databases.

In the past, the predominant way of building knowledge graphs was to use hand coded knowledge and an inference engine that could leverage higher-level RDF-based standards such as RDFS, OWL and SKOS. Now organizations are using machine learning to build seed concept graphs using natural language processing (NLP). There is also a strong shift to use the more flexible labeled-property-graphs (LPGs) to do similar reasoning (http://bit.ly/WHY-ARTIFICIAL-REASONING).

◆ https://medium.com/@dmccreary/how-knowledge-graphs-promote-fake-news-362947220ea8
◆ https://medium.com/@dmccreary/blockologies-a-pattern-language-for-ai-data-flows-de9f4507547



63 ◆ Exploring how sensor repositories may be helpful. See Figure 11 on page 28 – http://bit.ly/SIGNALS-SIGNALS

◆ Di ◆ Data-informed ◆ exploration of the tessellated facets of our elusive quest for meaning

“SIGNALS” contains a series of essays spewing amorphous thoughts:

1. SITS
2. SIP-SAR
3. SARS♠AG
4. ART
5. PEAS -------- **You are here**

Please review “SIGNALS”

● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

Alternate http://bit.ly/SIGNALS-SIGNALS

SIGNALS is part of the collection of essays (book)

“IoT is a Metaphor” (see Commentary E)

Please review “IoT is a Metaphor”

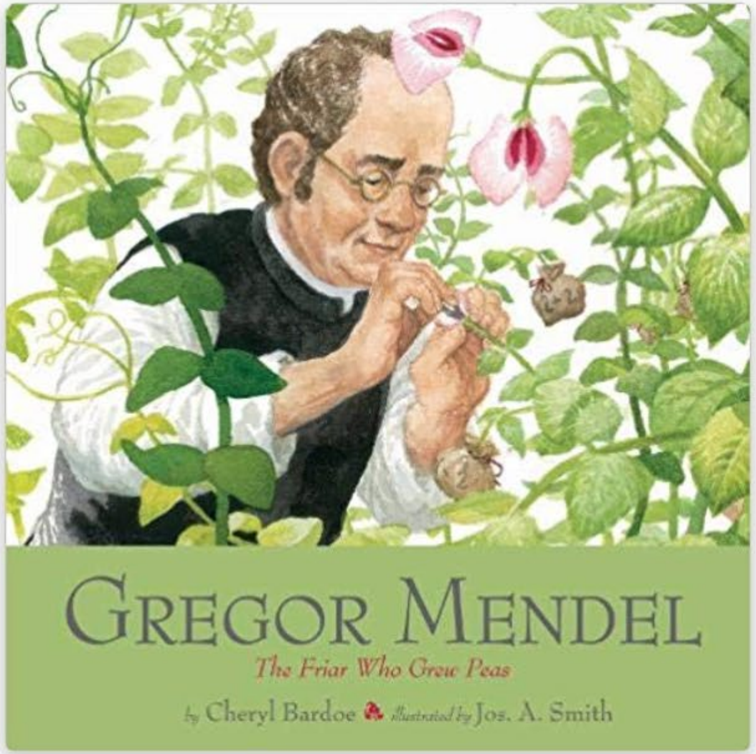
● Download PDF from MIT Library

https://dspace.mit.edu/handle/1721.1/111021

The page numbers in the PDF do not match the document page numbers at the bottom of the pages. This PDF is made up of a few separate documents which were then combined. Hence, the temporary mis-match of page numbers. Apologies.

PEAS is mnemonic borrowed from the literature on Agent based systems (ABS) designed (modeled) to address systems performance (P) in the context of environment (E) of operation, events or processes or systems, to be actuated (A), based on information from primary sources, for example, sensor (S) data. http://bit.ly/P-E-A-S and https://link.springer.com/chapter/10.1007/978-981-10-8258-0_8

Gregor Mendel: The Friar Who Grew Peas



PEAS

Platform for the Agro-Ecosystem

Eric Scott McLamore[1] and Shoumen Palit Austin Datta[1,2,3]

[1] Nano-Bio Sensors Lab, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611
[2] Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
[3] MDPnP Lab, Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, Cambridge, MA 02139

- **P**erformance
- **E**nvironment
- **A**ctuators
- **S**ensors

PEAS PLATFORM for the Agro-Ecosystem

Decision Science for Agriculture and Agri-Business

http://bit.ly/P-E-A-S

https://prezi.com/go_0ow_w-3n6/agroecosystem-analysis-aesa/

– **P**erformance

KIDS

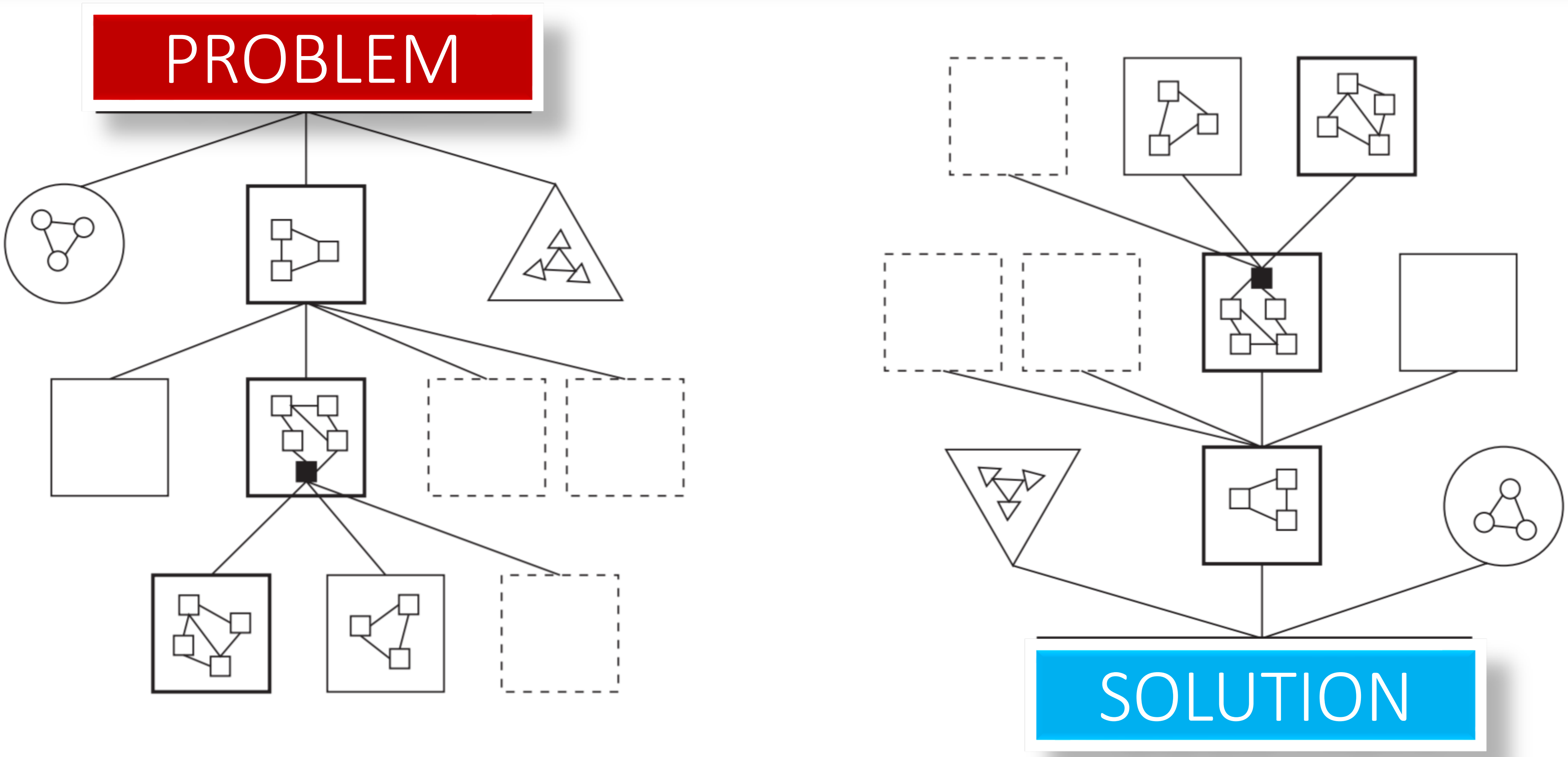
– **S**ensors

Knowledge-Informed Decision as a Service

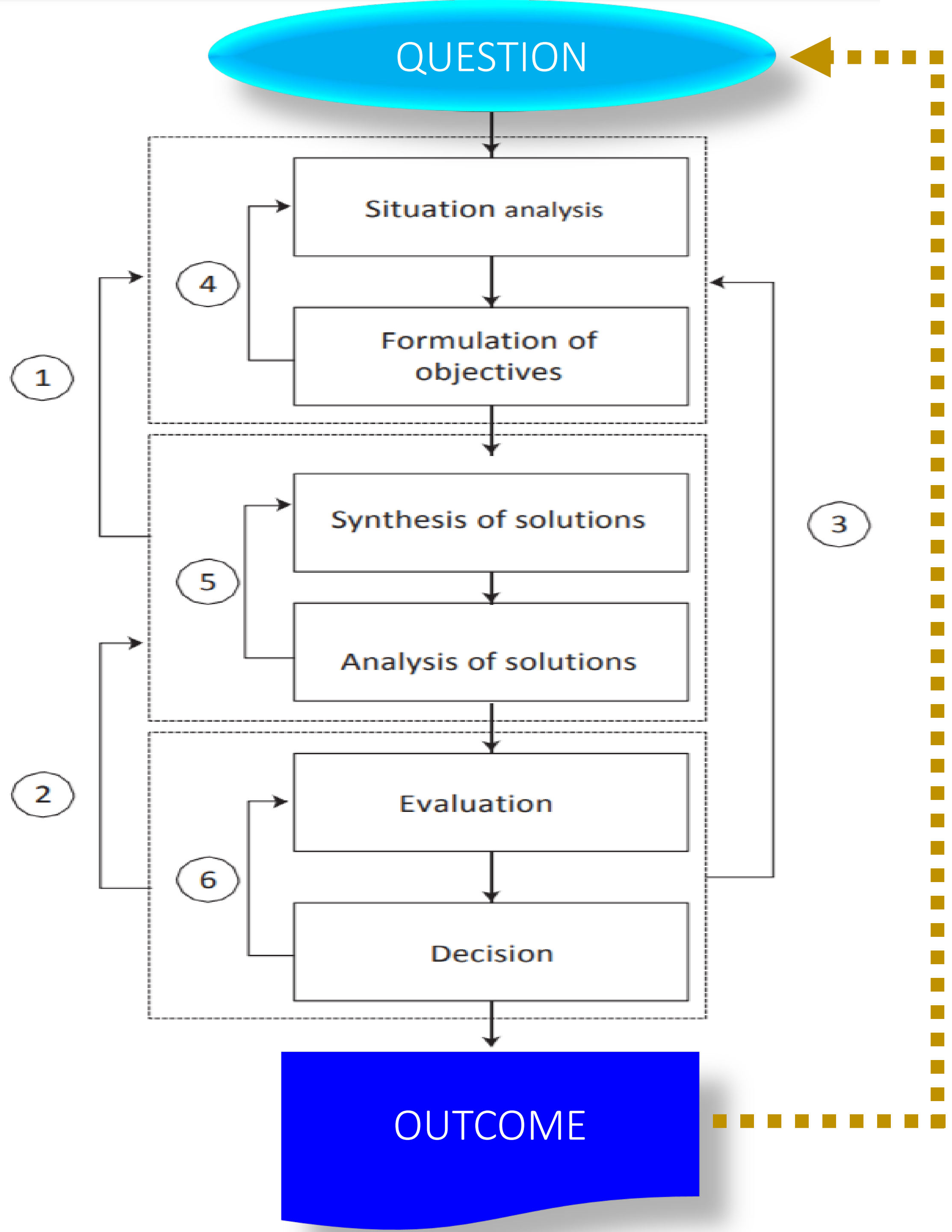
KIDS

The outcome and the value of the service is the key performance indicator.

De-construction and Re-construction



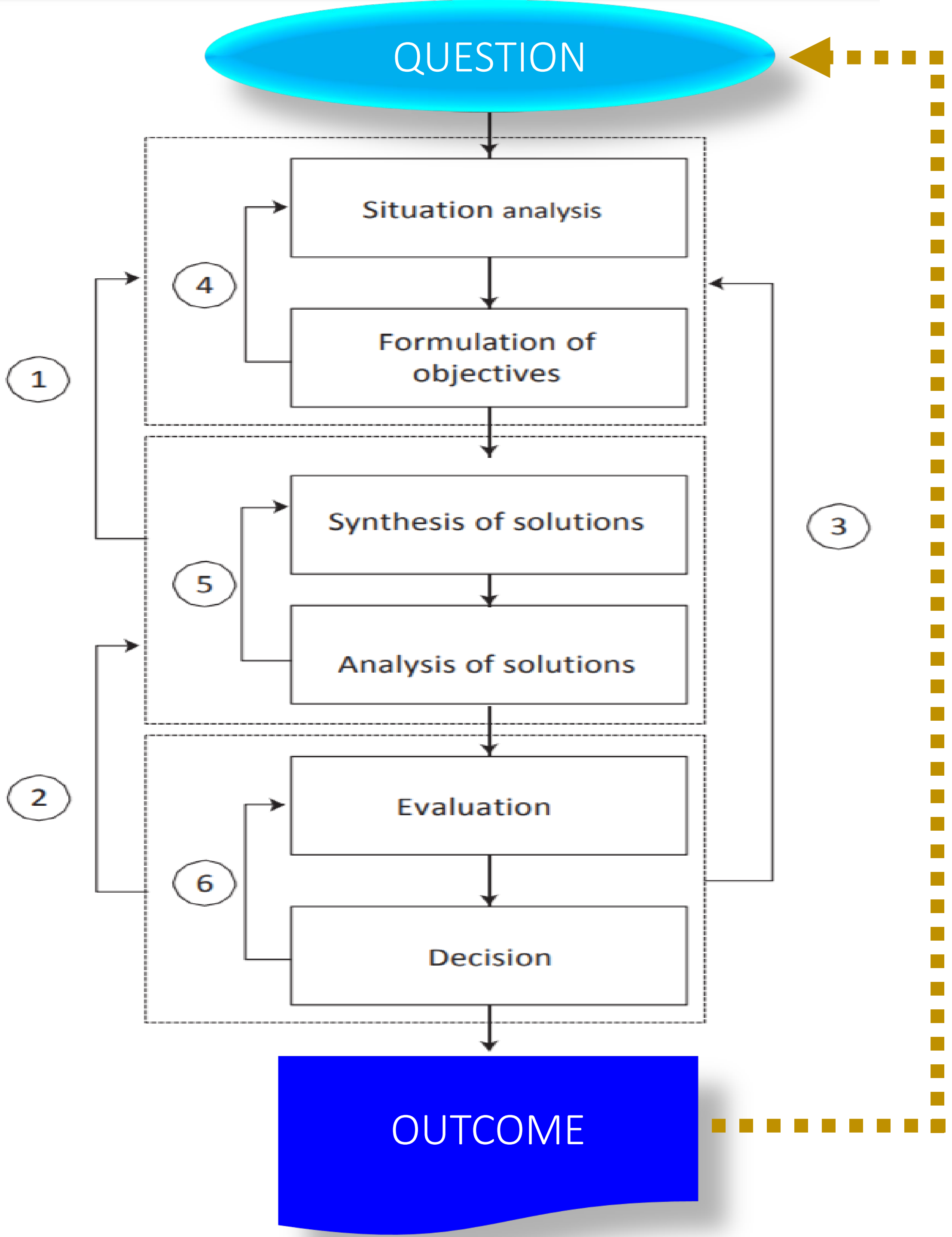
A Systems Engineering Approach? https://link.springer.com/book/10.1007/978-3-030-13431-0



QUESTION
Situation analysis
Formulation of objectives
Synthesis of solutions
Analysis of solutions
Evaluation
Decision
OUTCOME
1
2
3
4
5
6

A Systems Engineering Approach?

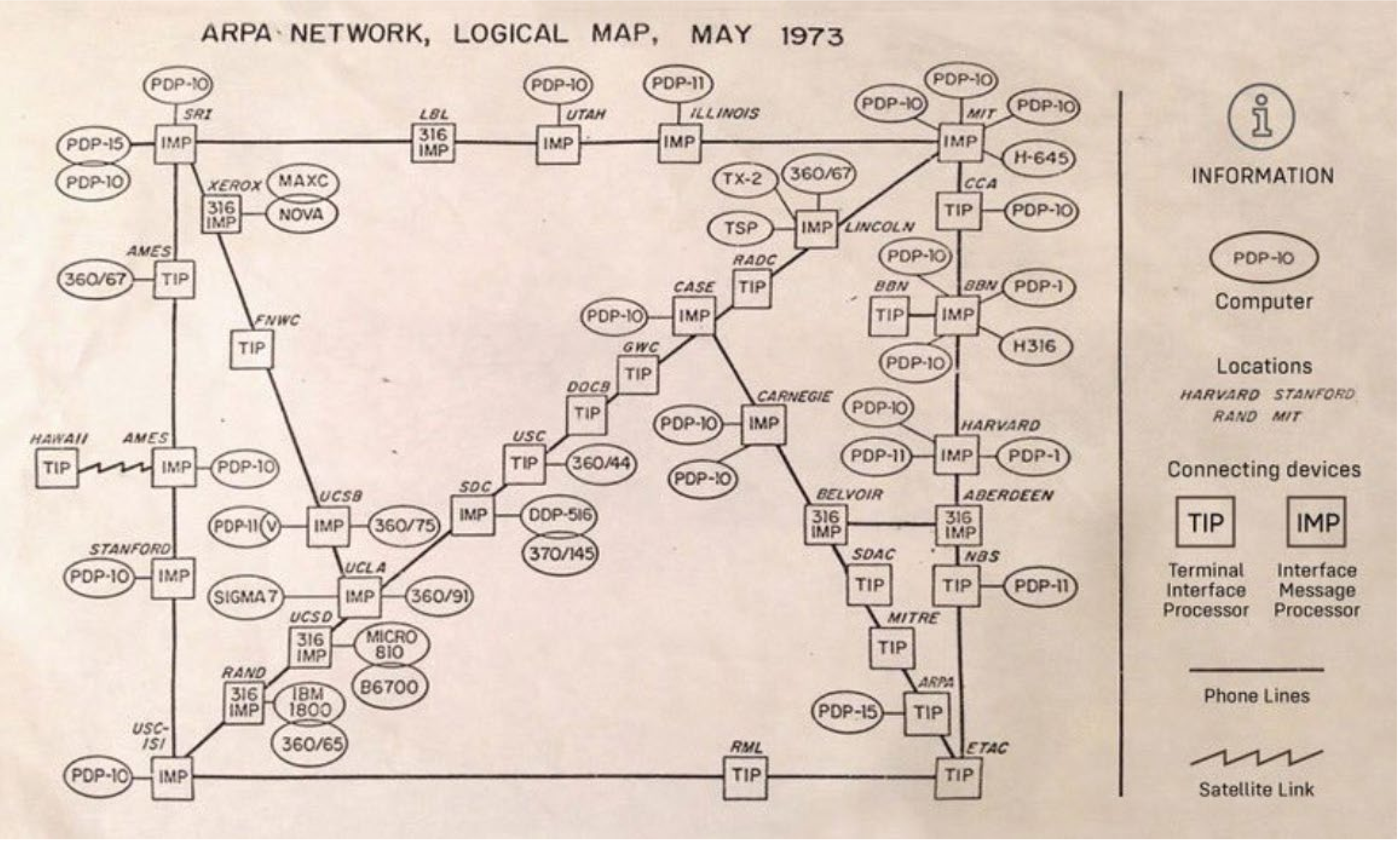
– What is the problem? Is it the correct problem to address?
– Boundaries of the problem space (dynamic vs static).
– Principal influences/mechanisms relative to the context of the problem and problem space.
– Needs a new or re-configured solution?
– What are the solution/system boundaries?
– Requirements of the solution space (system, design goals)
– Feasibility (contextual, technical, economic, social, ecological)
– Solution space to be reconstructed based on existing system or create/innovate architecture to execute solution system?



What is

KIDS

Knowledge-Informed Decision as a Service


ARPA NETWORK, LOGICAL MAP, MAY 1973
SRI
PDP-10
PDP-15
PDP-10
IMP
LBL
316 IMP
UTAH
PDP-10
IMP
ILLINOIS
PDP-11
IMP
MIT
PDP-10
PDP-10
PDP-10
H-645
IMP
XEROX
MAXC
NOVA
316 IMP
TX-2
360/67
TSP
IMP
LINCOLN
CCA
TIP
PDP-10
AMES
360/67
TIP
RADC
TIP
CASE
PDP-10
IMP
BBN
TIP
BBN
IMP
PDP-10
PDP-1
H316
PDP-10
FNWC
TIP
GWC
TIP
DOCB
TIP
CARNEGIE
PDP-10
IMP
PDP-10
HARVARD
PDP-10
PDP-11
IMP
PDP-1
HAWAII
TIP
AMES
IMP
PDP-10
USC
TIP
360/44
UCSB
PDP-11
V
IMP
360/75
SDC
IMP
DDP-516
370/145
BELVOIR
316 IMP
ABERDEEN
316 IMP
STANFORD
PDP-10
IMP
UCLA
SIGMA 7
IMP
360/91
SDAC
TIP
NBS
TIP
PDP-11
UCSD
316 IMP
MICRO 810
B6700
MITRE
TIP
RAND
316 IMP
IBM 1800
360/65
ARPA
PDP-15
TIP
USC-ISI
PDP-10
IMP
RML
TIP
ETAC
TIP
INFORMATION
PDP-10
Computer
Locations
HARVARD STANFORD
RAND MIT
Connecting devices
TIP
IMP
Terminal Interface Processor
Interface Message Processor
Phone Lines
Satellite Link

PLATFORM

KIDS

Knowledge-Informed Decision as a Service

PLATFORM

KIDS is an open plan platform concept. Platforms are comprised of multiple applications and integrated solutions with embedded tools and databases that function as complete, seamless environments. Product innovation platforms are intended to support groups of users collaborating across various levels, domains, business units, and the ecosystem. These capabilities are increasingly needed throughout the entire extended enterprise in almost every vertical, agnostic of the type of application or function or users, including farmers, meat packers, produce growers, retail stores, customers, suppliers, and business partners. Developing open platform tools and technologies are not limited to any one domain because graph networks can overlay and configured for use, almost anywhere. KIDS also includes error correction, search engine algorithms, NLU/NLP (natural language processing), automated feature engineering, drag and drop functions, data analytics, workflows, and open services with plug & play interfaces. Human-computer interactions and data interoperability between system of systems are key elements in the KIDS model.

What are the questions?

ABOUT AGRO-ECOSYSTEM

This discussion is about FEWS. But we focus on a tiny part of the science and engineering issues at the nexus of food, energy, water and sanitation (FEWS). Our emphasis is on agriculture and food.
<u>KIDS aims to answer questions from end-users.</u>

For KIDS, food growers and farmers in the field, are the customers.

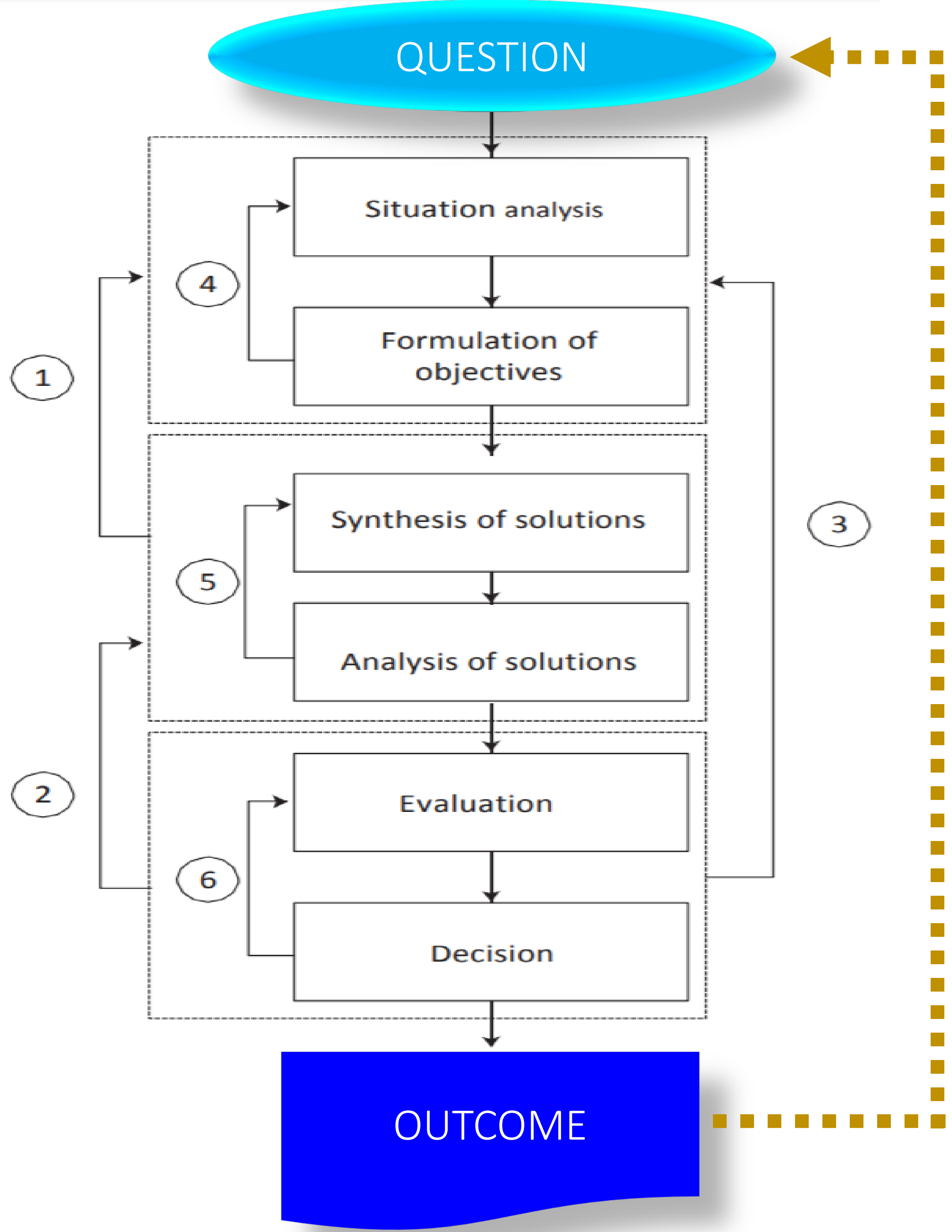
An educated consumer is the best customer.

Example of question from end-user

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Systems engineering approach can guide but it is woefully inadequate. Cannot stay in the “box” if we wish to answer.

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?



How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

The complexity of this question indicates the challenge for decision support system

Systems engineering approach may need several cycles of
deconstruction and reconstruction to analyze the question and disassemble
the sub-systems, components and data, necessary to attempt to answer the question.

Beyond the horizon of data-informed decision as a service (decision support system)

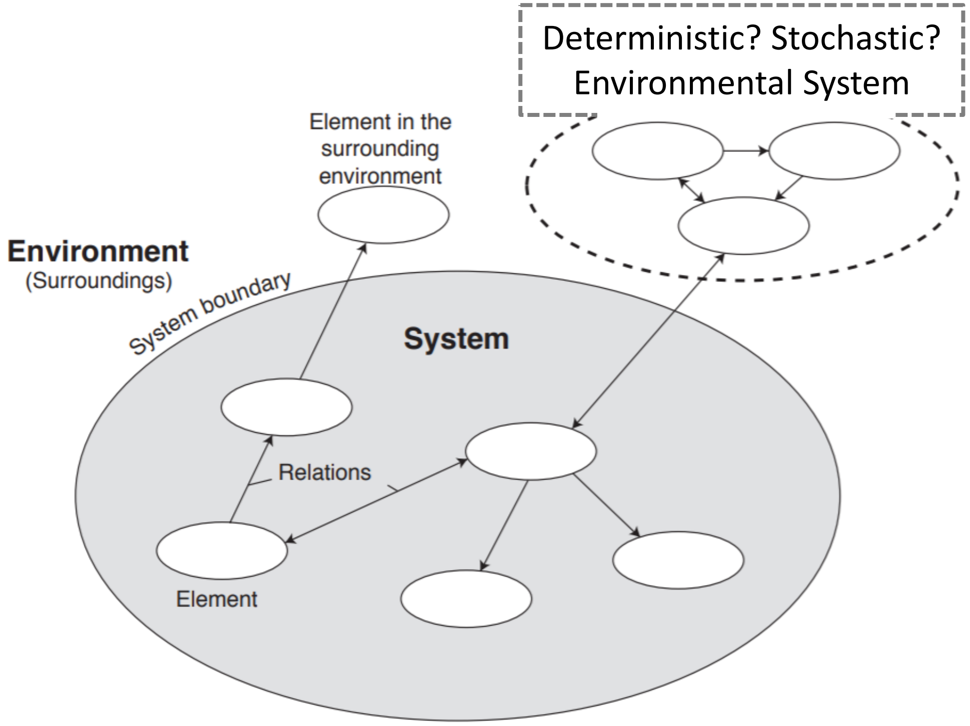
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

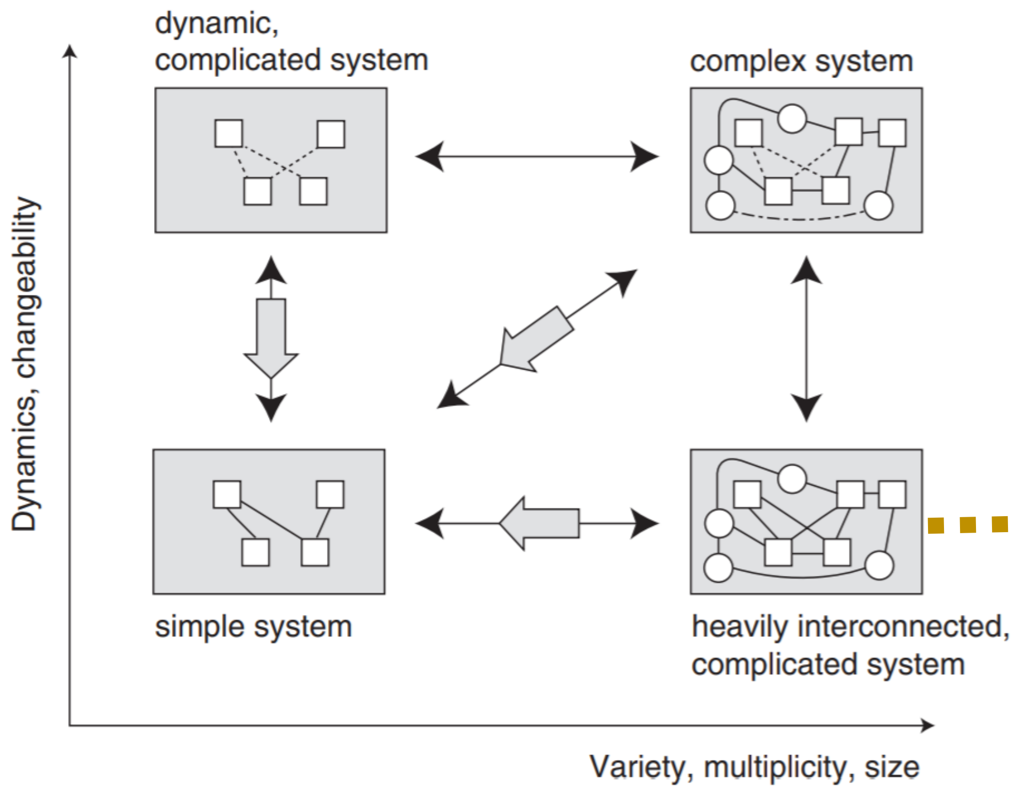
"*sacrificing my values*"

Prevailing decision systems are **data-informed**. Even extraction of actionable "information" stretches the reality. The semantics of this question represents an ecosystem of social "values" with respect to "sacrifices" which are personal in context of the user and her community. There are no tools or systems that can even attempt to answer the first part of the question to any degree of user satisfaction. In the short term, any answer may fail to meet an appreciable quality of service [QoS] level for which the user may recognize the value and may be willing to pay a fee (to receive the service). The best we are capable of delivering is the data-informed decision as a service (**DIDA'S**) which may be relevant to the cost and quality of the waste water which the user is seeking.

Beyond the horizon of data-informed decision as a service (decision support system)

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?





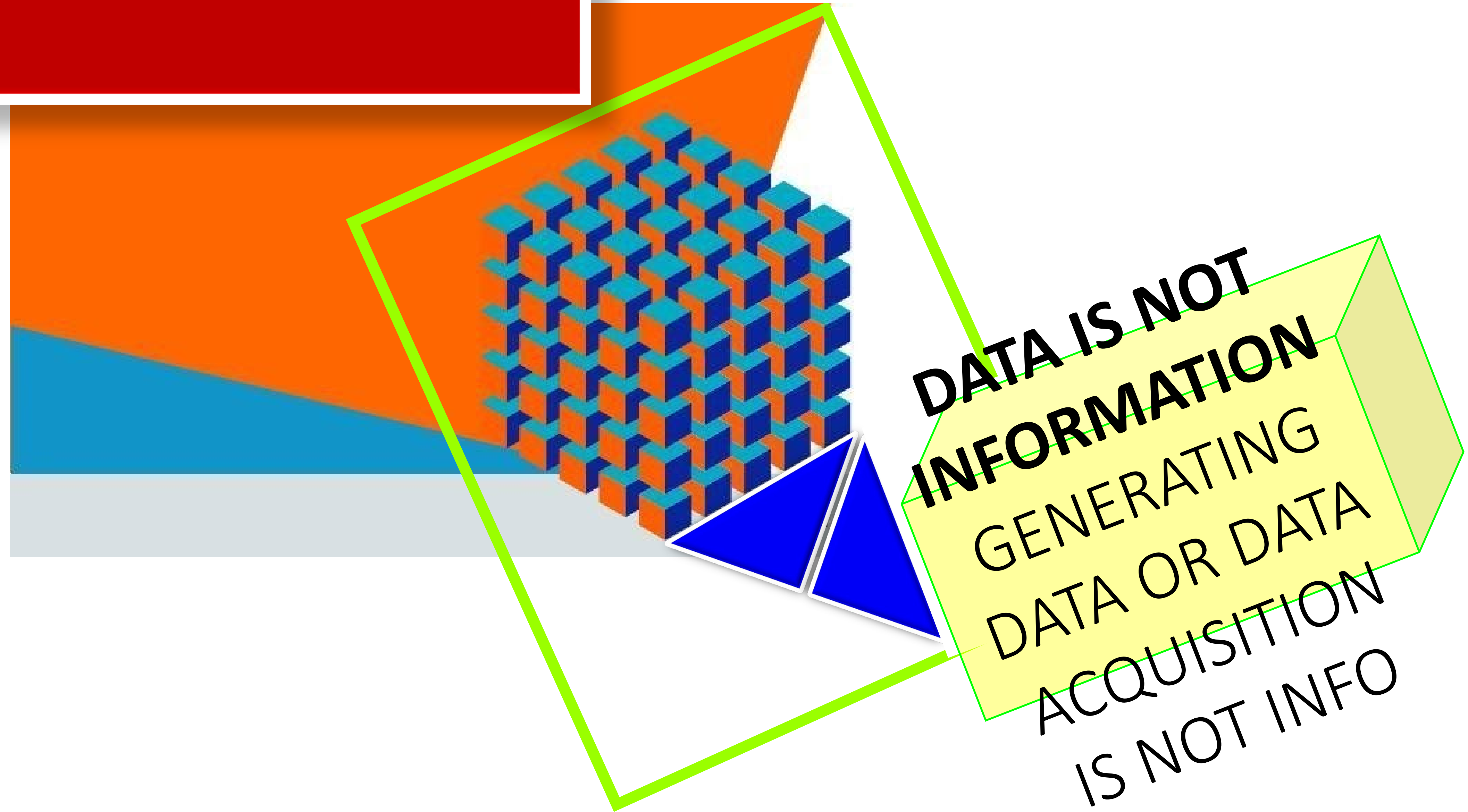
From a systems perspective, we have a heavily interconnected, complicated system

What happened to KIDS?

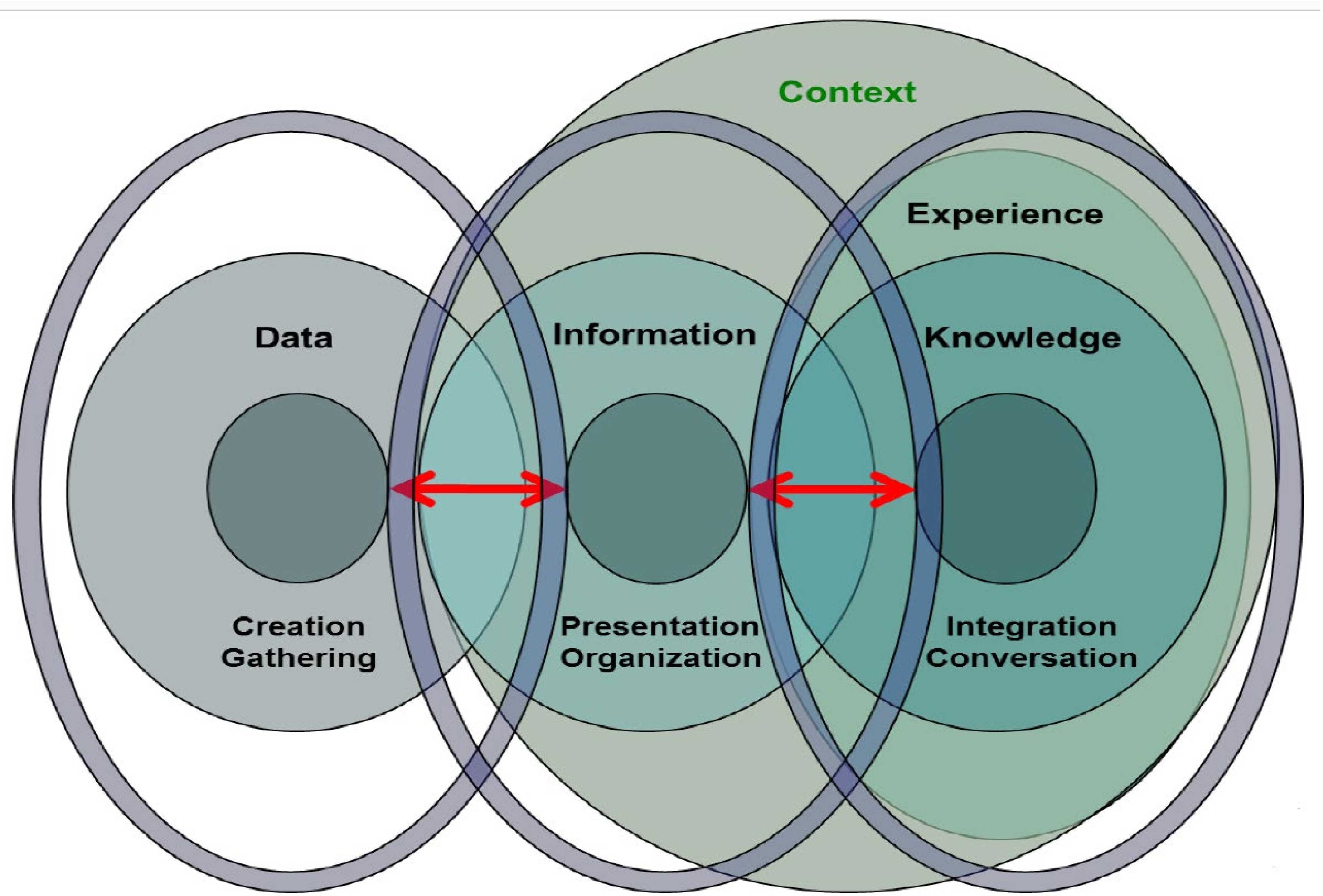
We initiated this discussion with **KIDS**, knowledge-informed decision as a service, but the user's question is compelling us to admit systemic inadequacies. Hence, we are stepping down, considerably, to recognize that the best outcome, at the present, may be limited to **DIDA'S** or data-informed decision as a service and stretch (?) our abilities to extract actionable information.

SUMMARY

1) Data ≠ Information ≠ Knowledge
2)
3)
4)



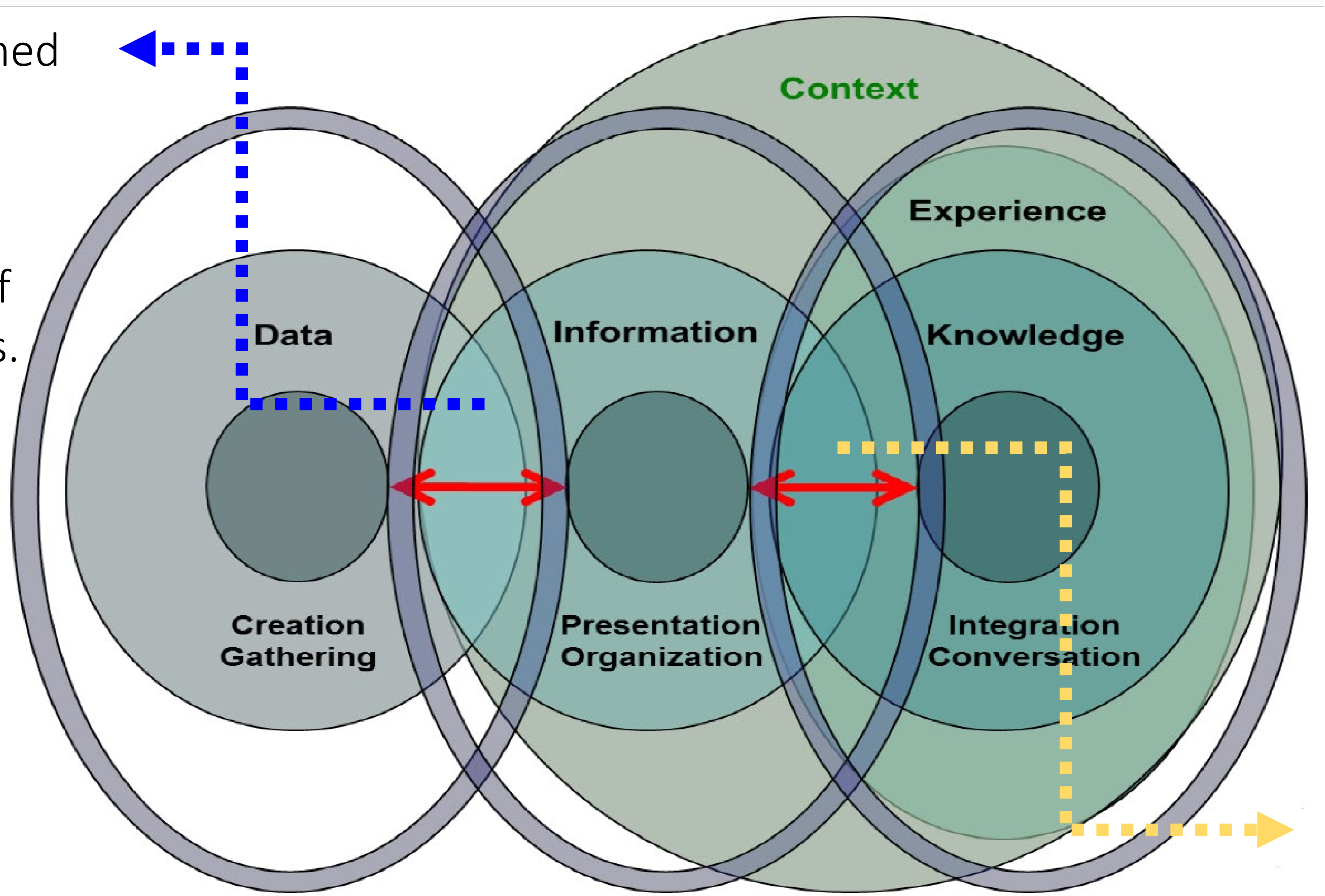
In terms of knowledge from data, our current status is analogous to 1950's, when TV was fuzzy black, greyish white, grainy & dull.



Cartoon: Jim Hendler

In terms of knowledge from data, our current status is analogous to 1950's, when TV was fuzzy black, greyish white, grainy & dull.

Data-informed may be the best case scenario at this stage of our systems.



Knowledge-informed is the Holy Grail. It may be decades away from reality. KIDS may aspire to reach this zone if artificial reasoning can escape the AI mis-information assault and move beyond classical expert systems.

Mind the Gap

There is a

vast

chasm between

data-informed

vs

knowledge-informed

http://www.kr.tuwien.ac.at/staff/tkren/pub/2008/rowschool2008.pdf

KIDS

Knowledge-Informed Decision as a Service

Convergence → *A Sense of the Future*

PLATFORM

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

DIDA'S

Data-Informed Decision as a Service

First, successfully deploy DIDA'S and create tools to extract actionable information. Then we may re-visit how to create KIDS, knowledge-informed decision as a service.

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

In the context of DIDA'S (Data-Informed Decision as a Service), re-visit the user's question with respect to cost, quality of wastewater and wastewater treatment.

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

In the context of DIDA'S (Data-Informed Decision as a Service), re-visit the user's question with respect to cost, quality of wastewater and wastewater treatment.

If we dissect deeper with our reductionist approach, we determine the need to answer the user's question in terms of water quality. What data DIDA'S may require for water quality?

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

In the context of DIDA'S (Data-Informed Decision as a Service), re-visit the user's question with respect to cost, quality of wastewater and wastewater treatment.

If we dissect deeper with our reductionist approach, we determine the need to answer the user's question in terms of water quality. What data DIDA'S may require for water quality?

SENSOR DATA

DIDA’S

In the context of DIDA’S (Data-Informed Decision as a Service), re-visit the user’s question with respect to performance, environment, actuators and sensors (PEAS).

DIDA’S

- Performance
- Environment
- Actuators
- Sensors

In the context of DIDA’S (Data-Informed Decision as a Service), re-visit the user’s question with respect to performance, environment, actuators and sensor data.

DIDA'S

- Performance
- Environment
- Actuators
- Sensors

http://bit.ly/P-E-A-S

DIDA'S

KIDS

- **P**erformance
- **E**nvironment
- **A**ctuators
- **S**ensors

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-825-techniques-in-artificial-intelligence-sma-5504-fall-2002/

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

In the context of DIDA'S (Data-Informed Decision as a Service), re-visit the user's question with respect to cost, quality of wastewater and wastewater treatment.

If we dissect deeper with our reductionist approach, we determine the need to answer the user's question in terms of water quality. What data DIDA'S may require for water quality?

1. Data from multiple sensors for water quality monitoring
2. Cost and pricing data for comparative analysis
3. Wastewater treatment tools and technologies

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Even the apparently simpler part of the question requires **multiple** sources of **data** and **convergence** of information to provide a sufficiently data-informed service to the user.

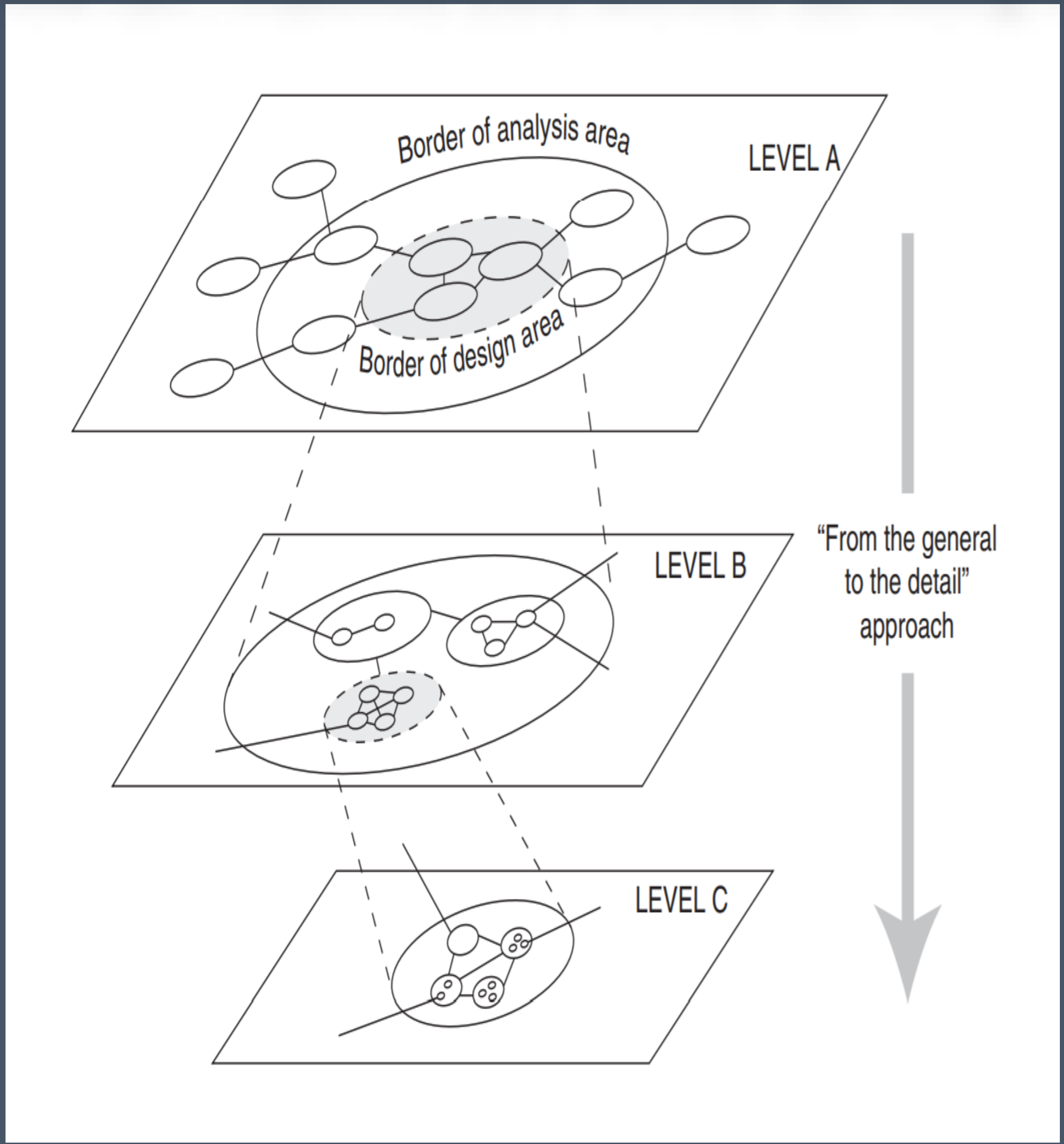
1. Data from multiple sensors for water quality monitoring
2. Cost and pricing data for comparative analysis
3. Wastewater treatment tools and technologies

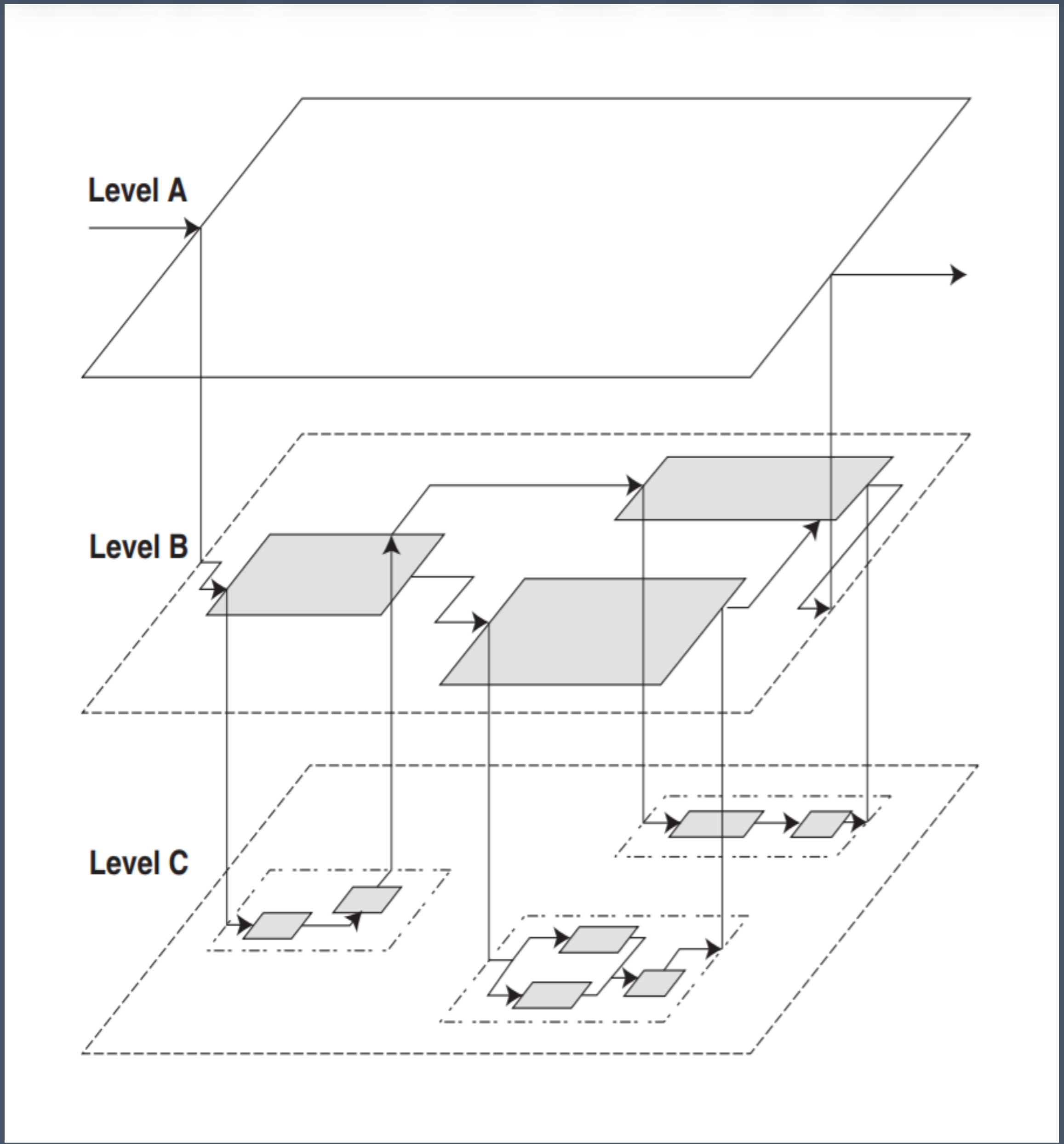
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Data fusion and convergence of information are only a part of the data-informed ***service*** users expect. The tasks are to delineate relationships germane to the question, select relevant data, connect, catalyze data fusion, synthesize information, extract "actionable" information, and deliver to a mobile device, in time, contextually relevant recommendation, of value, to the end-user.

1. Data from multiple sensors for water quality monitoring
2. Cost and pricing data for comparative analysis
3. Wastewater treatment tools and technologies

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

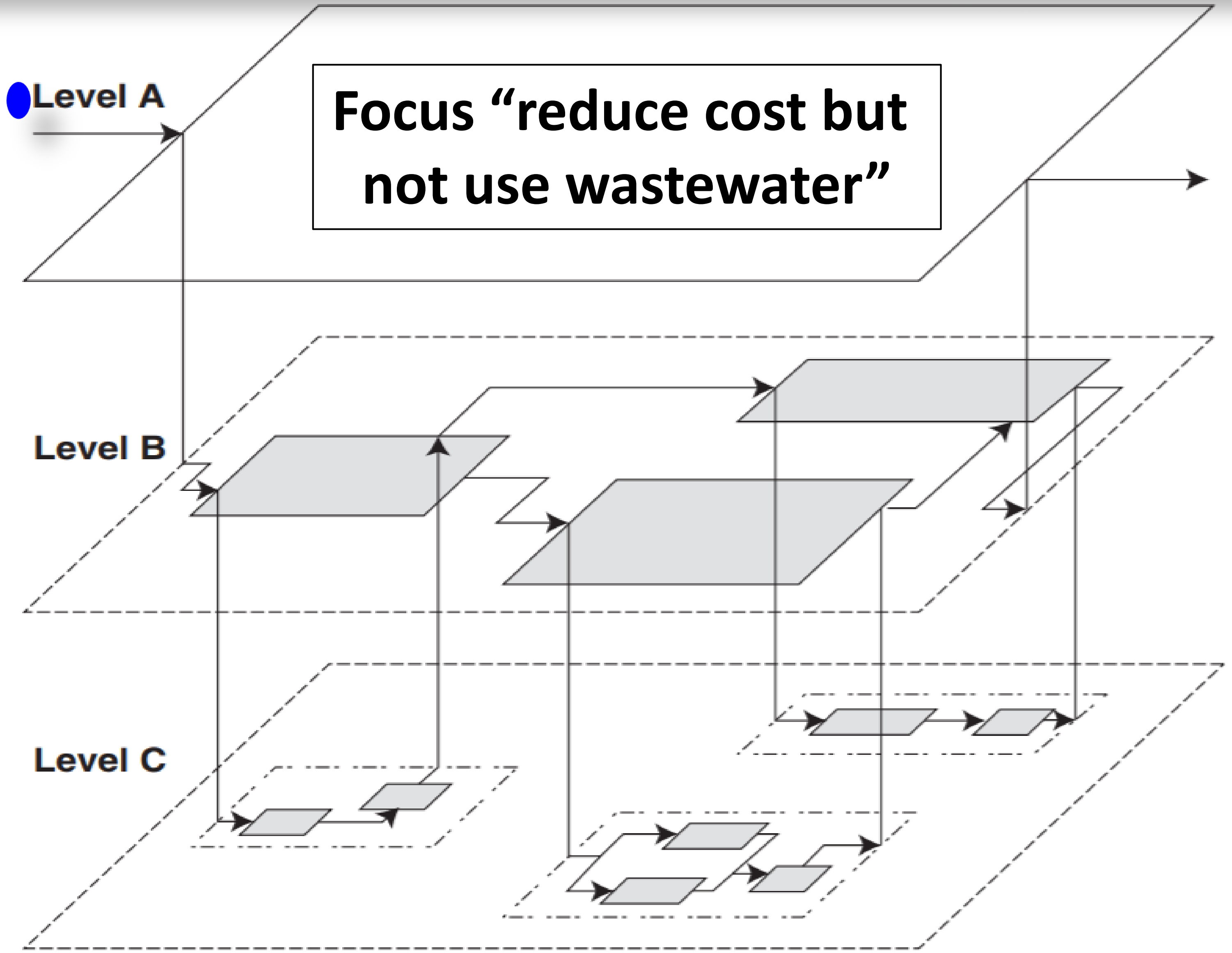




From a systems engineering approach, deconstruct the question in terms of data granularity.

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

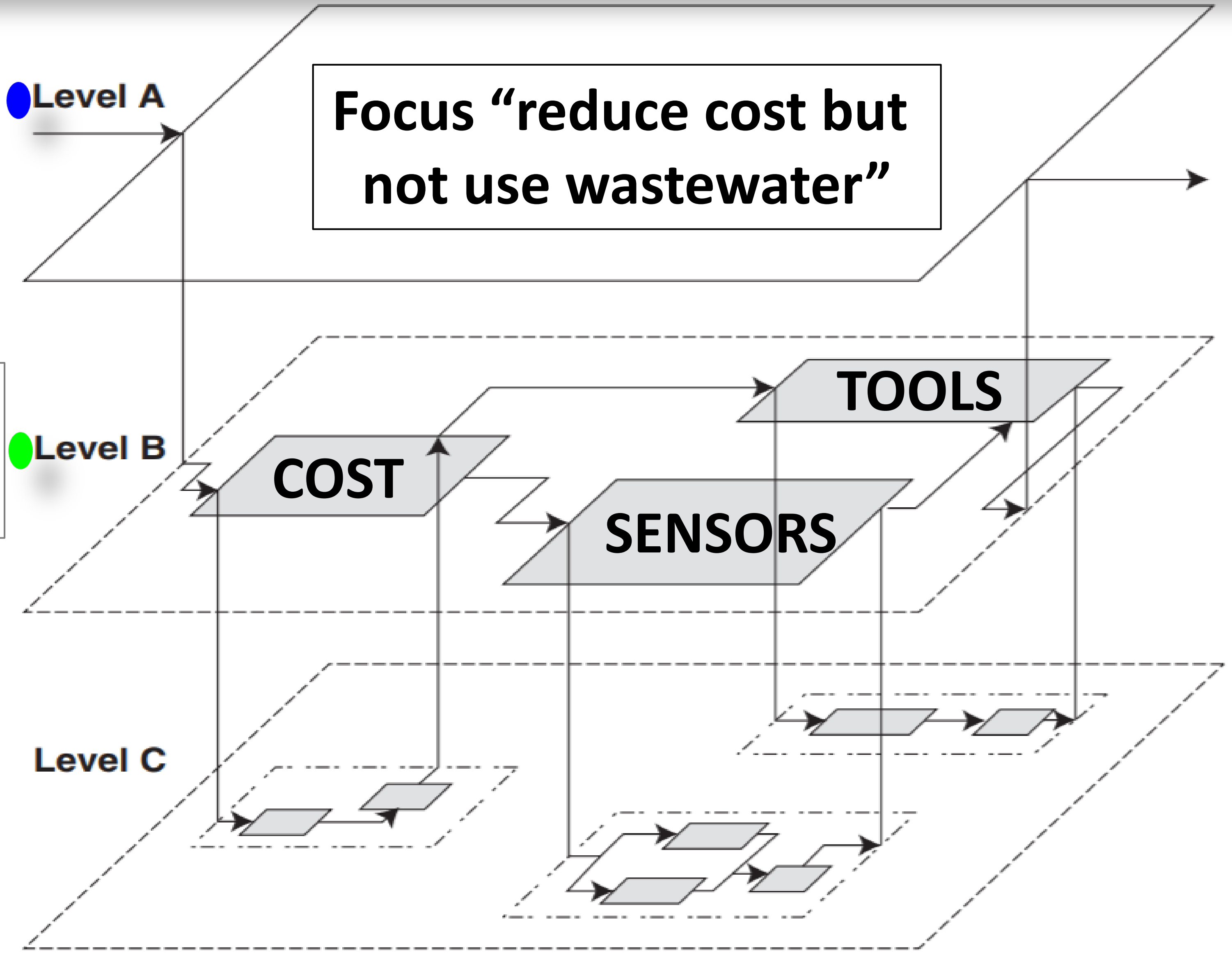
Focus “”reduce cost but not use wastewater



How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Focus “”reduce cost but not use wastewater

-Factors affecting cost
-Sensors to determine water quality
-Tools necessary for treatment of wastewater

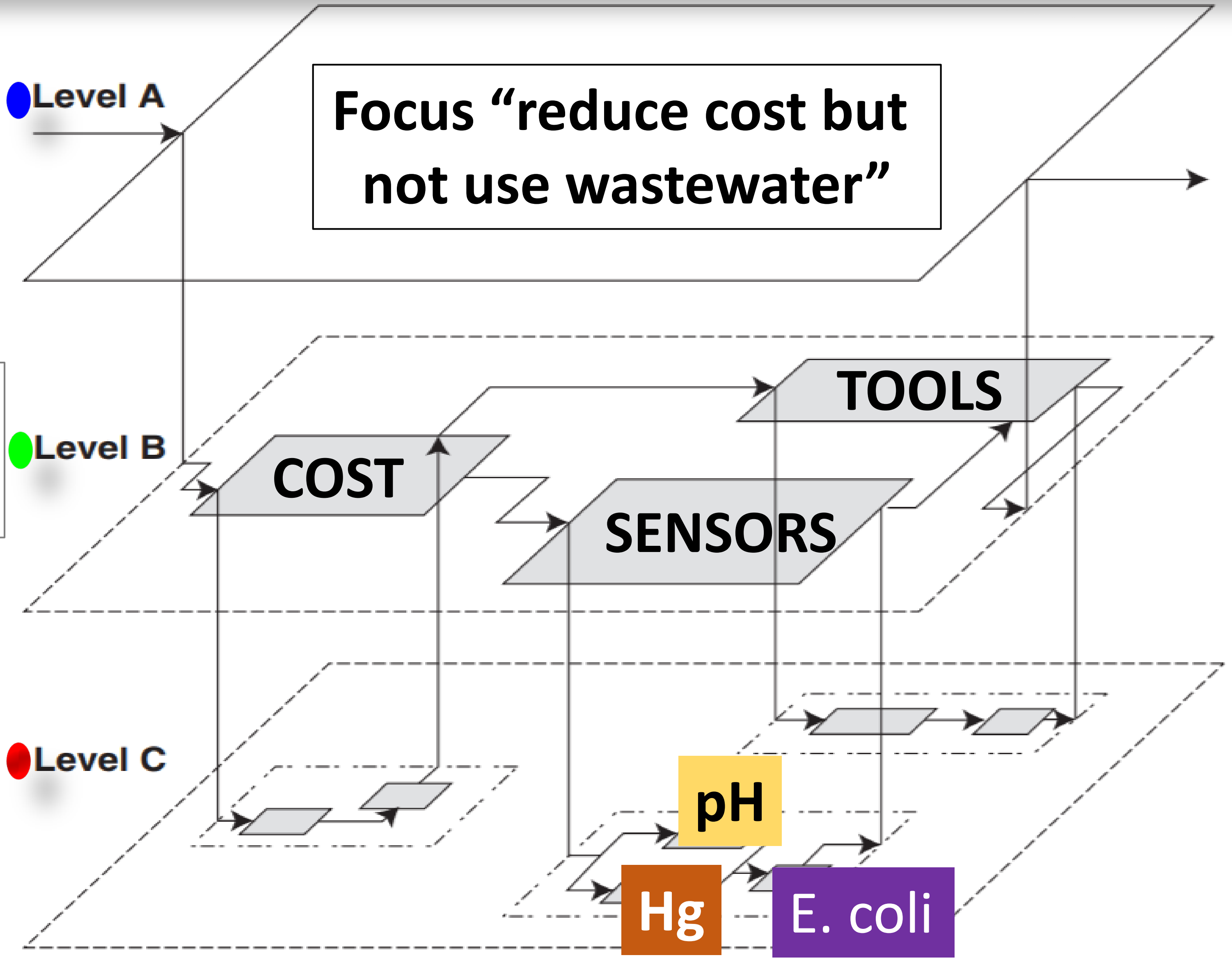


How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Focus “”reduce cost but not use wastewater

-Factors affecting cost
-Sensors to determine water quality
-Tools necessary for treatment of wastewater

- Sensors to determine water quality
  - Heavy metal pollution (Mercury)
  - Microbial contaminants (E. coli)
  - Chemical characteristics (pH)



Granular Deconstruction – Lowest Common Denominator

Raw Data Source

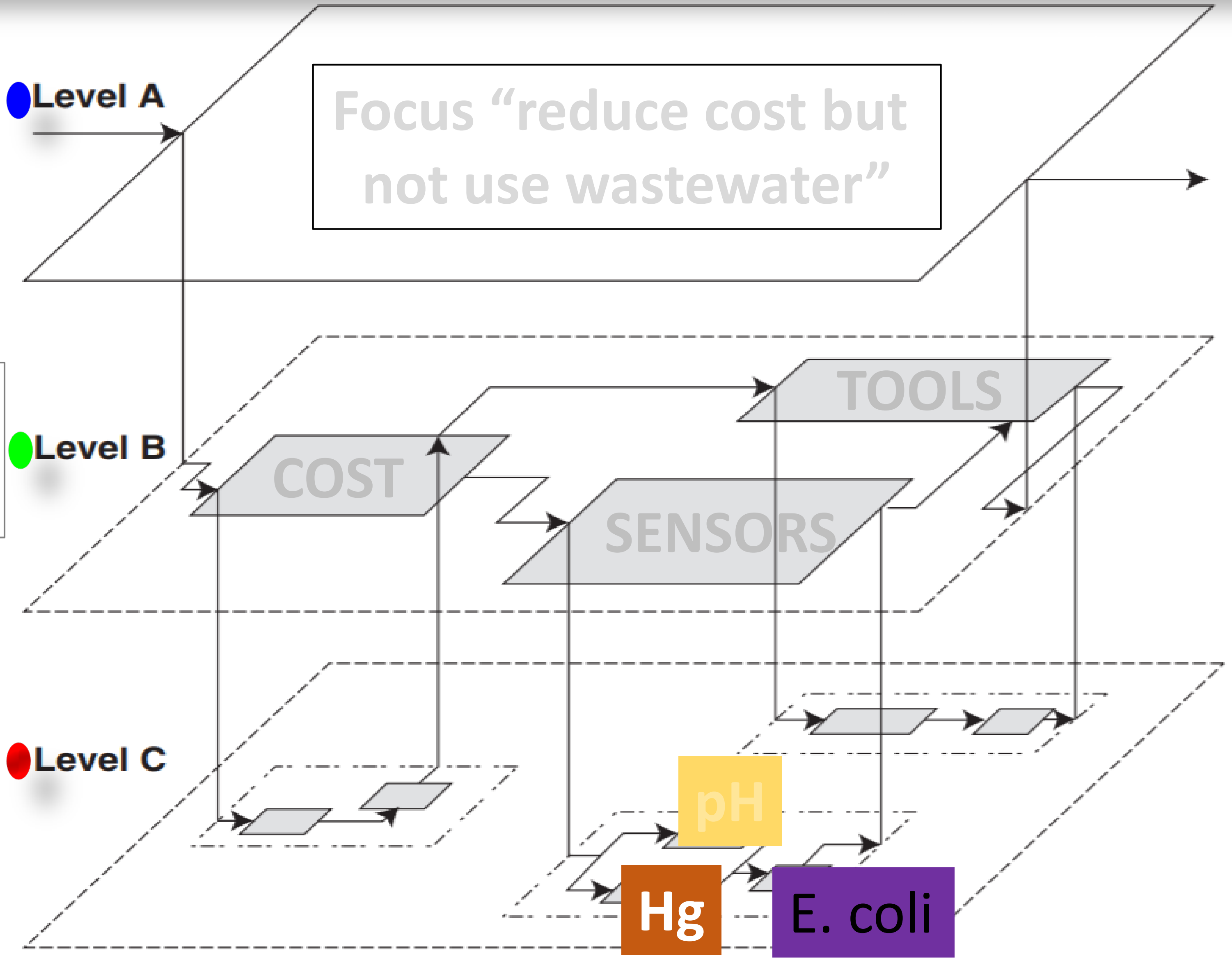
Denominator

Granularity of Deconstruction – Where is the data source?

Focus “”reduce cost but not use wastewater

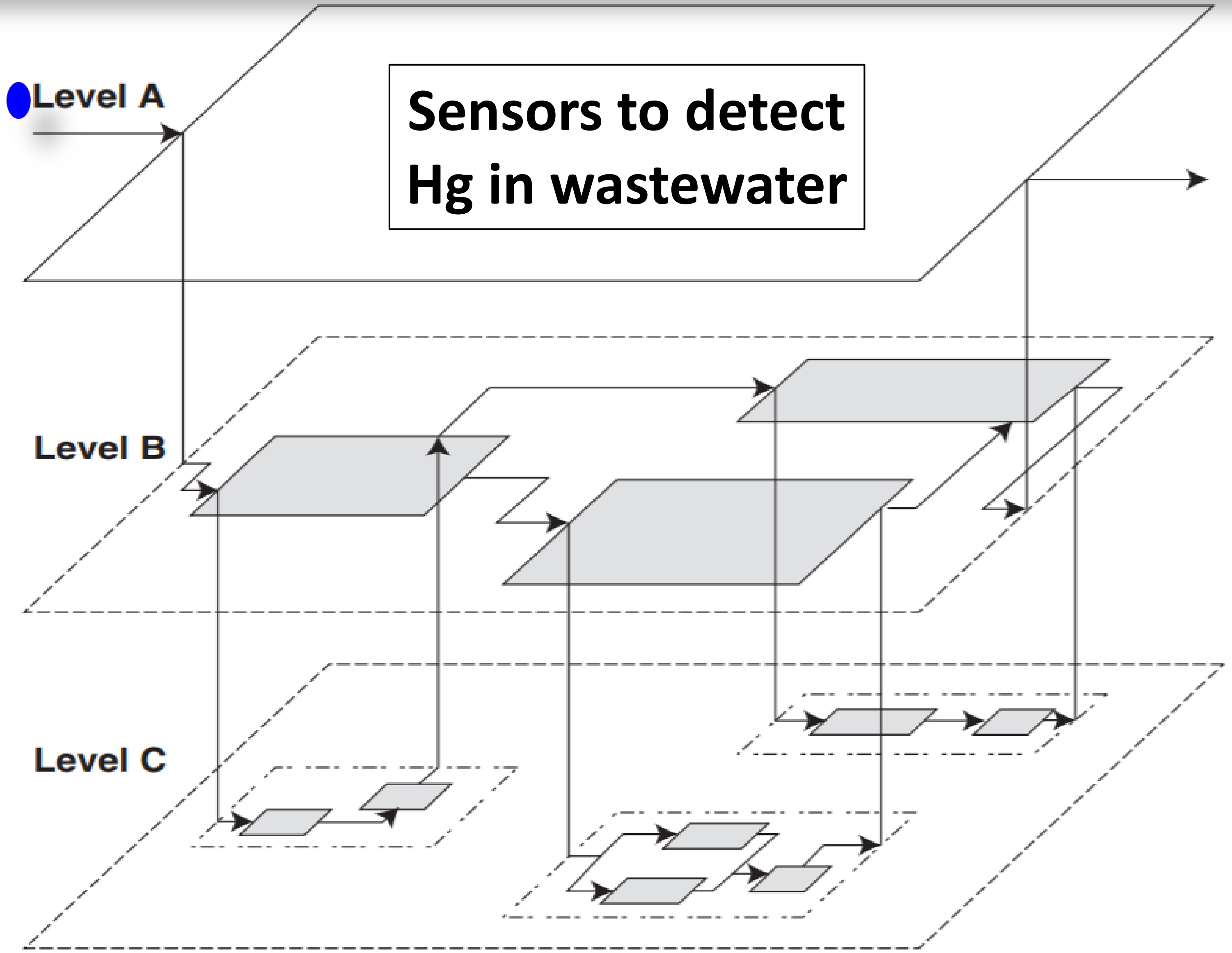
-Factors affecting cost
-Sensors to determine water quality
-Tools necessary for treatment of wastewater

- Sensors to determine water quality
  - Heavy metal pollution (Mercury)
  - Microbial contaminants (E. coli)
  - Chemical characteristics (pH)



Granularity of Deconstruction – Sensors to Detect Mercury

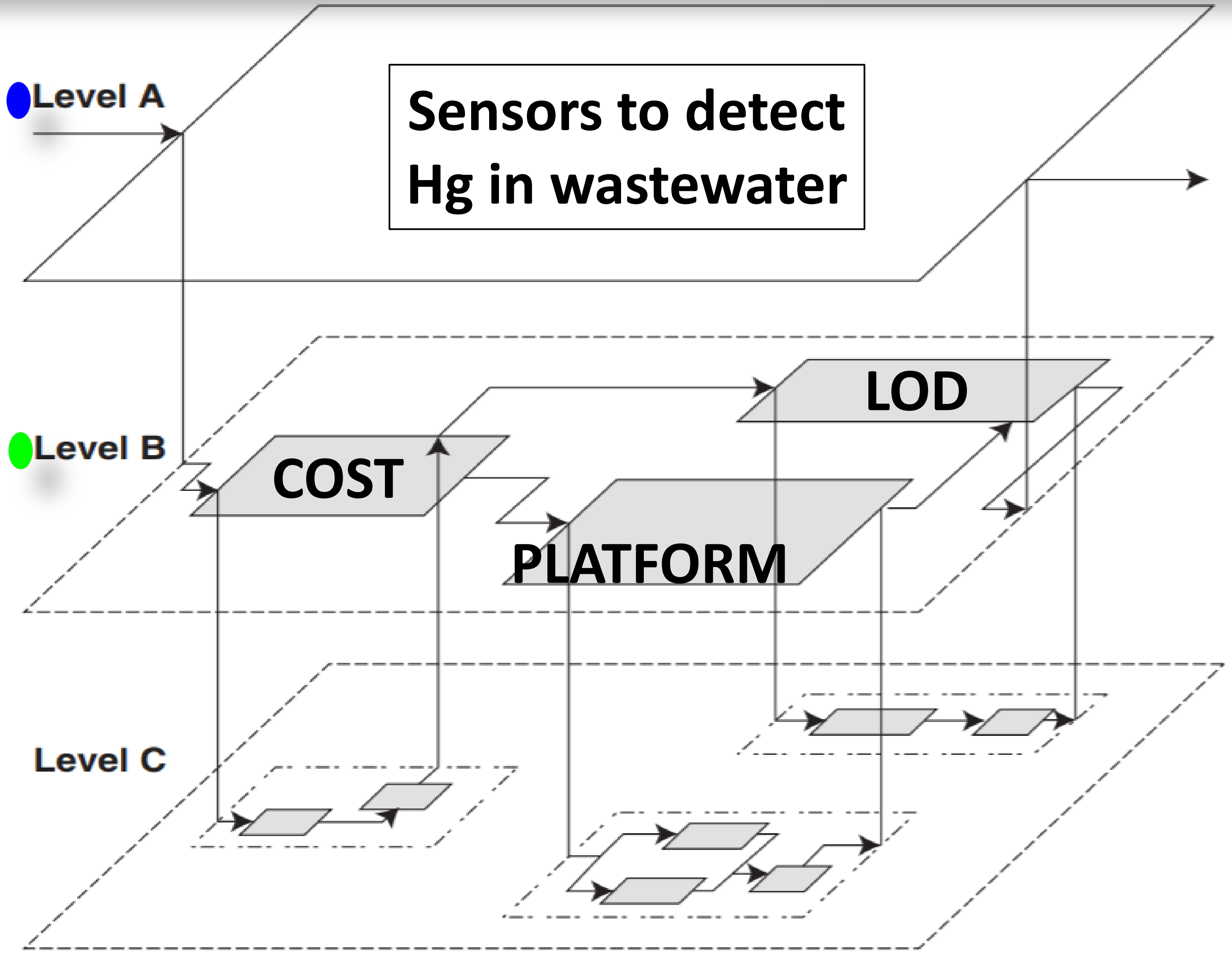
Sensors to detect Mercury (Hg) in wastewater



Granularity of Deconstruction – Sensors to Detect Mercury

Sensors to detect Mercury (Hg) in wastewater

-Cost of sensors
-Limit of Detection (LOD) necessary for use
-Which platform is best for the specific use

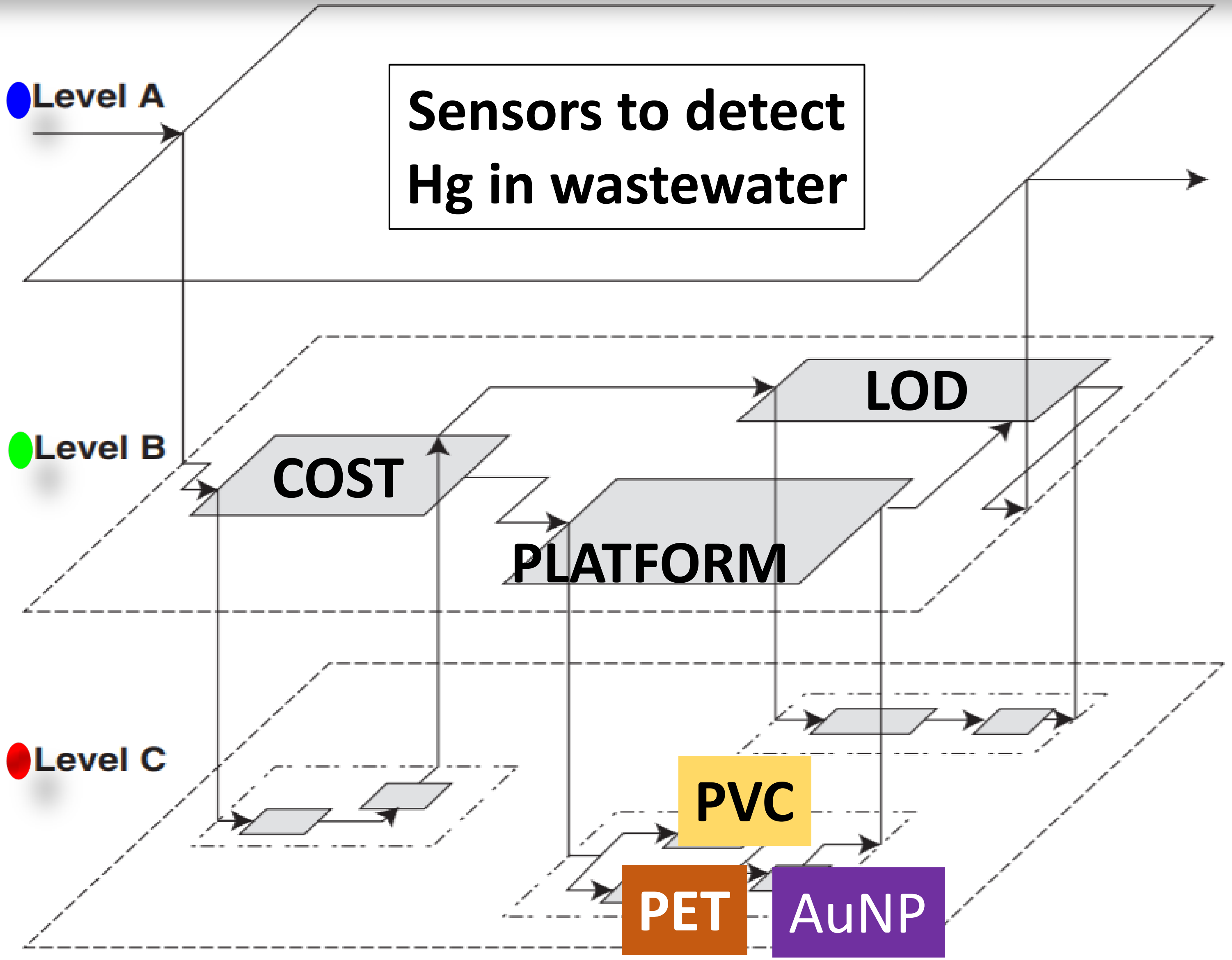


Granularity of Deconstruction – Sensors to Detect Mercury

Sensors to detect Mercury (Hg) in wastewater

-Cost of sensors
-Limit of Detection (LOD) necessary for use
-Which platform is best for the specific use

- Platforms for Mercury (Hg) sensor
  - Poly(vinyl chloride) (PVC) membrane
  - Photoinduced electron transfer (PET)
  - Gold-nanoparticle (AuNP)



https://onlinelibrary.wiley.com/doi/abs/10.1002/elan.1140090609
https://pubs.acs.org/doi/10.1021/ja037995g#
https://pubs.acs.org/doi/10.1021/nl0100116

Granularity of Deconstruction: Several types of sensors to detect Mercury

A	B	C	D	E	F	G	H	I	J
Device number	**MW [Da]**	**Category**	**Target**	**Recognition-transduction scheme**	**Platform**	**Range (LOD)**	**Range (max tested)**	**Selectivity (interferent species tested)**	**Response time [sec]**
1	201	heavy metal	mercury (Hg2+)	AuNP		1.00E-08	NR	NR	60
2	201	heavy metal	mercury (Hg2+)	MIP??	Sol gel	5.00E-06	NR	NR	600
3	201	heavy metal	mercury (Hg2+)	Rhodamine		1.00E-07	NR	NR	60
4	201	heavy metal	mercury (Hg2+)			5.00E-07	NR	95%	60
5	201	heavy metal	mercury (Hg2+)	foldamer	micelle	5.00E-07	NR	99%	60
6	201	heavy metal	mercury (Hg2+)	corroloe derivative	PVC	5.60E-06	NR	NR	300
7	201	heavy metal	mercury (Hg2+)	tetraarylborate		3.00E-07	NR	NR	60
8	201	heavy metal	mercury (Hg2+)			1.00E-07	NR	poor over Ag+	60
9	201	heavy metal	mercury (Hg2+)	polythiophene		3.00E-05	NR	90%	**60**
10	201	heavy metal	mercury (Hg2+)	thiosemicarbazone		5.00E-06	NR	NR	60
11	201	heavy metal	mercury (Hg2+)	dansylcarboxamide		1.00E-05	5.00E-04	NR	60
12	201	heavy metal	mercury (Hg2+)	quenching		3.00E-06	5.50E-05	excellent	60
13	201	heavy metal	mercury (Hg2+)	DNAzyme		2.40E-09	NR	excellent (transition/heavy metals)	60
14	201	heavy metal	mercury (Hg2+)	chromo-ionophore assembly	PVC	3.40E-08	NR	poor (heavy metals)	60
15	201	heavy metal	mercury (Hg2+)	AuNP		5.00E-09	1.00E-05	excellent (transition/heavy metals)	600
16	201	heavy metal	mercury (Hg2+)			1.00E-08	2.00E-04	excellent (transition/heavy metals)	60
17	201	heavy metal	mercury (Hg2+)	Rhodamine 6G	AuNP	6.00E-11	3.60E-08	excellent (transition/heavy metals)	60
18	201	heavy metal	mercury (Hg2+)	Cholic acid		5.00E-08	NR	good (MeHg/transition/heavy metals)	60
19	201	heavy metal	mercury (Hg2+)	thiacalixarene		2.00E-06	8.50E-06	good (poor over Ag+)	60
20	201	heavy metal	mercury (Hg2+)			7.00E-07	NR	poor over Cu+	60
21	201	heavy metal	mercury (Hg2+)	anthraquinone/urea		5.0E-05	2.0E-04	poor	60
22	201	heavy metal	mercury (Hg2+)	anthracene/ionophore hybrid	PET	1.0E-06		poor over Fe3+	60
23	201	heavy metal	mercury (Hg2+)	oligonucleotide	AuNP	1.0E-07	1.0E-06	poor over Pb3+	60
24	201	heavy metal	mercury (Hg2+)	oligonucleotide		4.2E-08	6.7E-07	moderate	60
25	201	heavy metal	mercury (Hg2+)			5.0E-08		excellent (transition/heavy metals)	60
26	201	heavy metal	mercury (Hg2+)	phosphorescent iridium(III) complex		2.0E-05		excellent (transition/heavy metals)	60
27	201	heavy metal	mercury (Hg2+)	MerR protein		1.0E-08		NR	60
28	201	heavy metal	mercury (Hg2+)			1.0E-06		NR	60

Which sensor to choose? Which sensor has the lowest limit of detection?

Which sensor to choose? Which sensor has the lowest limit of detection?

Users wish to explore sensor categories and attributes ?

Which sensor to choose? Which sensor has the lowest limit of detection?

Users wish to explore sensor categories and attributes ?

End-users, as well as experts, may benefit from information about different sensors, by categories and list of attributes, which may be suitable for use.

Which sensor to choose? Which sensor has the lowest limit of detection?

SENsor SEarch Engine

Which sensor to choose? Which sensor has the lowest limit of detection?

SENsor **SE**arch **E**ngine

SENSEE

Delving deeper into the granularity of the data necessary for DIDA'S to be sufficiently data-informed, we arrive at one data source:

SENSORS
for
DIDA'S

Sensor data as a source of data for data-informed decision as a service (DIDA'S)

Sensor data still remains a key denominator when we move from data-informed (DIDA'S) to knowledge-informed (KIDS)

SENSORS
for
DIDA'S KIDS

Sensors for knowledge-informed decision as a service (KIDS)

SKIDS

In granular terms, DIDA’S and KIDS, still needs to choose sensor type.

SENsor **SE**arch **E**ngine

SENSEE

In granular terms, the outcome from DIDA'S and KIDS, depends on data.

Data from Sensors

SENSEE

At the most granular level, first we need to *choose the sensor* and then proceed to harvest *data* from specific sensor(s).

FOR SOME OF THE QUESTIONS, THIS IS A PRE-REQUISITE FOR DIDA'S and KIDS.

Hence, we start searching for suitable sensor categories and attributes.

SENsor SEarch Engine

SENSEE

Then, we seek data from sensors (relevant to the real world questions).

Data from Sensors

SENSEE

Before we can think about information or digital design

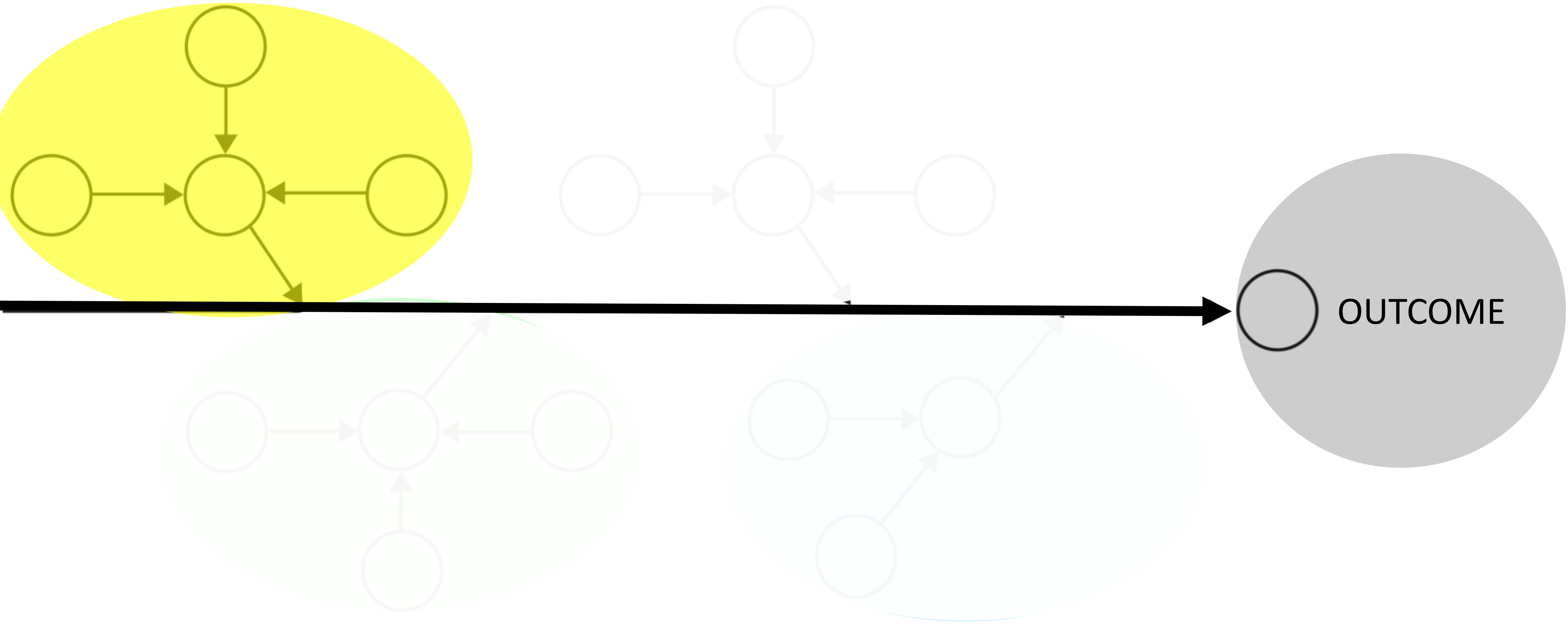
We need sensors, and sensor data, to fuel the outcome.

Data-informed (DIDA'S)

Knowledge-informed (KIDS)

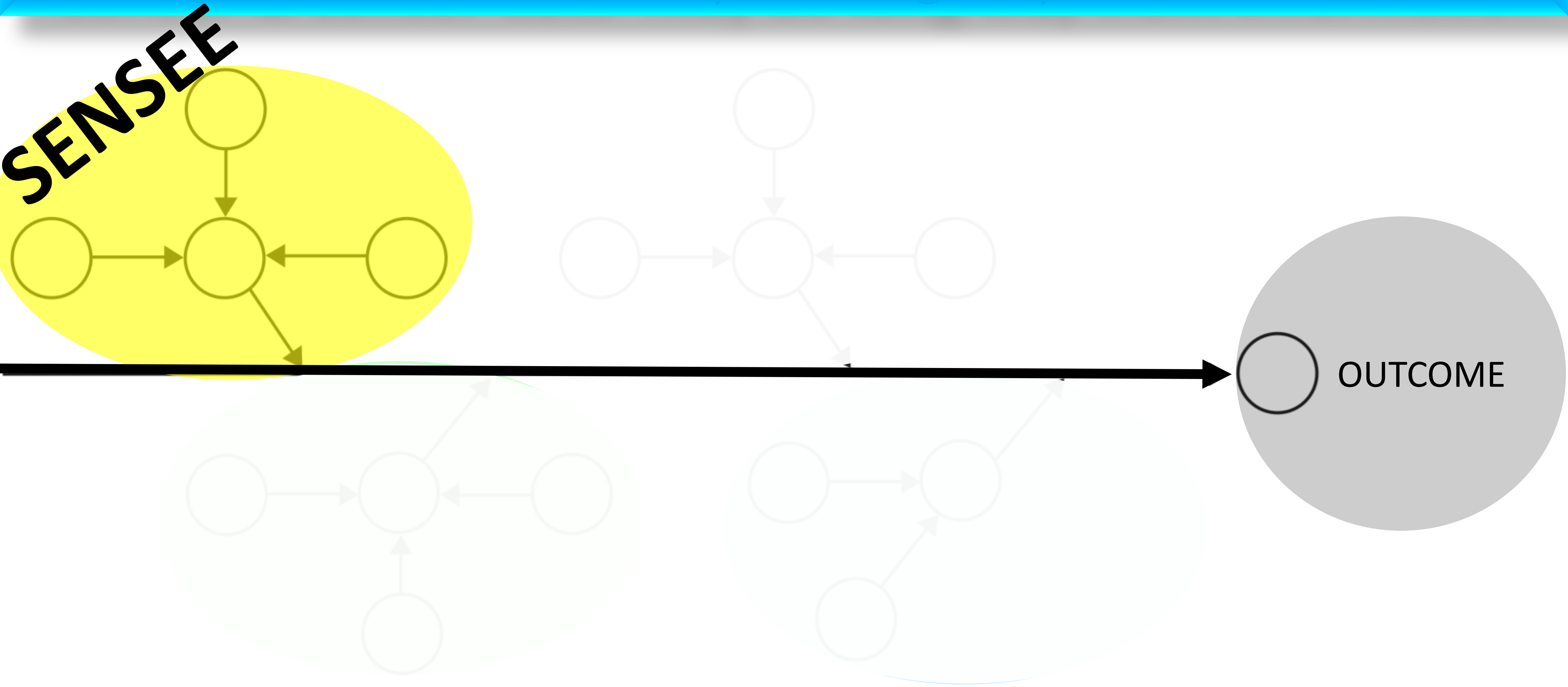
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Sensor & sensor data



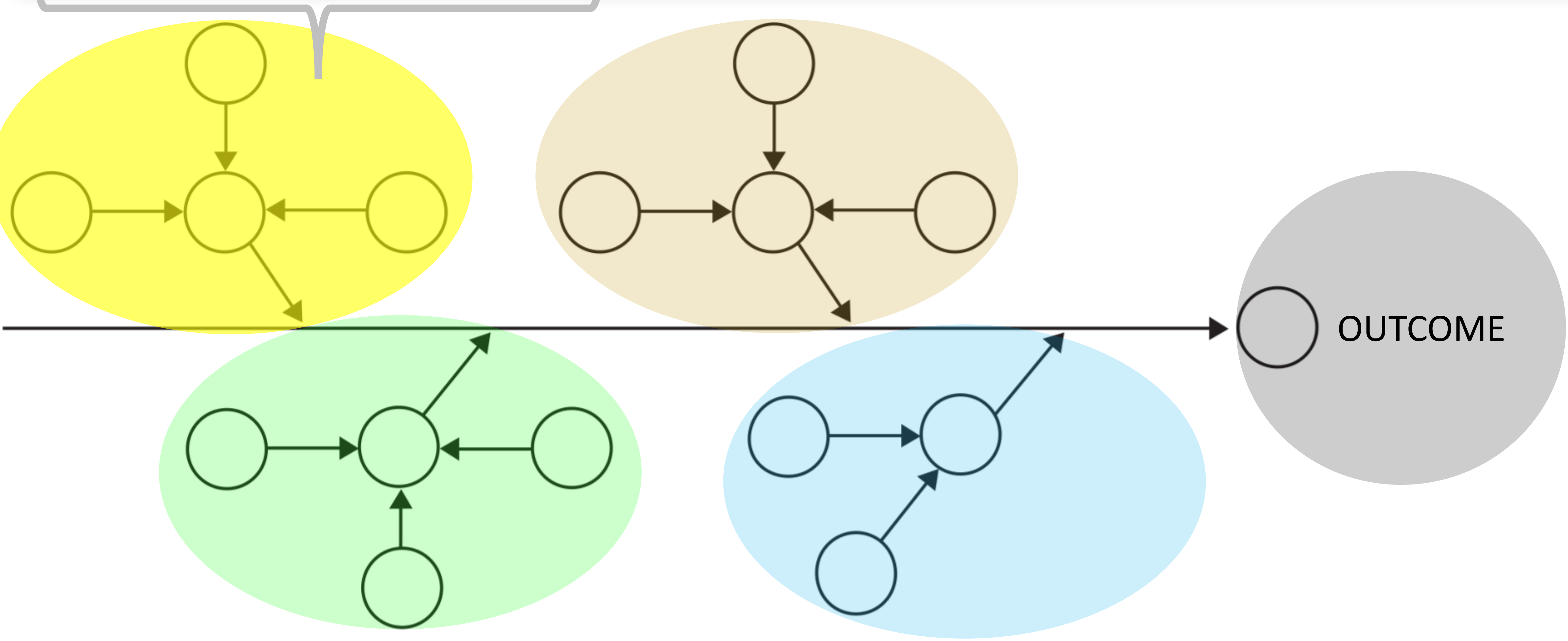
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Sensor & sensor data



How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

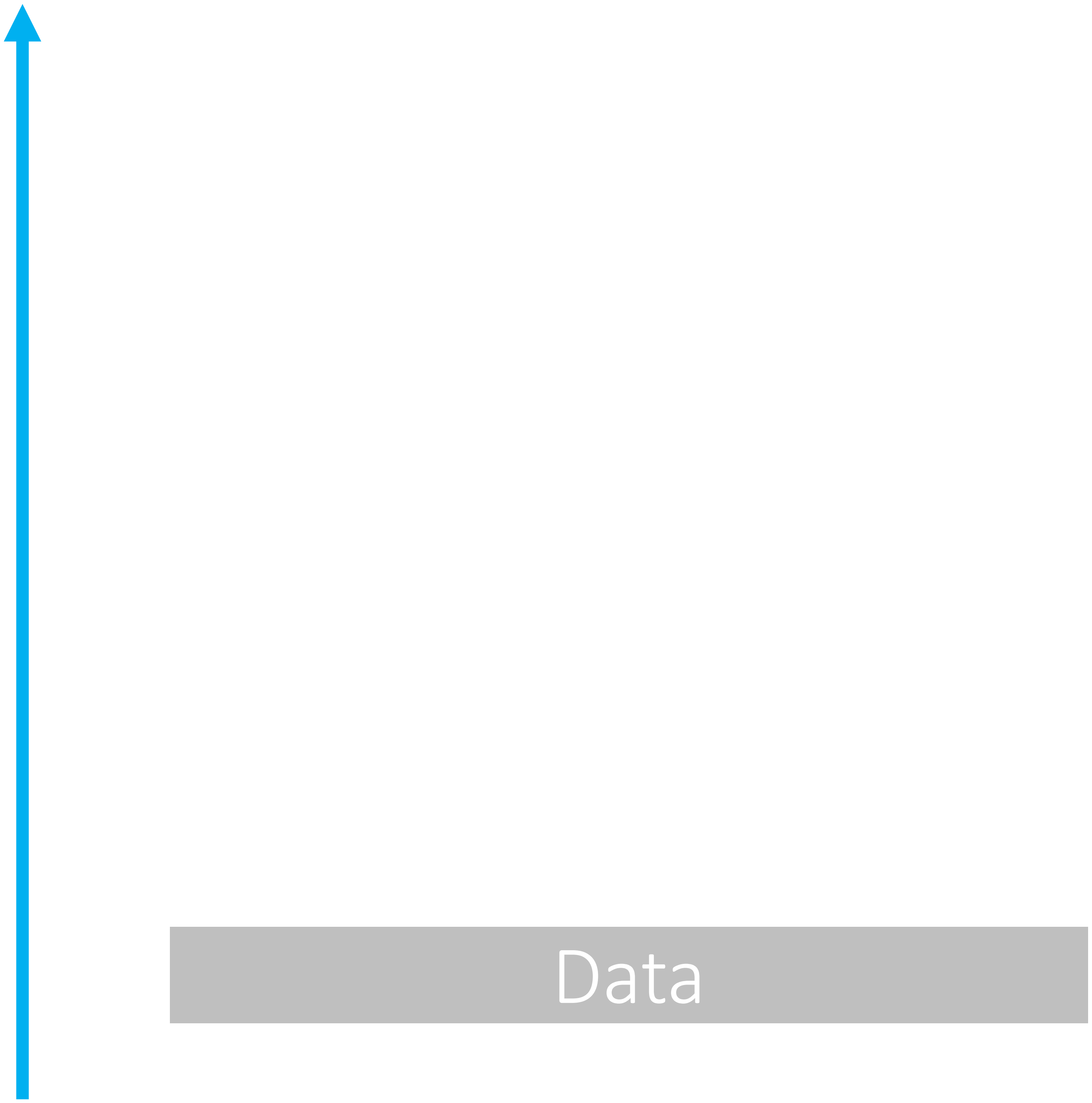
Sensor & sensor data may be long way from information



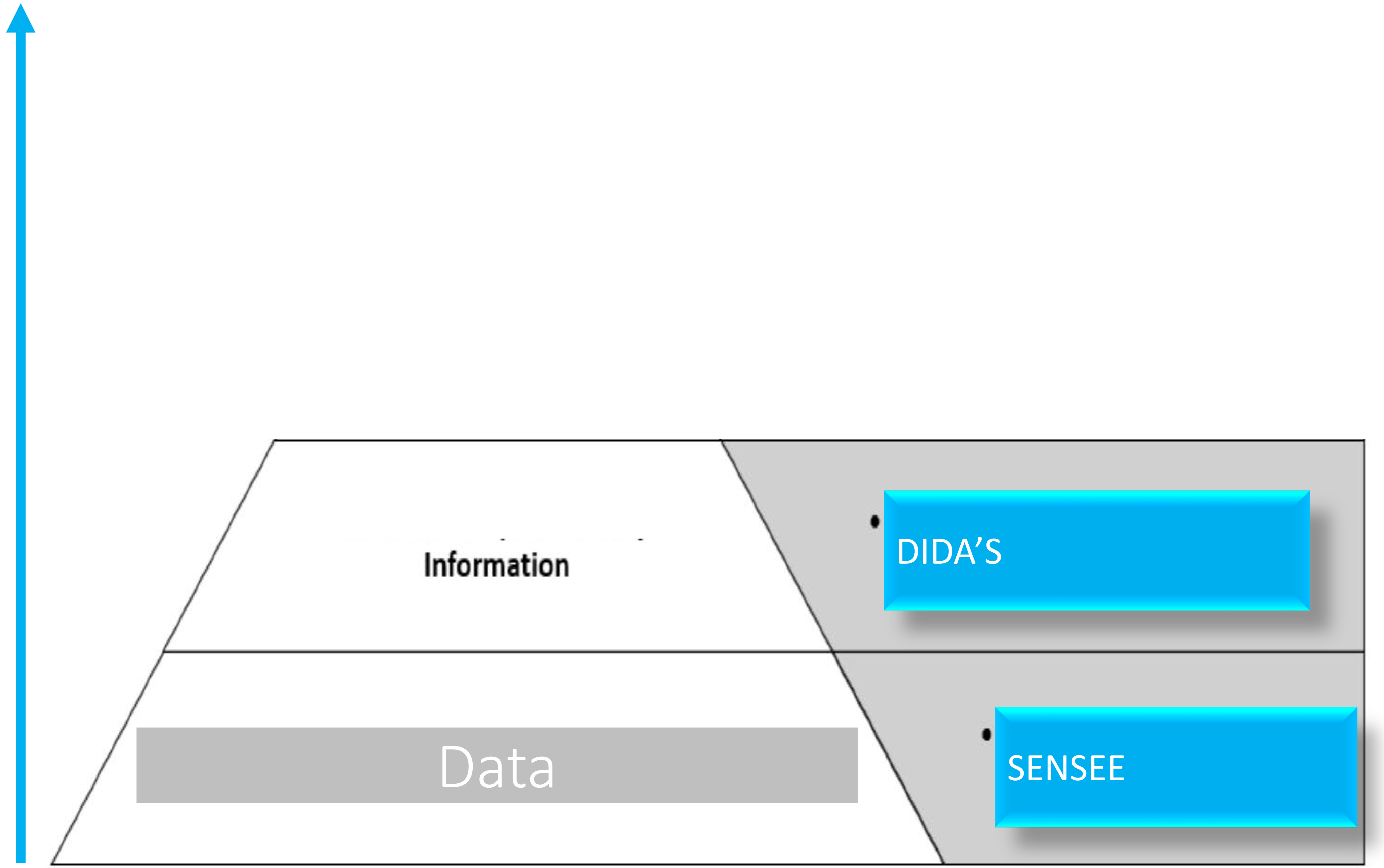
http://www.ihi.org/education/IHIOpenSchool/resources/Assets/CauseandEffect_Instructions.pdf

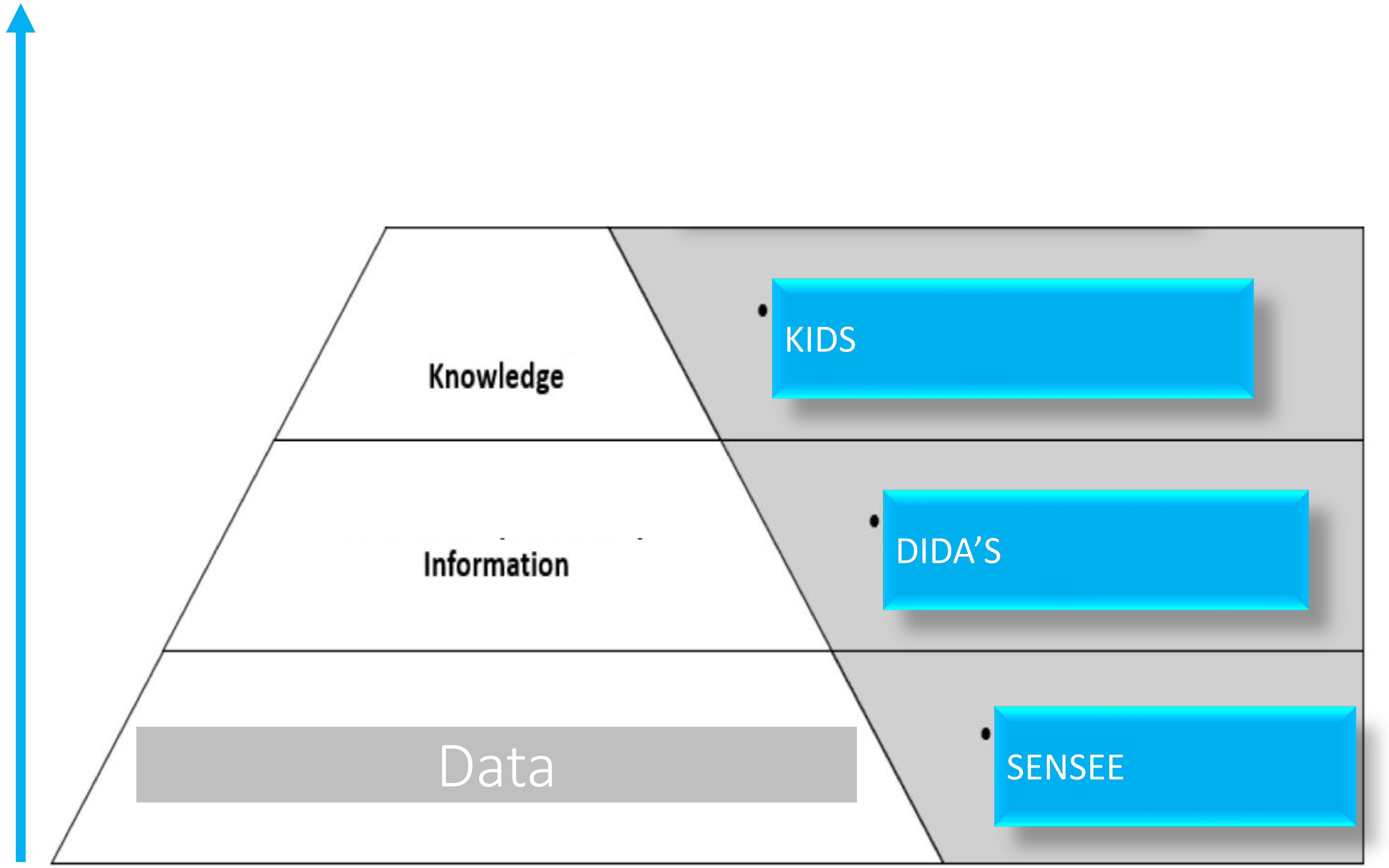
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

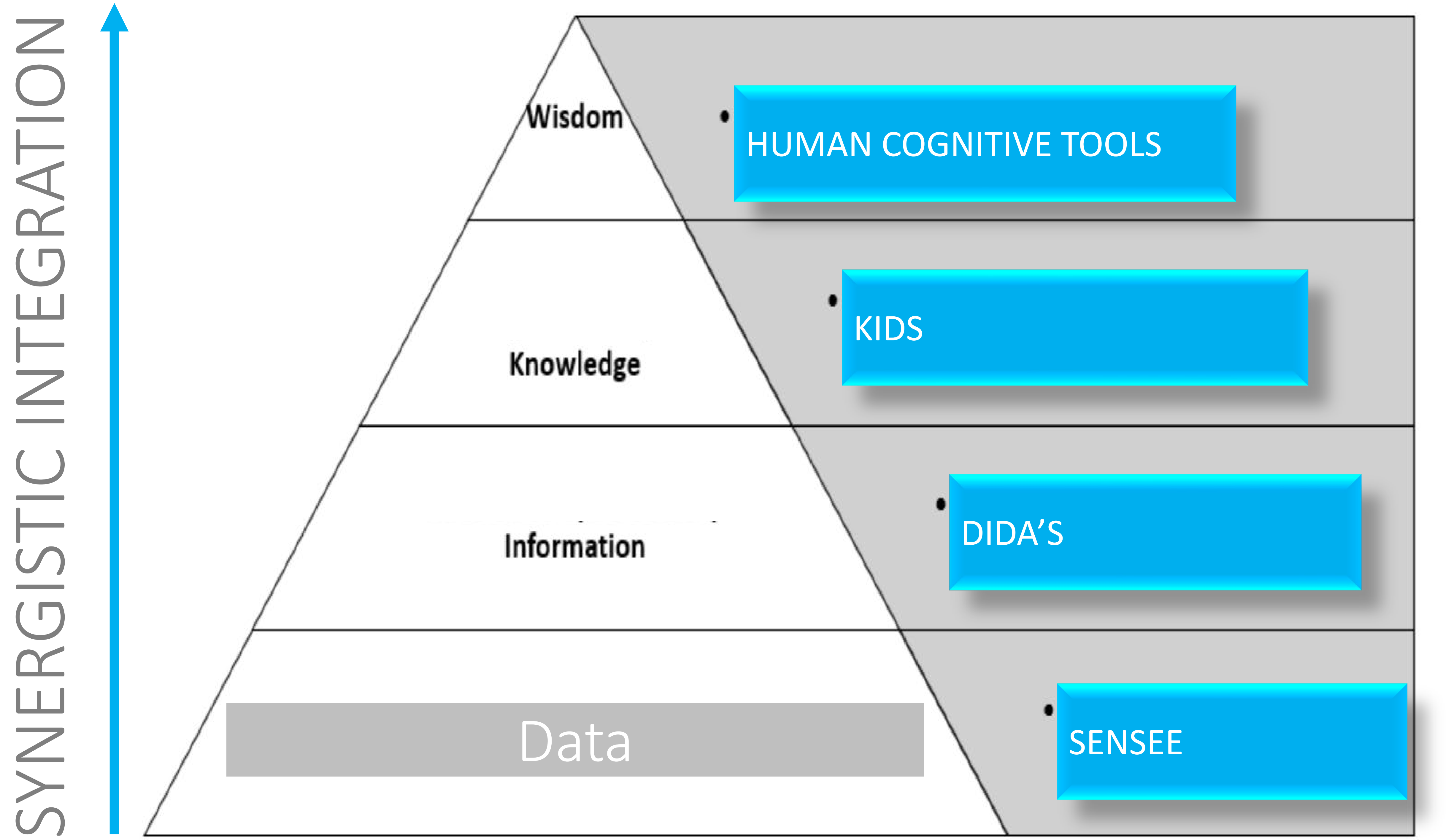
To summarize the steps

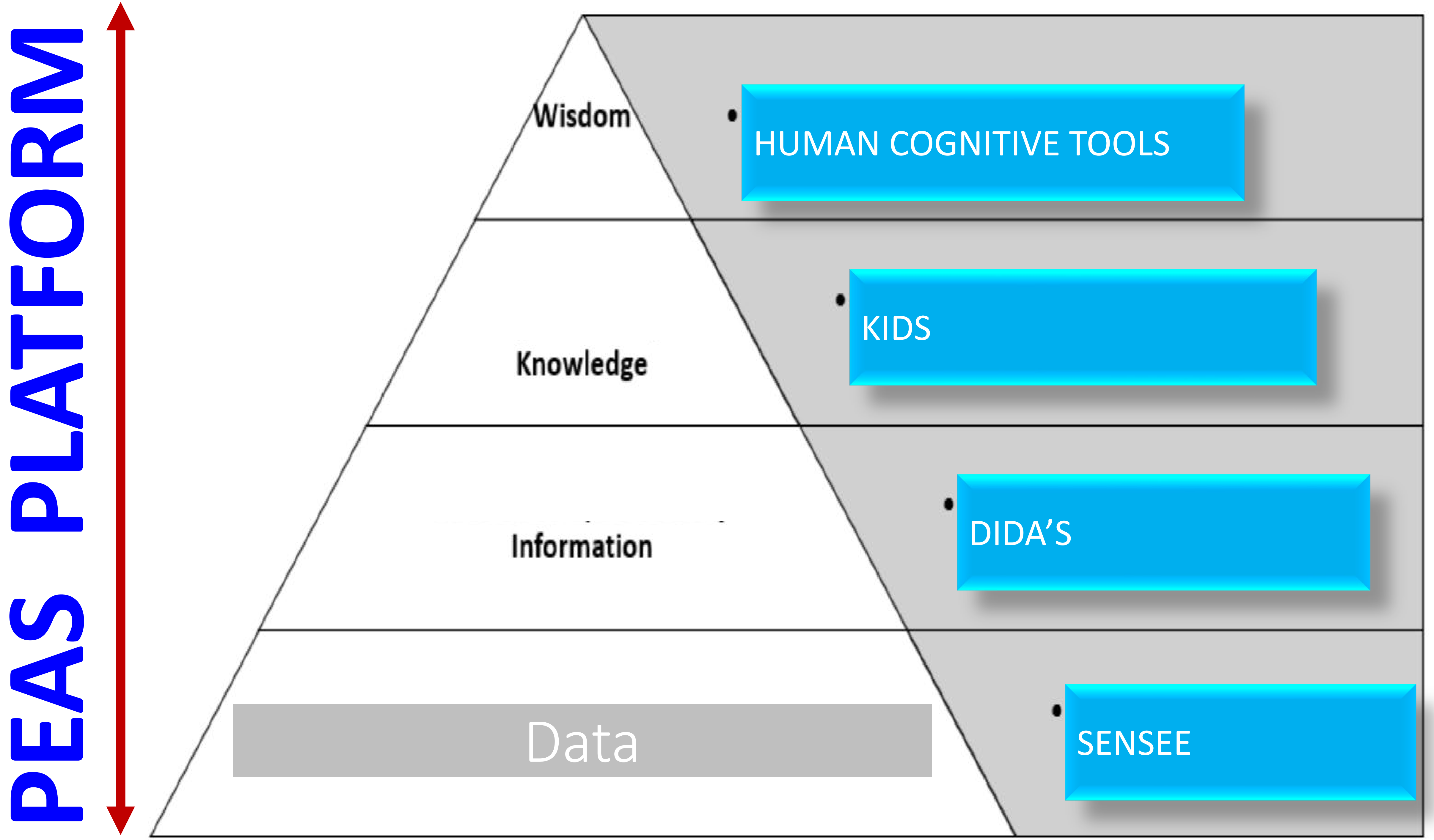

Data


Data
SENSEE


Information
Data
DIDA’S
SENSEE


Knowledge
Information
Data
KIDS
DIDA’S
SENSEE


SYNERGISTIC INTEGRATION
Wisdom
Knowledge
Information
Data
HUMAN COGNITIVE TOOLS
KIDS
DIDA'S
SENSEE


PEAS PLATFORM
Wisdom
Knowledge
Information
Data
HUMAN COGNITIVE TOOLS
KIDS
DIDA'S
SENSEE

Digital Transformation

is about the life cycle of data as it transforms to information and contributes to better decisions

Are we immersed in **data** swamps?
Actions depend on **information**.
Informed by **knowledge**.
Learn from **experience**.

Unbeknownst to us, we are in a perpetual quest for knowledge. Every day, in every action, we undertake the journey from data and information to knowledge. In the process we are learning, in each step, from our experiences, no matter how small and agnostic of the scenarios (social actions, academic activities, business pursuits, ideas and opinions).

Convergence

every step of the journey

Social Domain

Situation

Time constraint

Social Domain

Situation **Time constraint**

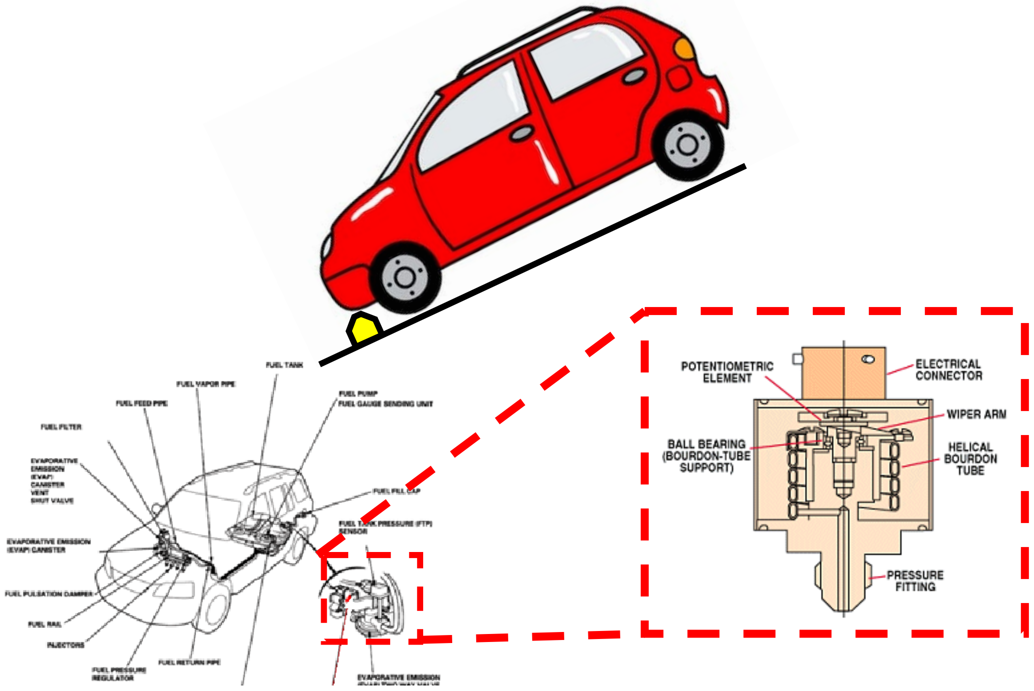
Engineering Domain

Sensor reading *Local knowledge*

Social Domain

Situation **Time constraint**

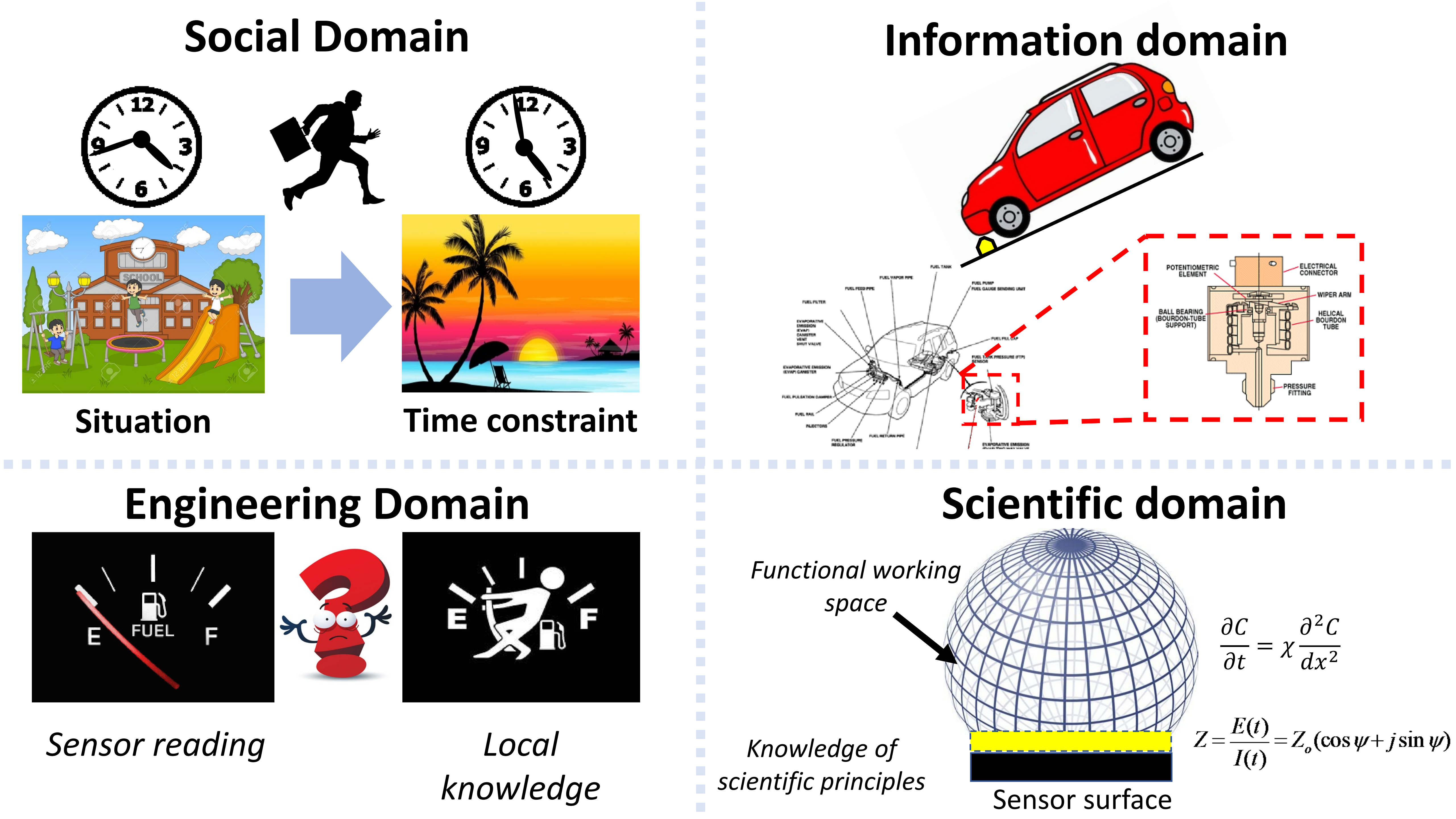
Information domain



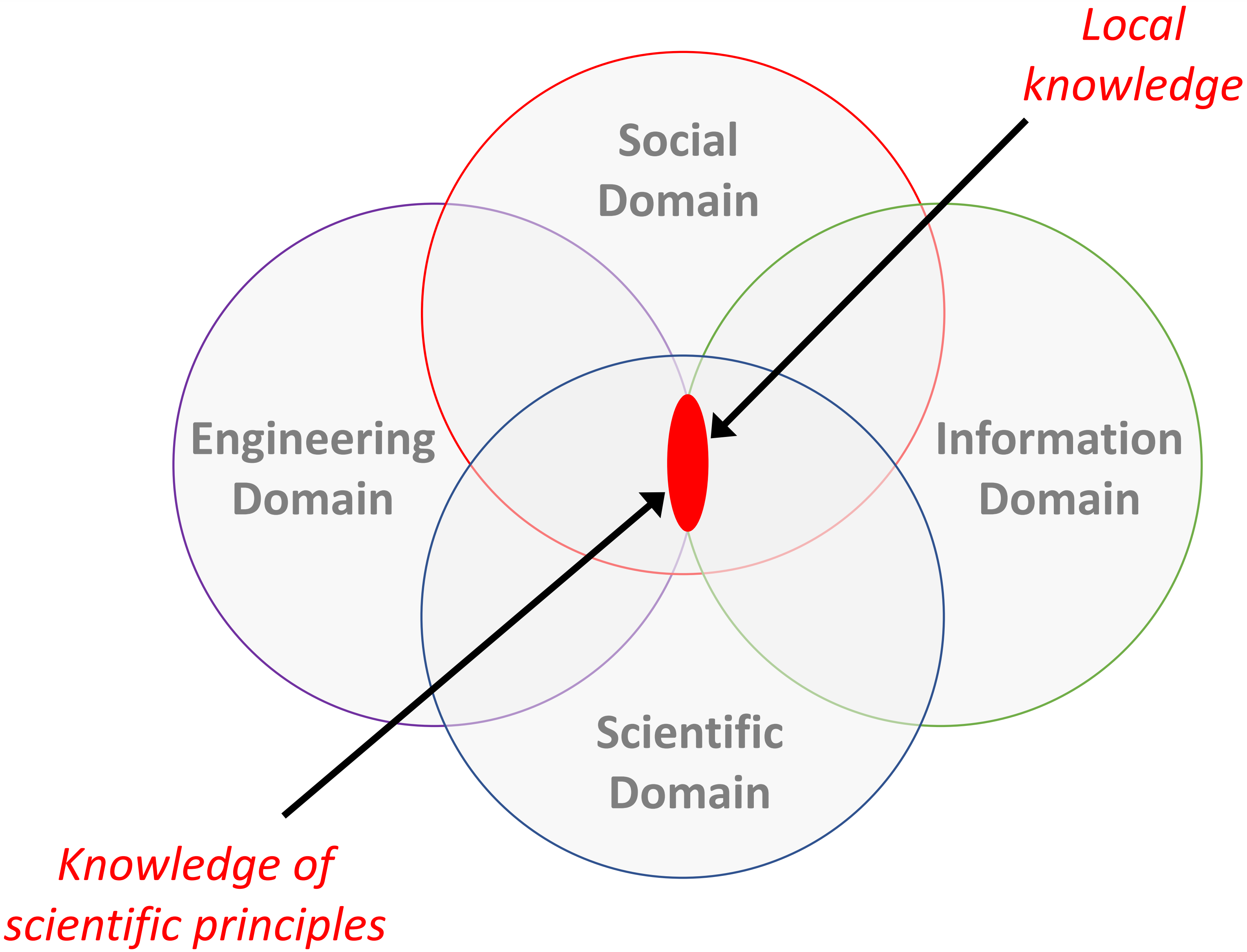
Engineering Domain

Sensor reading

Local knowledge


Social Domain
12
3
6
9
Situation
Time constraint
Information domain
POTENTIOMETRIC ELEMENT
ELECTRICAL CONNECTOR
WIPER ARM
BALL BEARING (BOURDON-TUBE SUPPORT)
HELICAL BOURDON TUBE
PRESSURE FITTING
Engineering Domain
E
FUEL
F
E
F
Sensor reading
Local knowledge
Scientific domain
Functional working space
$\frac{\partial C}{\partial t} = \chi \frac{\partial^2 C}{dx^2}$
Knowledge of scientific principles
Sensor surface
$Z = \frac{E(t)}{I(t)} = Z_o(\cos\psi + j\sin\psi)$

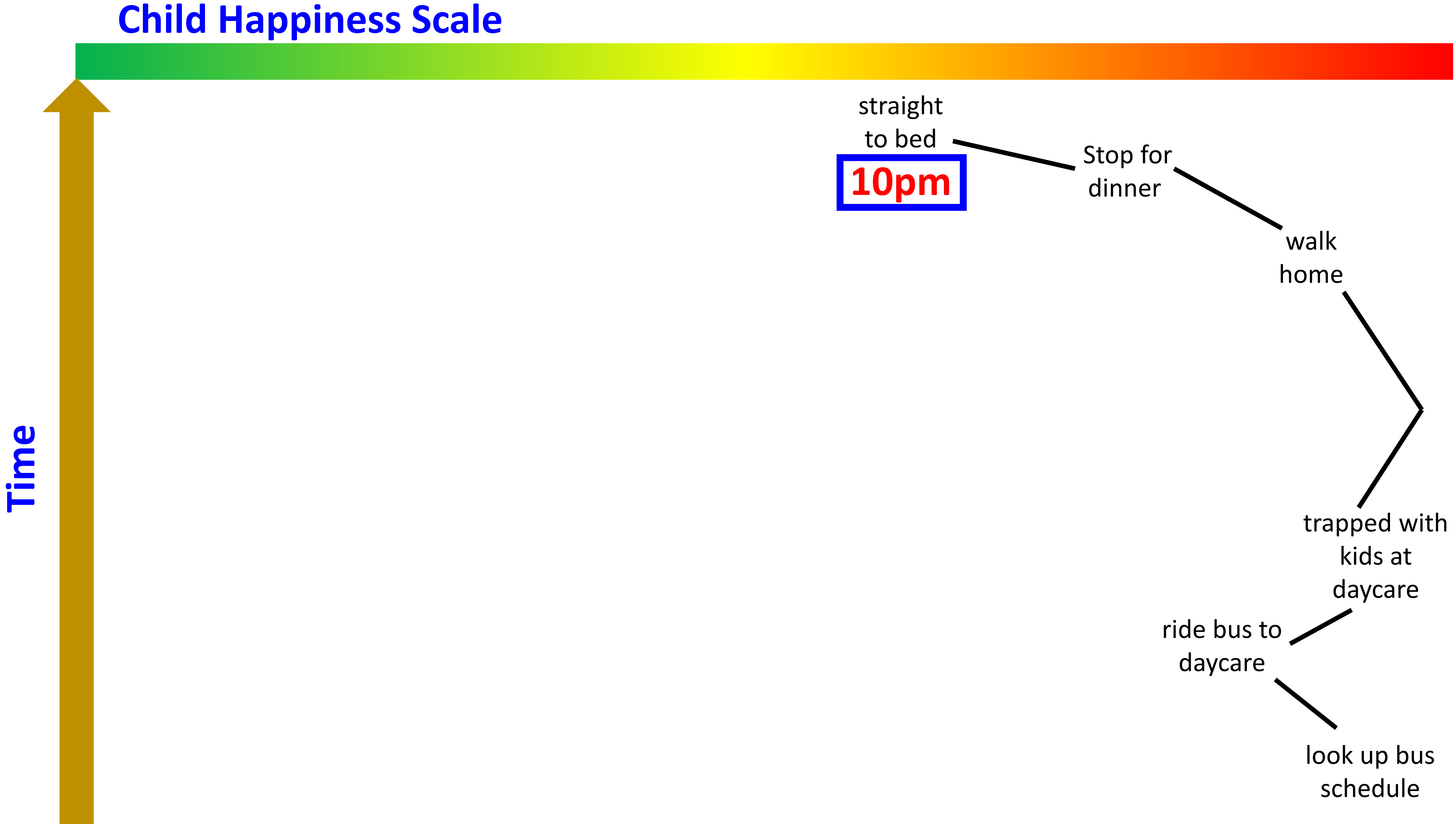
Convergence

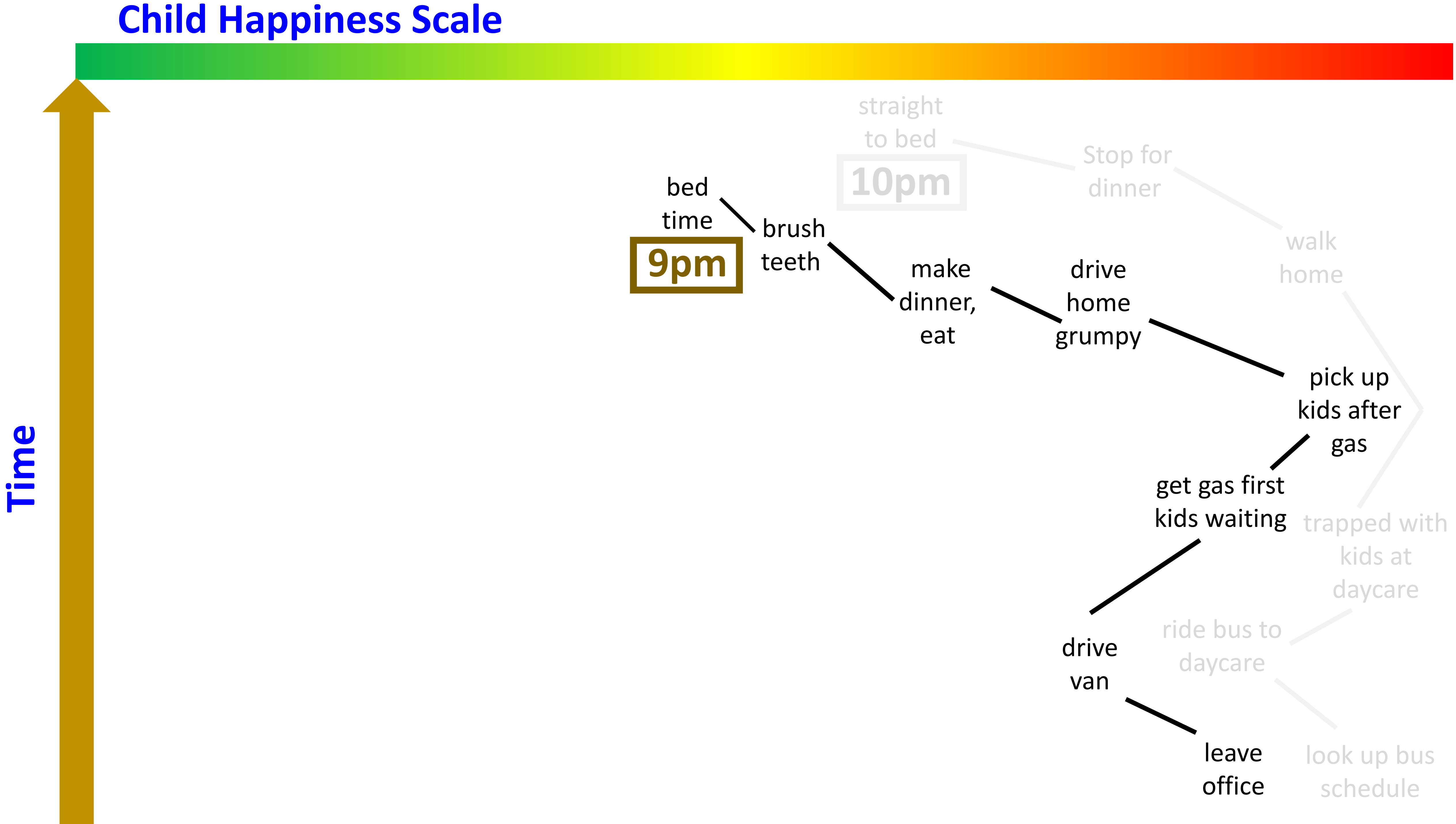


R2C2 – Think Convergence

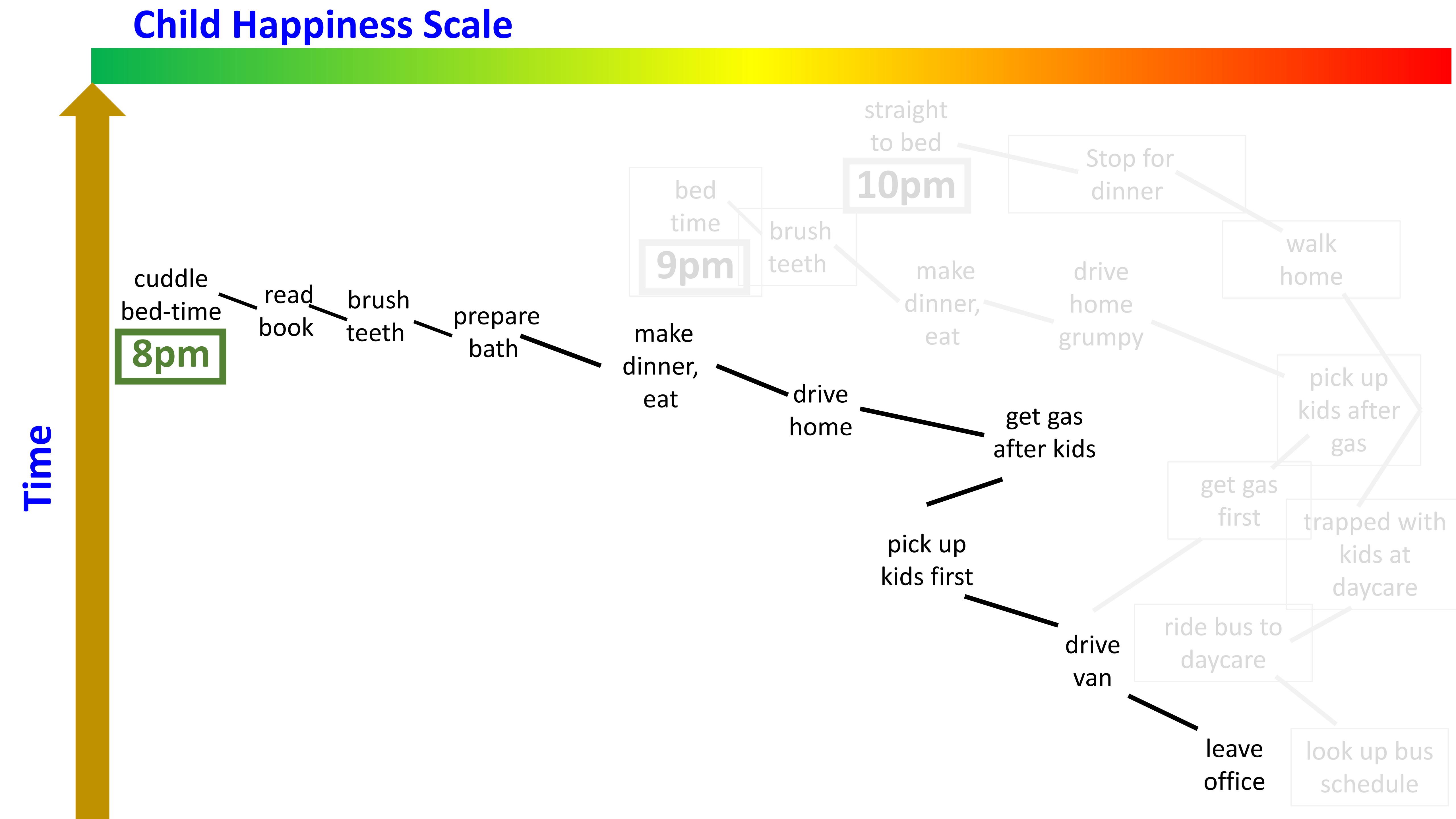
For humans, it is child's play

For systems, it's still a difficult task to select what is relevant, relative to the context, and connect (**R2C2**), distributed data, to extract information, to aid decisions or execute action for the situation, based on actionable knowledge.

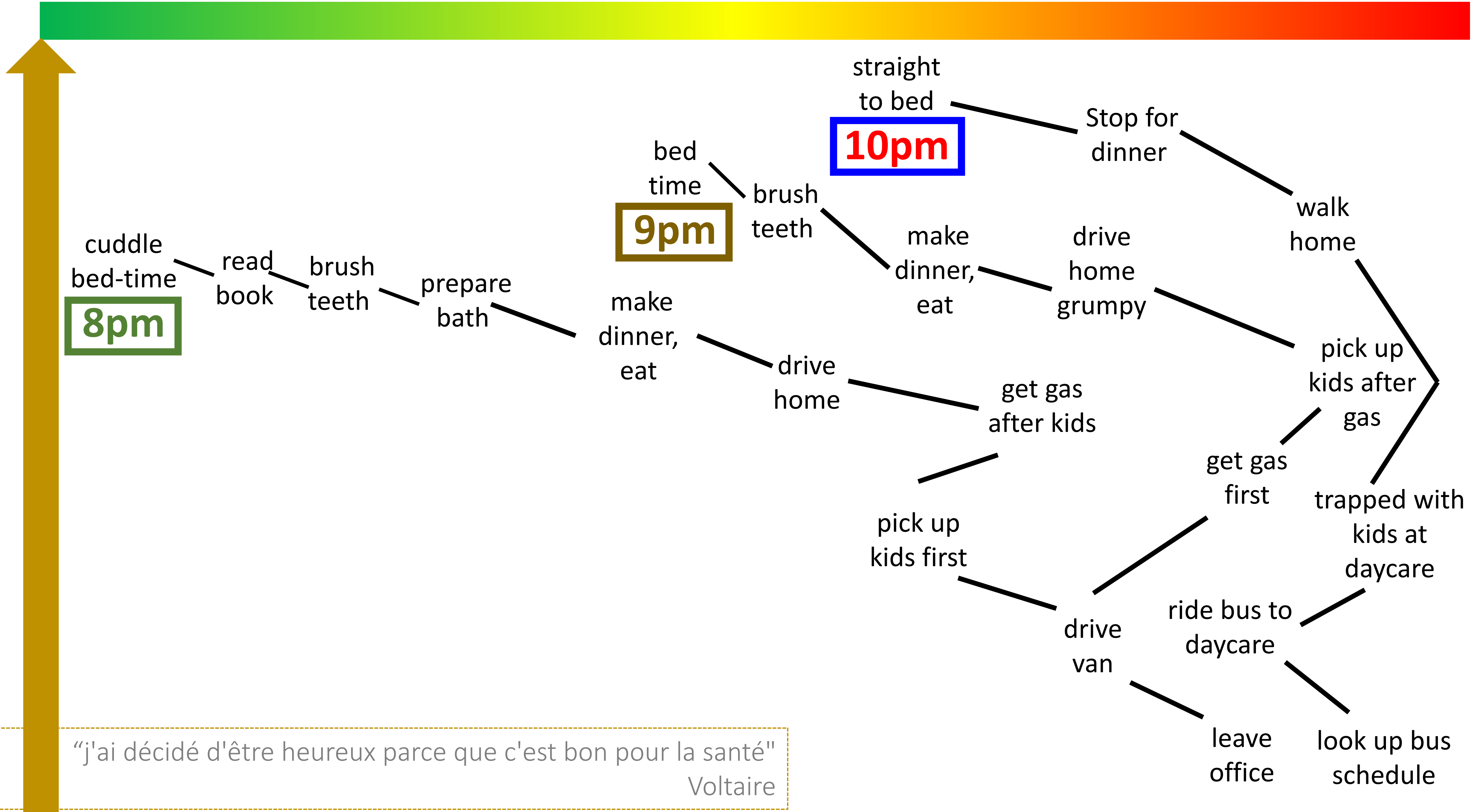

Child Happiness Scale
Time
straight to bed
10pm
Stop for dinner
walk home
trapped with kids at daycare
ride bus to daycare
look up bus schedule



Child Happiness Scale
Time
straight to bed
10pm
Stop for dinner
walk home
bed time
9pm
brush teeth
make dinner, eat
drive home grumpy
pick up kids after gas
get gas first kids waiting
trapped with kids at daycare
drive van
ride bus to daycare
leave office
look up bus schedule

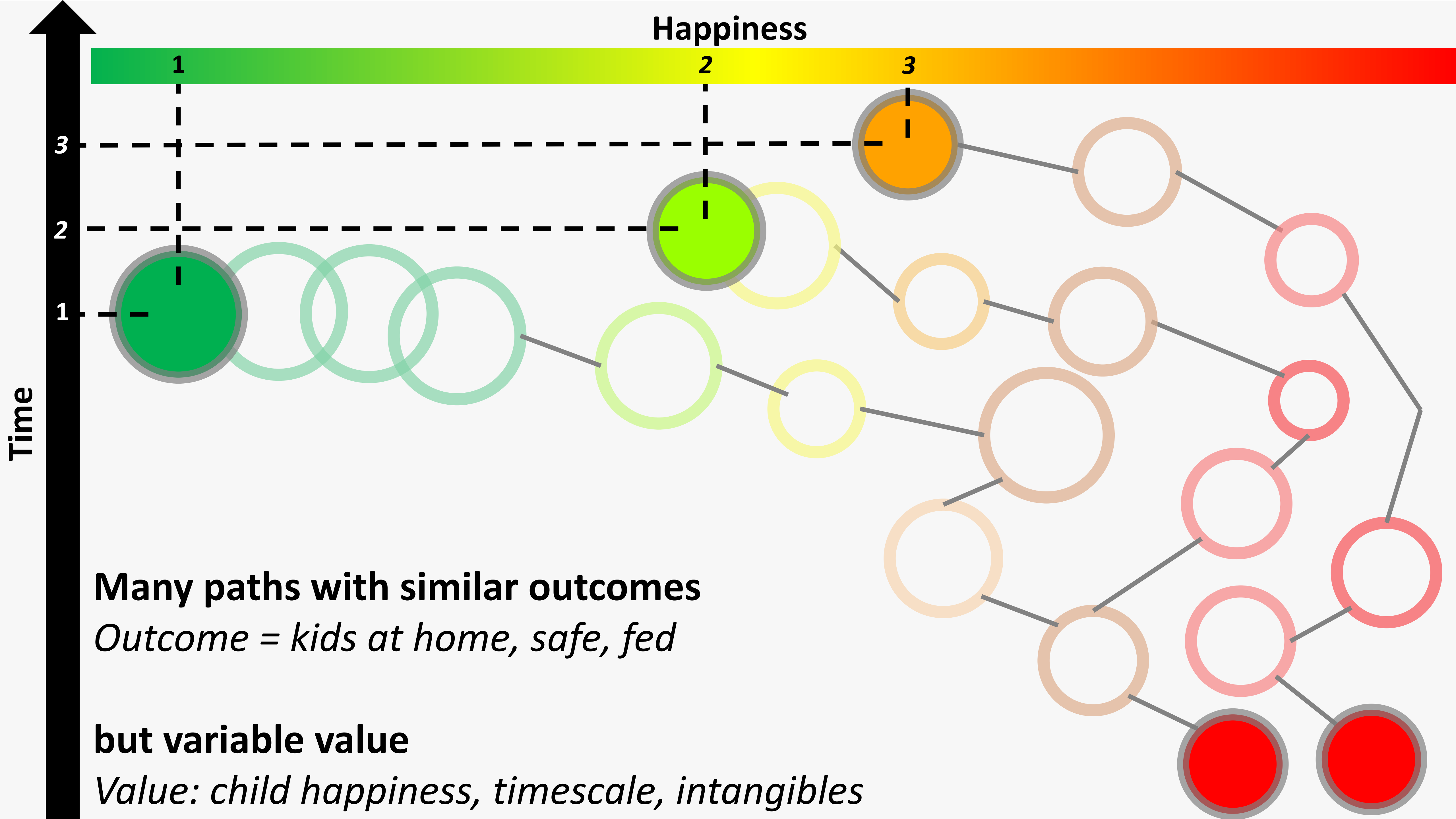

Child Happiness Scale
Time
cuddle bed-time
8pm
read book
brush teeth
prepare bath
make dinner, eat
drive home
get gas after kids
pick up kids first
drive van
leave office
bed time
9pm
brush teeth
straight to bed
10pm
Stop for dinner
walk home
make dinner, eat
drive home grumpy
pick up kids after gas
get gas first
trapped with kids at daycare
ride bus to daycare
look up bus schedule

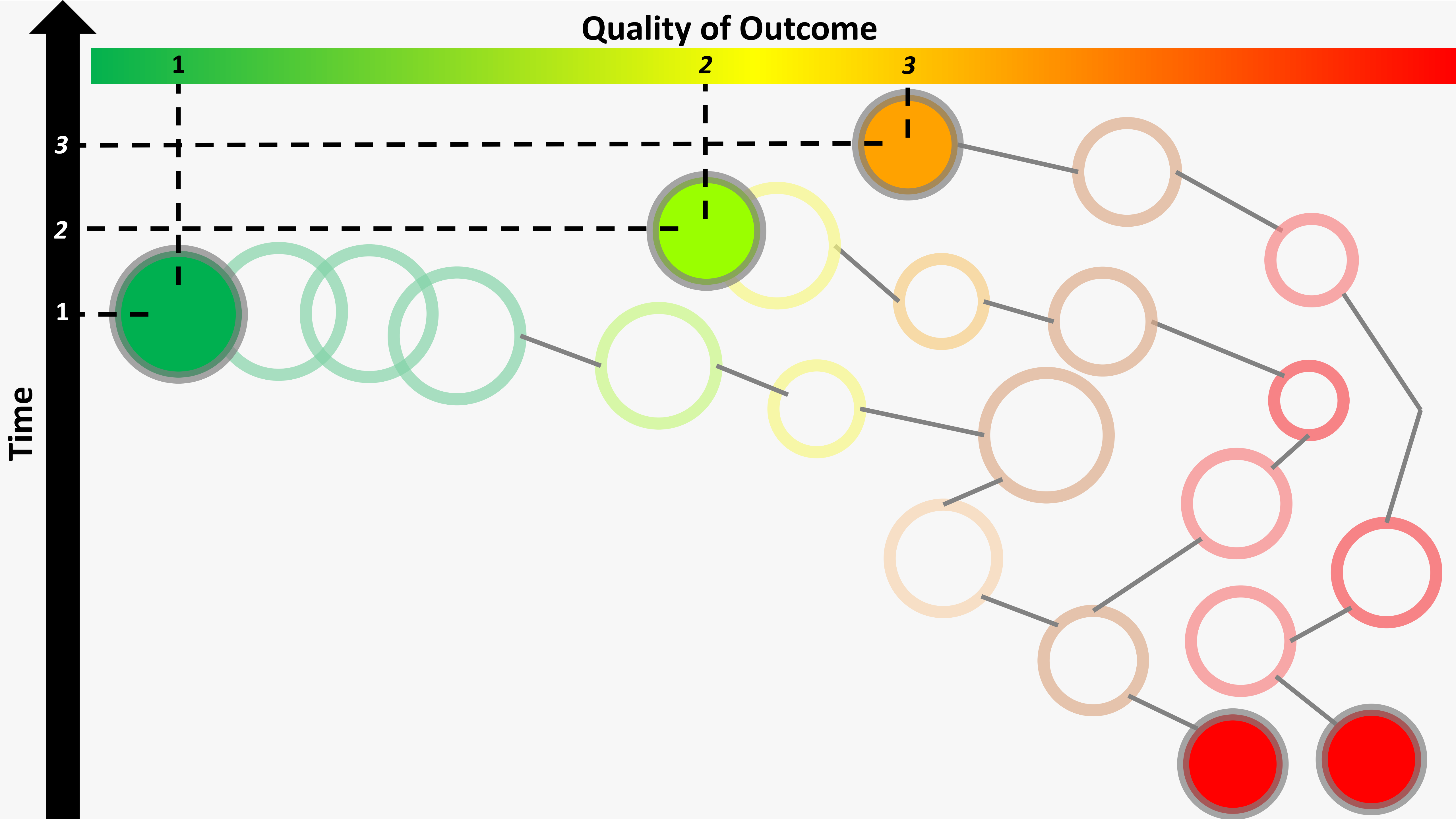
Child Happiness Scale (clap along if you feel like happiness is the truth)

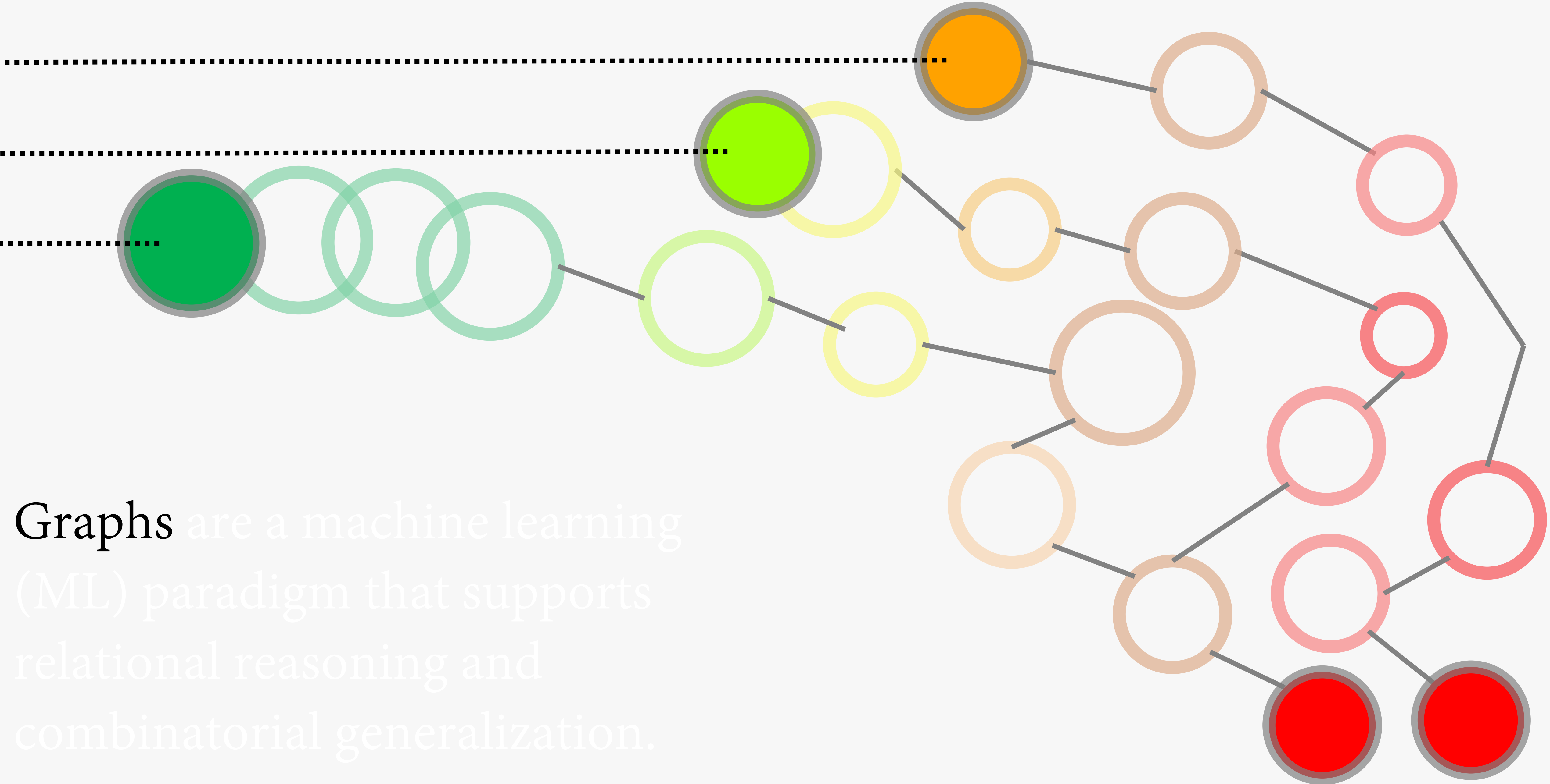


“j'ai décidé d'être heureux parce que c'est bon pour la santé"

Voltaire

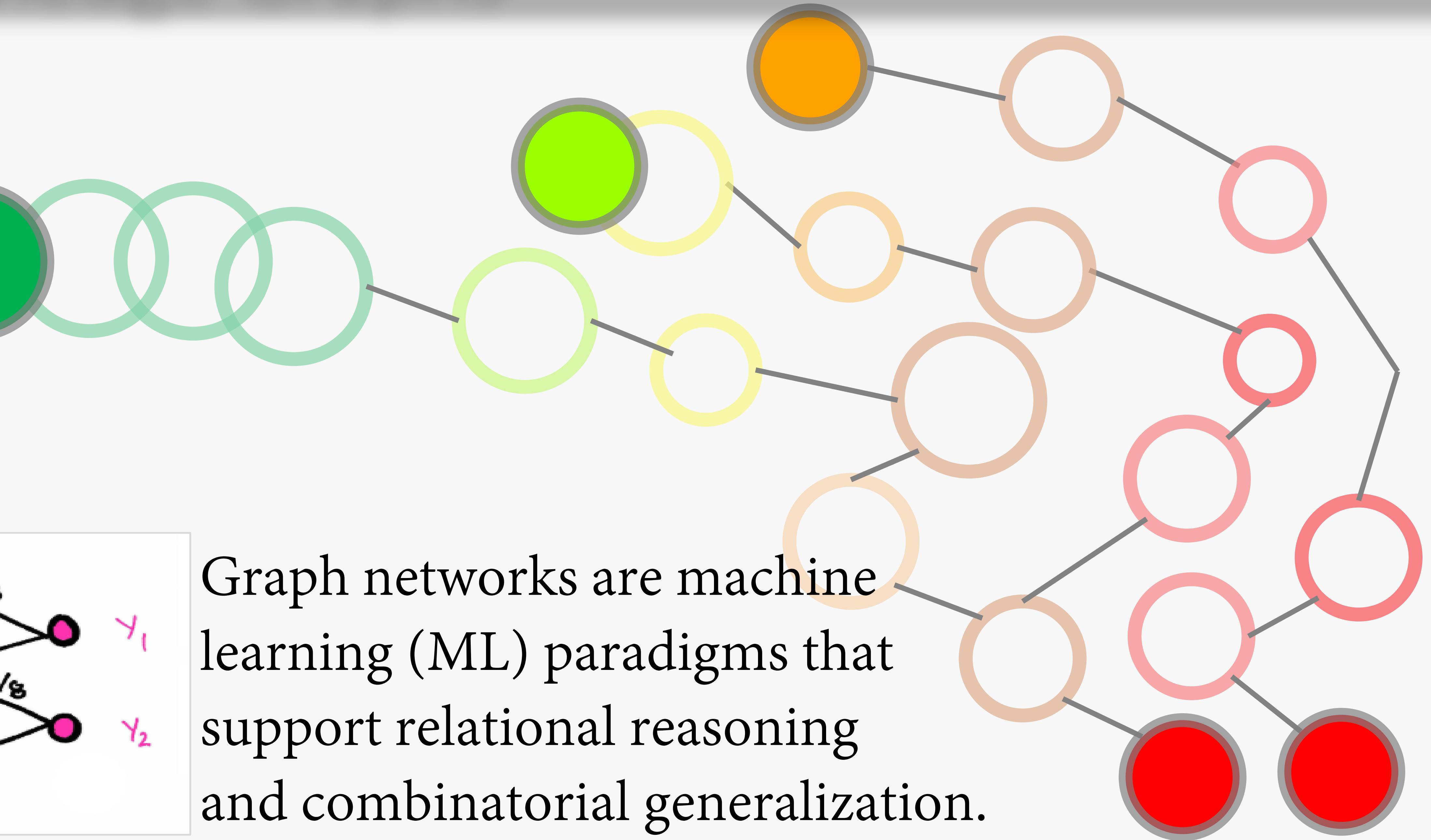

Happiness
1
2
3
Time
3
2
1
Many paths with similar outcomes
Outcome = kids at home, safe, fed
but variable value
Value: child happiness, timescale, intangibles

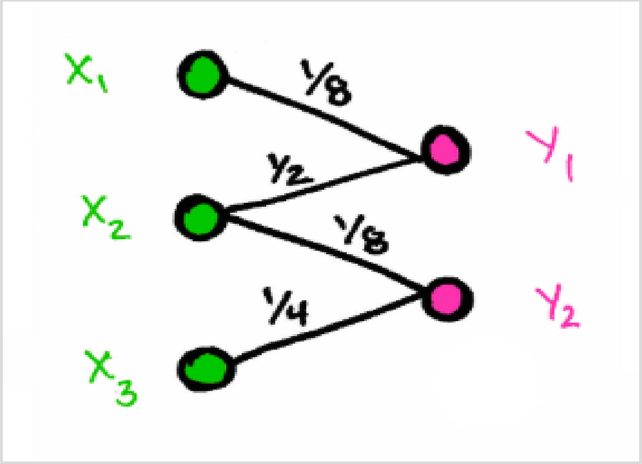

Quality of Outcome
1
2
3
Time
3
2
1



Graphs are a machine learning (ML) paradigm that supports relational reasoning and combinatorial generalization.

Knowledge Graphs

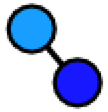




Graph networks are machine learning (ML) paradigms that support relational reasoning and combinatorial generalization.

http://www.mkbergman.com/2244/a-common-sense-view-of-knowledge-graphs/

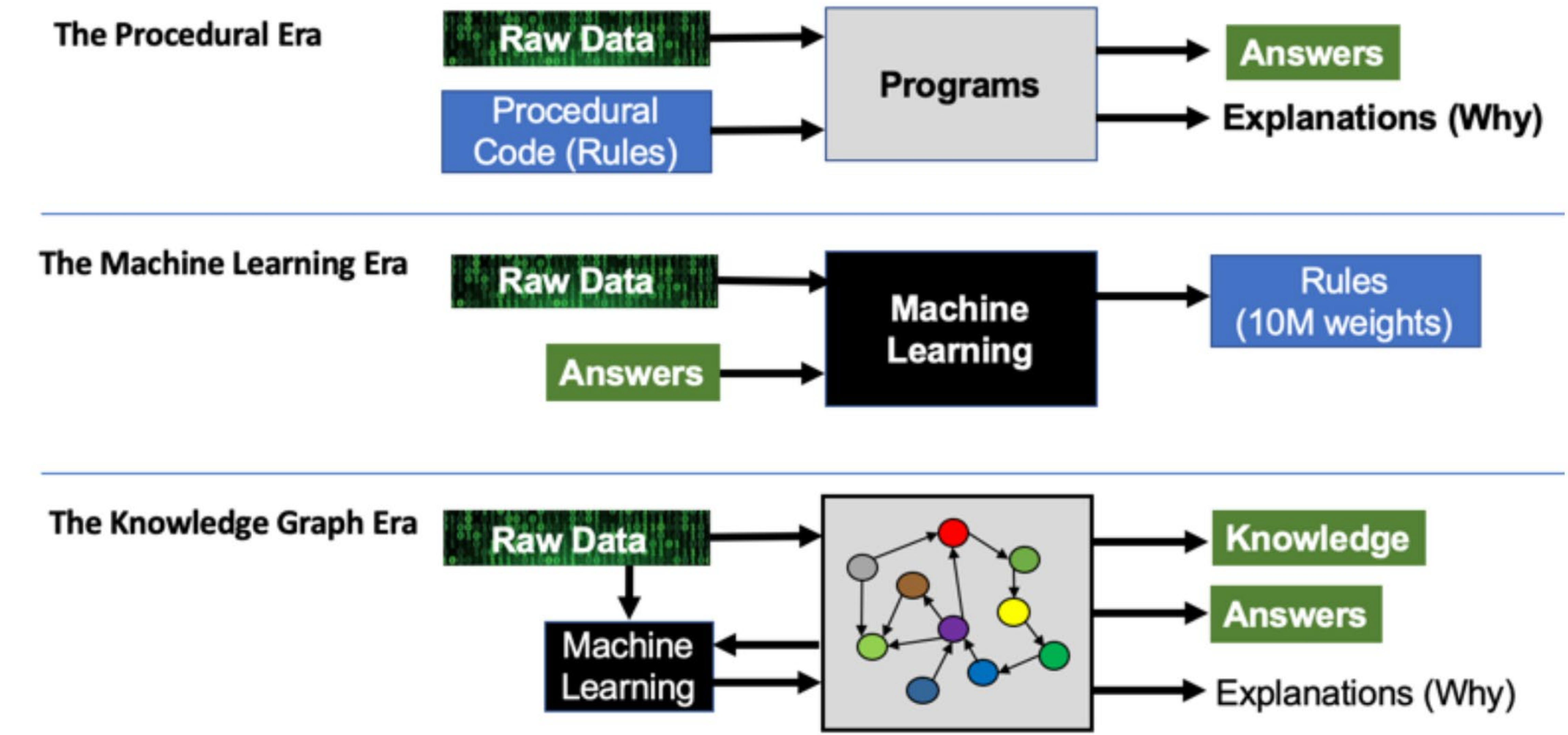
BOOK - http://bit.ly/Knowledge-Representation



https://www.springer.com/gp/book/9783319980911

http://www.mkbergman.com/wp-content/themes/ai3v2/images/2012Posts/ontology_build.gif

Knowledge Graphs: The core of the 3rd era of computing

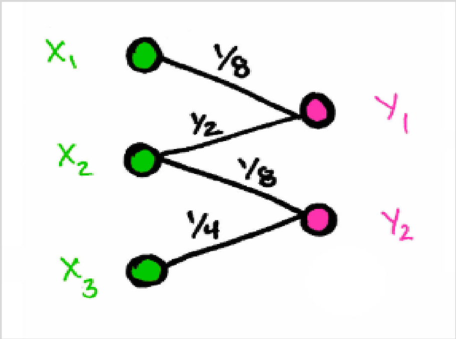


@dmccreary

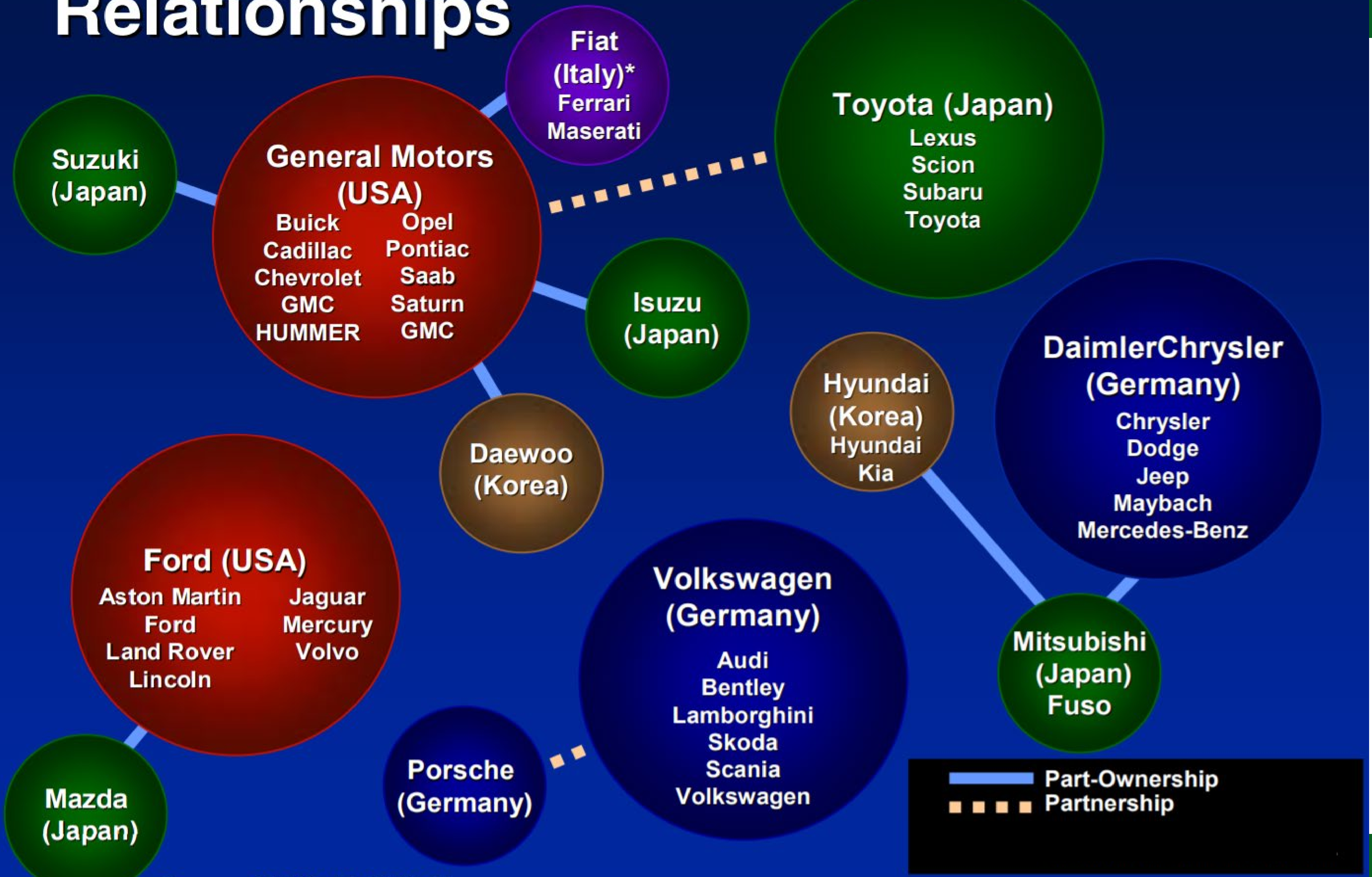
Use machine learning to continuously enrich knowledge

RELATIONSHIPS

between entities are critical in the context of making decisions. Data available to the human mind (or the system) in the context of entities or objects or processes, under consideration, are used directly (as data) or converted to information (by humans) to fuel decisions (outcomes). Knowledge graphs are a form of **bio-mimicry** tool, to enable non-human computer systems to understand relationships between entities, objects, processes, people, and things (think IoT, internet of things). Resource Description Framework (**RDF)** is a standard to describe resources and is based on principles of linguistics (noun, verb, subject, predicate).

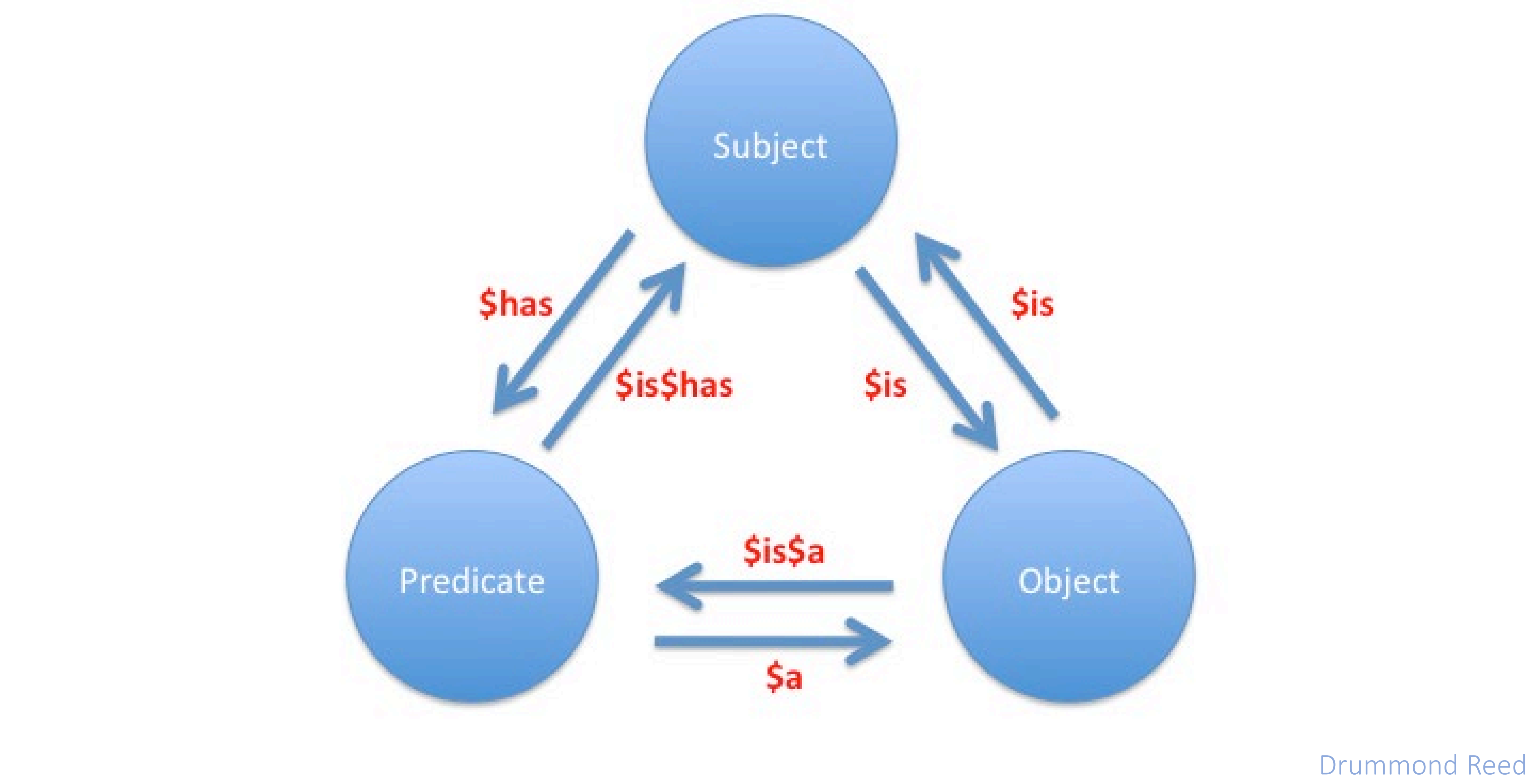


Relationships



Sources: *AAA World*, AV World ● https://www.researchgate.net/publication/228383625_Supply_chain_networks

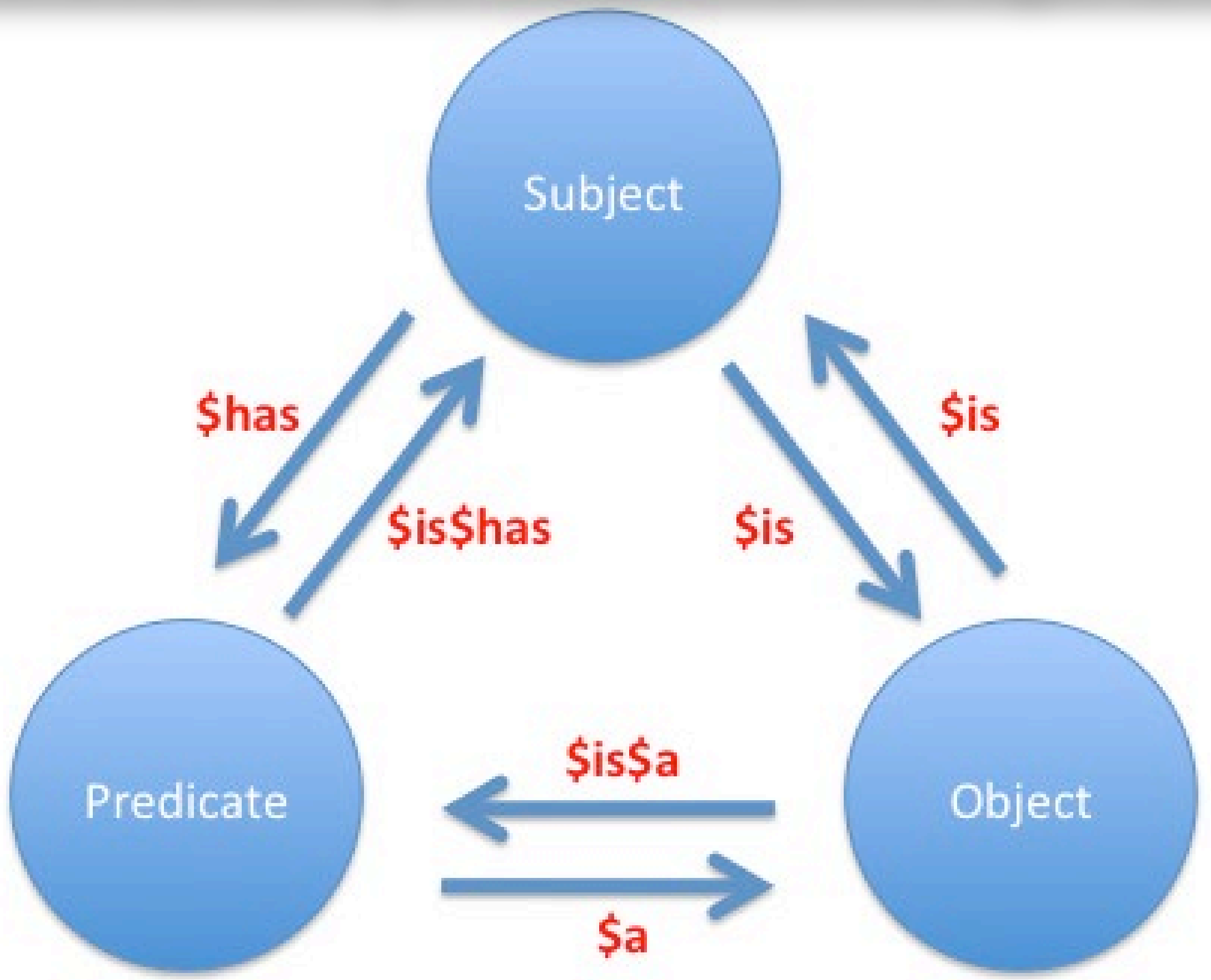
RDF expresses RELATIONSHIPS as “triples” which are based on principles of linguistics (noun, verb, subject, predicate).



Resource Description Framework (RDF) is a standard to describe resources. It is written in XML and machine-readable.

Linguistics, Relationships and Knowledge Graph Networks

In the “Golden Triangle" of metagraph relationships: hasa (which is literally $has/$has/$a) defines a subset of $has relationships in which the predicate is also an object. Asserting a predicate **as an object** is different than asserting it **as a predicate**. Neither implies that the other exists. They have a logical relationship, it is the same predicate involved in both cases, but asserting it as a predicate does not mean it is also an object, and asserting is as an object does not need it to be also a predicate.



$ is a subsegment delimiter (Global Context Symbols)
https://wiki.oasis-open.org/xri/XriThree/GcsDelimiter

Drummond Reed drummond.reed@xdi.org https://lists.oasis-open.org/archives/xdi/201004/msg00013.html

Are computational standards, syntax semantics and ontologies influenced by linguistic bias?

Subject	Predicate	
	linking verb	subject complement noun or adjective
The aliens	were	killers.
The rats	are	ugly.
The griff	is	grafunkulous.

- The predicate is the action
- Action verbs are easy to identify, but remember **verbs of being**: am, is, are, were, was
- A sentence *can* have more than one predicate

Subject and Predicate

My EnglishTeacher .eu

What is a Subject?

- *A subject is the person or thing that is doing an action, or the person or thing that is the focus of the sentence*

What is a Predicate?

- *The predicate of the sentence is the part that contains the action.*

https://english.eagetutor.com/spoken-english-grammar/subject-predicate-and-object

What happens when the RDF ontology creator speaks a language where the rules of English grammar are not applicable?

At the heart of the predicate is a verb. In addition to the verb, a predicate can contain direct objects, indirect objects, and various kinds of phrases.

A sentence has two parts: the subject and the predicate. The subject is what the sentence is about, and the predicate is a comment about the subject.

www.grammar-monster.com/glossary/predicate.htm

Examples of Predicates of Sentences

Here are some examples of predicates. In each example, the predicate of the sentence is shaded and the verb in the predicate is in bold.

- Elvis **lives**.
- Adam **lives** in Bangor.
- The telegram **contained** exciting news.
- The girls in our office **are** experienced instructors.
- They **are** experienced instructors, who acquired their experience in France.

Predicates in Clauses

A clause contains a subject and predicate too. The examples below are all clauses not sentences. The predicate is shaded and the verb of the clause is in bold.

- who **lives** with her mother
  (The subject is the relative pronoun *who*.)
- which **was** somewhat unexpected
  (The subject is the relative pronoun *which*.)
- that **points** to the North Pole
  (The subject is the relative pronoun *that*.)

www.grammar-monster.com/glossary/predicate.htm

Why RDF may be just a part of the solution: Is linguistic bias embedded in the grammatical context of RDF triples?

Semantic Standards in other languages?

https://lists.w3.org/Archives/Public/www-archive/2005Feb/att-0050/eswc-i18n.pdf

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1892103/pdf/1471-2105-8-S3-S6.pdf

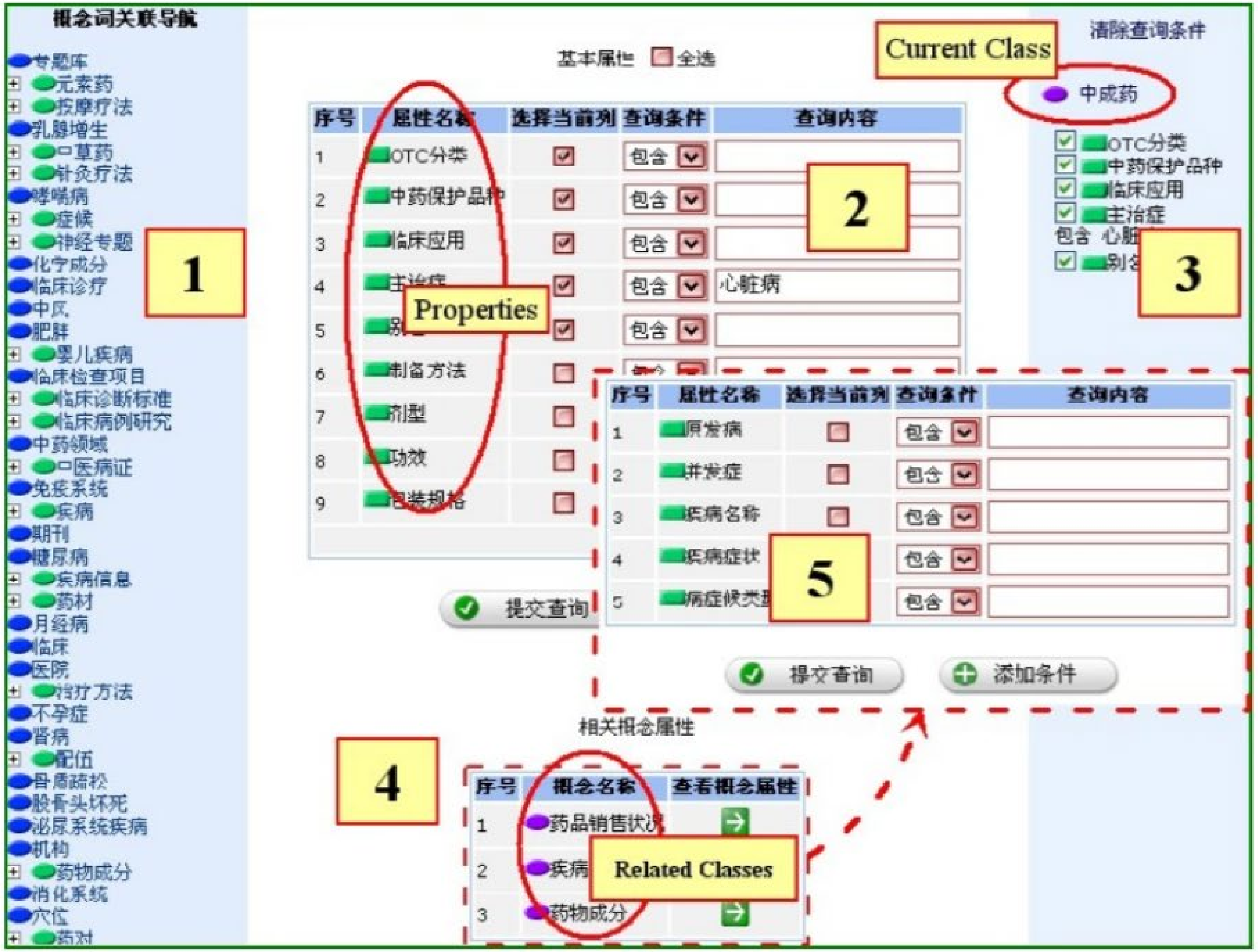
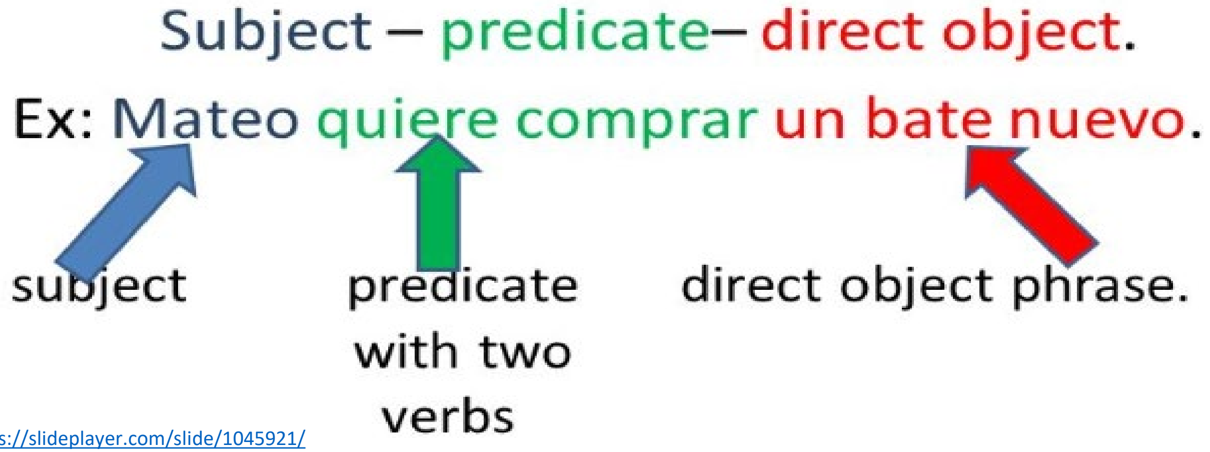


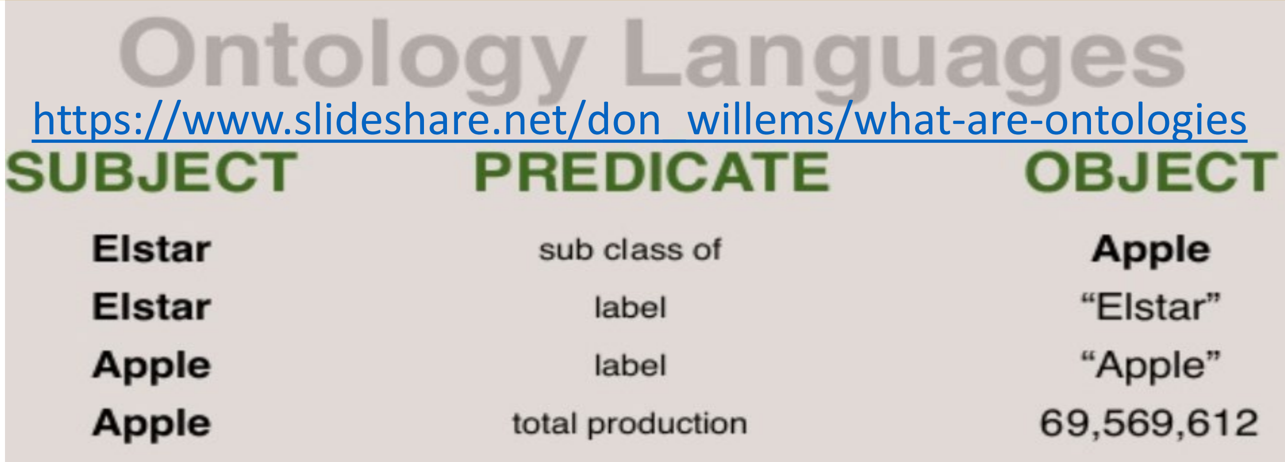
Figure 4
DartSearch. The default user interface for DartSearch.

Knowledge Graphs relevant to SENSEE, ART, & DIDA'S KIDS

When data is mapped against an OWL/RDF ontology, instances of the data are expressed based upon the idea of making statements about resources in the form of **subject–predicate–object** expressions. These expressions are known as *triples* in RDF terminology. The 'Subject' denotes the object, and the predicate denotes a single semantic trait or aspect of the object that can be a literal value or expressed as a relationship between the subject and another object that is the target of the relationship. For example, the notion "The soil has a pH of 8" in RDF triple is **subject** denoting "soil" and **predicate** denoting "pH" and an **object** denoting "8" which is the OWL/RDF take on using the object as the subject from the classical entity–attribute–value model within object-oriented design: object (soil), attribute (pH) and value (8). The object (soil) can also have another attribute (contains) that can point to another object (phosphate). The object (phosphate) might have an attribute (produces) another object (acidity). Yet again, the object (soil) might have an attribute (contains) another object (microbes). This is one reason why RDF triples, despite their shortcomings and potential for linguistic bias, enables the formation, to link a series of relationships, between two or more objects. The latter is the foundation on which directed graphs can be built. Hence, knowledge graphs.



https://slideplayer.com/slide/1045921/



Ontology Languages

https://www.slideshare.net/don_willems/what-are-ontologies

SUBJECT	PREDICATE	OBJECT
Elstar	sub class of	Apple
Elstar	label	"Elstar"
Apple	label	"Apple"
Apple	total production	69,569,612

Which / what nodes are the graphs connecting?

https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

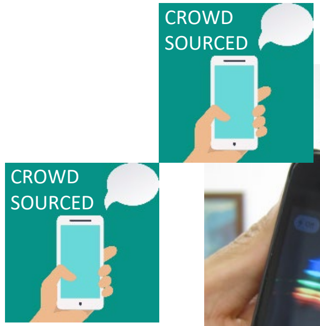
Which / what nodes are the graphs connecting?

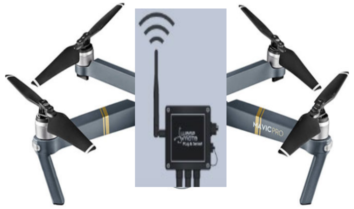
SENSORS

CHAT, TALK

DATA, MAPS

IoT ECOSYSTEM



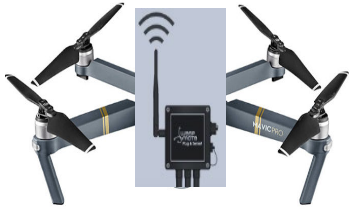


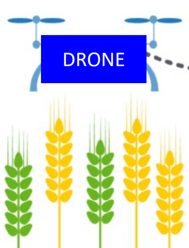
PORTABLE SPEC

https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Web of Knowledge Graph Networks for SENSEE, ART, DIDA’S, KIDS

Nodes connected by graphs

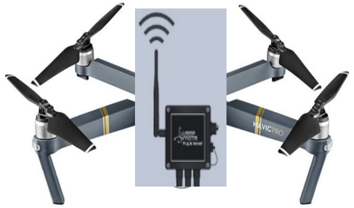




https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Web of Knowledge Graph Networks for SENSEE, ART, DIDA'S, KIDS

Must be connected to a mobile device?



https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Web of Knowledge Graph Networks for SENSEE, ART, DIDA'S, KIDS

Connected KIDS

https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Web of Knowledge Graph Networks are necessary for ART, DIDA'S, KIDS

Connect KIDS PEAS to other ecosystems

https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Web of Knowledge Graph Networks are necessary for ART, DIDA'S, KIDS

Chromosomes of Knowledge

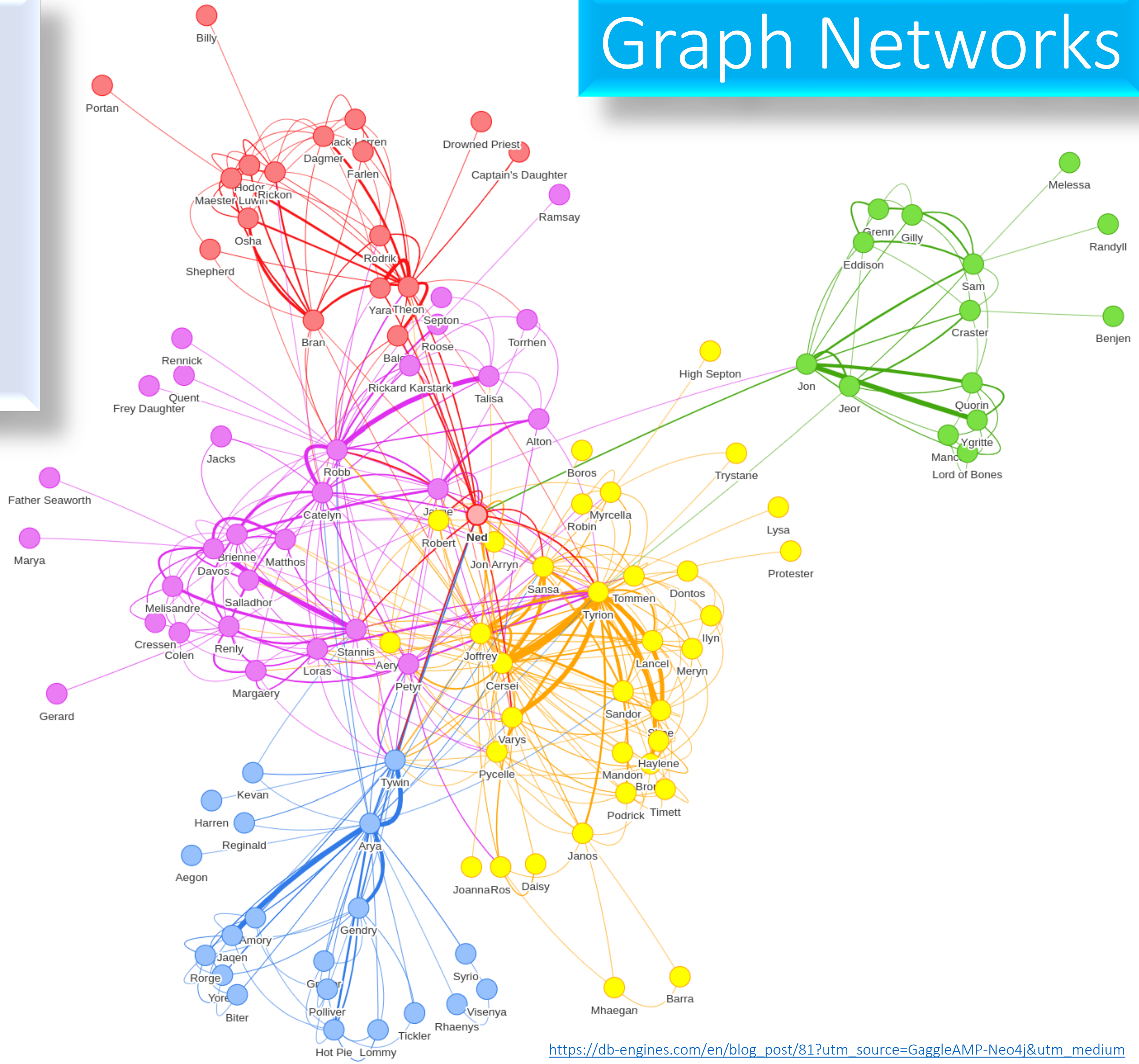
https://www.aac.or.at/smart-farming
https://www.findlight.net/blog/2018/09/01/spectroscopy-precision-agriculture/
http://trajectorymagazine.com/wp-content/uploads/2017/04/TRJ-009-Q1-2013_Final.pdf

Graph Networks

Increased adoption of tools based on graph theory. HTAP integrates graph transactions (OLTP) and analytic processing (OLAP) using graph databases and graph algorithms (relationships are key).

Graph algorithms provide one of the most potent approaches to analyzing connected data because their mathematical calculations are specifically built to operate on relationships. There are many types of graph algorithms and categories. The three classic categories consider the overall nature of the graph: pathfinding, centrality, and community detection. However, other graph algorithms such as similarity and link prediction algorithms consider and compare specific nodes.

- Pathfinding (and search) algorithms are fundamental to graph analytics and algorithms and explore routes between nodes. These algorithms are used to identify optimal routes for uses such as logistics planning, least cost routing, and gaming simulation.
- Centrality algorithms help us understand the roles and impact of individual nodes in a graph. They're useful because they identify the most important nodes and help us understand group dynamics such as credibility, accessibility, the speed at which things spread, and bridges between groups.
- Community algorithms evaluate related sets of notes, finding communities where members have more relationships within the group. Identifying these related sets reveals clusters of nodes, isolated groups, and network structure. This helps infer similar behavior or preferences of peer groups, estimate resiliency, find nested relationships, and prepare data for other analyses.
- Similarity algorithms look at how alike individual nodes are. By comparing the properties and attributes of nodes, we can identify the most similar entity and score differences. This helps build more personalized recommendations as well as develop ontologies and hierarchies.
- Link Prediction algorithms consider the proximity of nodes as well as structural elements, such as potential triangles between nodes, to estimate the likelihood of a new relationship forming or that undocumented connections exist. This class of algorithms has many applications from drug repurposing to criminal investigations.



https://db-engines.com/en/blog_post/81?utm_source=GaggleAMP-Neo4j&utm_medium

Web of Knowledge Graph – Select List of References

[0] BOOK - http://bit.ly/Knowledge-Representation

[1] P. Bonatti, S.Decker, A.Polleres and V. Presutti. Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web, Dagstuhl Seminar 18371, 2019.

[2] E. Marchi and O. Miguel. On the Structure of the Teaching-learning Interactive Process, International Journal of Game Theory, 3 (2), 1974.

[3] R. Bakker. Knowledge Graphs: Representation and Structuring of Scientific Knowledge, Ph.D. Thesis, University of Twente, Enschede. 1987.

[4] L. Ehrlinger and W. Wöß. Towards a Definition of Knowledge Graphs. SEMANTiCS, 2016.

[5] A Common Sense View of Knowledge Graphs. July 2019.http://www.mkbergman.com/2244/a-common-sense-view-of-knowledge-graphs/

[6] P. Domingos. "A Few Useful Things to Know about Machine Learning." Communications of the ACM 55(10): 78 – 87,2012

[7] A. Sheth, S. Parera, S. Wijaratne, K. Thirunarayan. "Knowledge will propel machine understanding of content: extrapolating from current examples." Proceedings of the International Conference on Web Intelligence. ACM, 2017.

[8] M. Noura, A. Gyrard, S. Heil, M. Gaedke. Automatic Knowledge Extraction to build Semantic Web of Things Applications. IEEE IoT Journal 2019

[9] S. Bhatt, S. Padhee, A. Sheth, K. Chen, V. Shalin, D. Doran, B. Minnery. "Knowledge Graph Enhanced Community Detection and Characterization." Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. ACM, 2019.

[10] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational ML for knowledge graphs. IEEE Proceedings 104(1), 11-33, 2015

[11] H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web, 8(3), pp.489-508, 2017

[12] S. Lalithsena, S. Perera, P. Kapanipathi, and A. Sheth. Domain-specific hierarchical subgraph extraction: A recommendation use case. In IEEE International Conference on Big Data (Big Data) (pp. 666-675). IEEE, 2017.

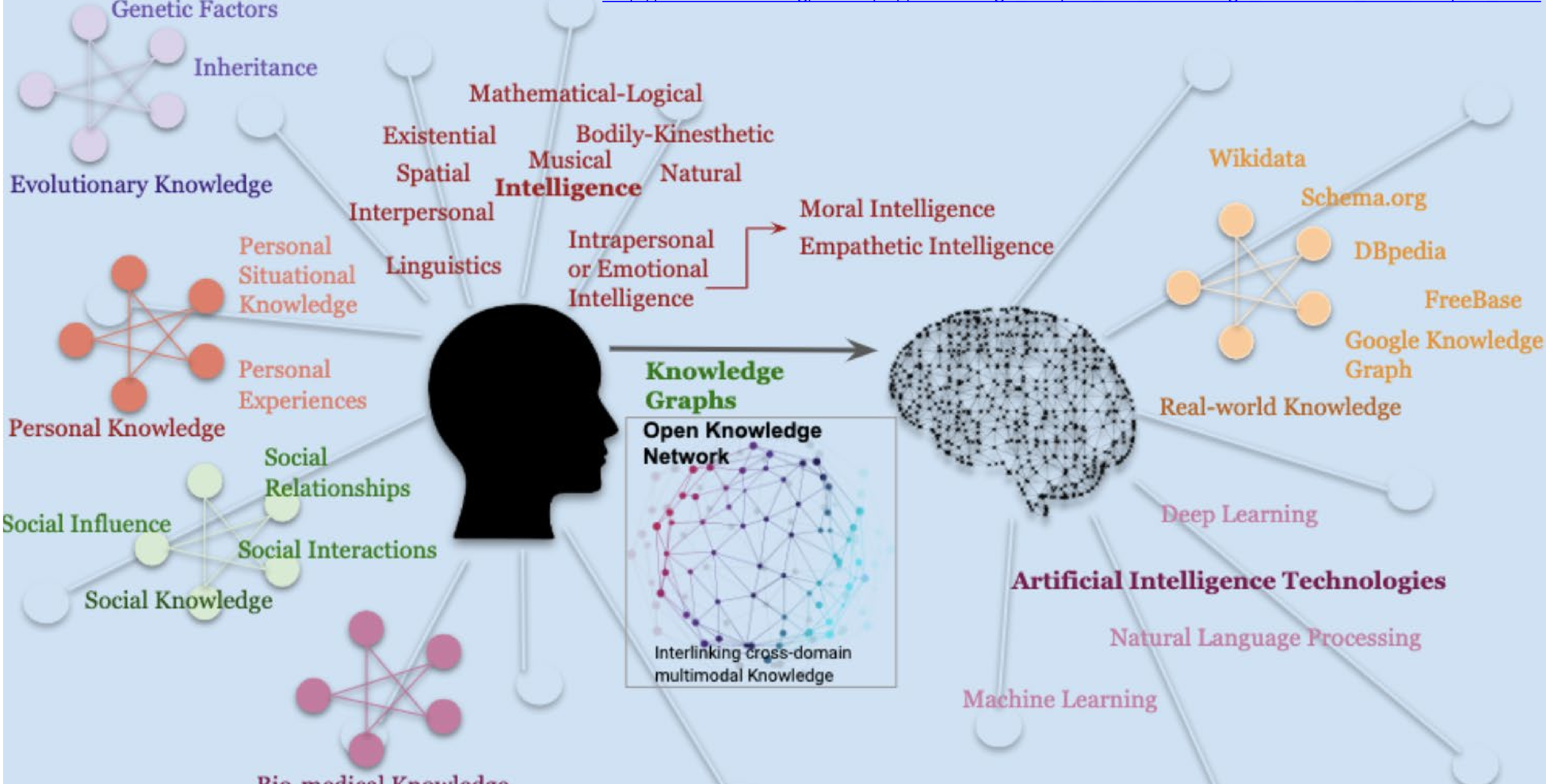
[13] R. Trivedi, H. Dai, Y. Wang, L. Song. "Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 3462-3471). JMLR. org, 2017.

[14] OKN: Open Knowledge Network: Creating the Semantic Information Infrastructure for the Future (Summary of the Big Data IWG Workshop),October 4–5, 2017. November 2018. https://www.nitrd.gov/news/Open-Knowledge-Network-Workshop-Report-2018.aspx

[15] P. Ekman. "Basic Emotions", in Dalgleish, T; Power, M (eds.), Handbook of Cognition and Emotion, Sussex, UK: John Wiley & Sons, 1999.

[16] Amit Sheth, Swati Padhee, Amelie Gyrard, 'Knowledge Graphs and Knowledge Networks - The Story in Brief," IEEE Internet Computing, July-Aug 2019 http://wiki.knoesis.org/index.php/Knowledge_Graphs_and_Knowledge_Networks:_The_Story_in_Brief

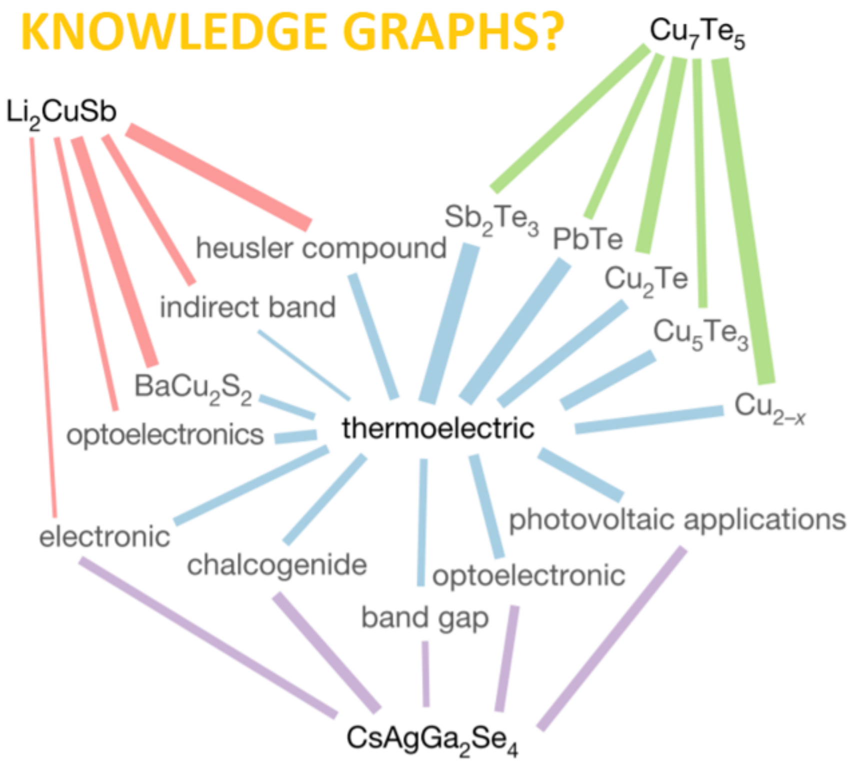
http://wiki.knoesis.org/index.php/Knowledge_Graphs_and_Knowledge_Networks:_The_Story_in_Brief



http://bit.ly/KR-SETH and http://bit.ly/KR-BERGMAN

Novel Paradigm for Machine-Assisted Graph Theoretic Approach to Mine Non-Obvious Relationships

The width of the edges between 'thermoelectric' and the context words (blue) is proportional to the cosine similarity between the word embeddings of the nodes, whereas the width of the edges between the materials and the context words (red, green and purple) is proportional to the cosine similarity between the word embeddings of context words and the output embedding of the material. The context words are top context words according to the sum of the edge weights between the material and the word 'thermoelectric'.



https://doi.org/10.1038/s41586-019-1335-8

Unsupervised word embeddings capture latent knowledge from materials science literature

Vahe Tshitoyan[1,3]*, John Dagdelen[1,2], Leigh Weston[1], Alexander Dunn[1,2], Ziqin Rong[1], Olga Kononova[2], Kristin A. Persson[1,2], Gerbrand Ceder[1,2]* & Anubhav Jain[1]*

https://www.nature.com/articles/s41586-019-1335-8

Novel Paradigm for Machine-Assisted Graph Theoretic Approach to Mine Non-Obvious Relationships

KNOWLEDGE GRAPHS?

Scientific progress relies on the confluence of efficient assimilation of existing knowledge in order to minimize re-invention. The methodology in this paper may create a tool to plumb the depths of the unknown unknowns, where catalysts for scientific breakthroughs often reside. The authors are incisive to point out that this approach may be "generalized to other language models, such that the probability of an entity (a material or molecule) co-occurring with words, that represent a target application or property, can be treated as an indicator of performance." Entity-relationship mode remains the "bread and butter" of context-awareness while RDF is a more general model of entities (nodes) and relationships. Thus, the paper strengthens the notion that knowledge graphs may aid in unleashing new ideas. Context-aware embeddings such as NLP BERT or ELMo may improve predictions. This document (PEAS) and the accompanying ideas (SIGNALS) are in quest of these tools. The paper indicates the potential for new research at the nexus of natural language processing, linguistics, semantics, and science, to advance knowledge discovery.

https://www.nature.com/articles/s41586-019-1335-8

Artificial Reasoning Outcomes → Reasonable Expectations

It is easy to illustrate, but quite difficult for systems to claim real ‘knowledge’ discovery.

Can graph networks catalyze data to reveal information?

But, beware of snake oil sales and stupidity

BEWARE OF STUPIDITY

www.foxbusiness.com/features/after-watson-ibm-looks-to-build-brain-in-a-box

https://emoshape.com/emoshape-enhances-its-cutting-edge-emotion-chip-with-the-addition-of-cloud-service/

EMOSHAPE
EMOTIONS NEXT FRONTIER

Emotion Synthesis | Products | Order | News

Emoshape Enhances Its Cutting Edge Emotion Chip with the Addition of Cloud Service

On: Jun 03 / Author: Patrick Levy-Rosenthal / Categories: Uncategorized /

The Lowest Common Denominator

We are immersed in **data** swamps.
Actions depend on **information**.
Informed by **knowledge**.
Learn from **experience**.

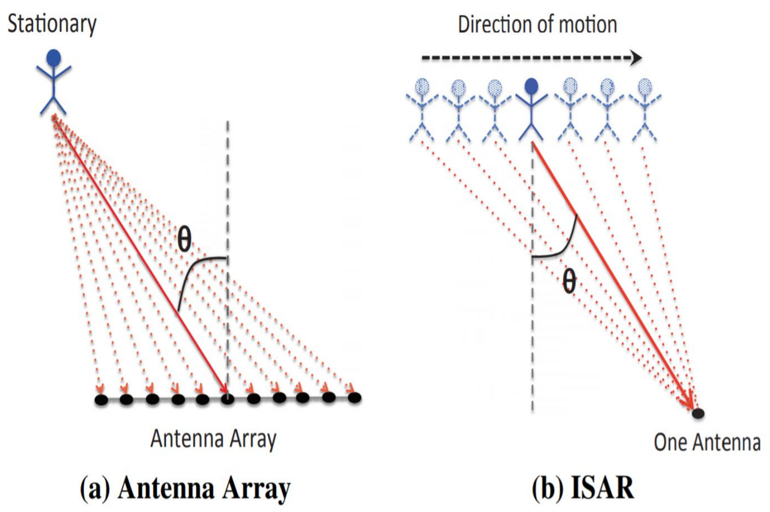
SENSEE → Choose sensors and then harvest DATA from specific SENSORS

SENSE

S E N S E

Don’t bind.
Reflect

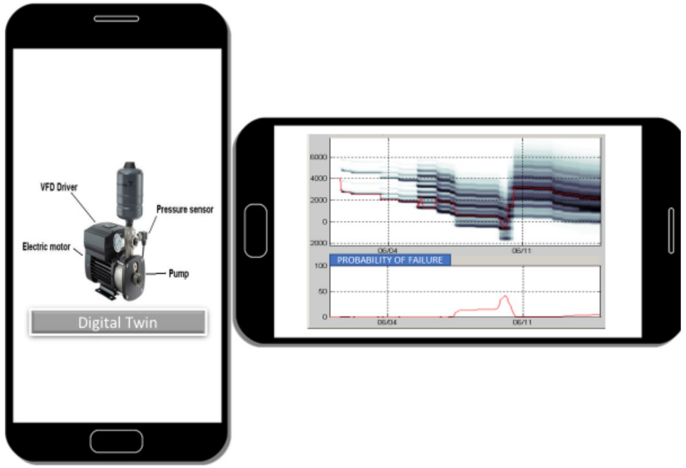
CONVERGE



SENSE RF REFLECTION

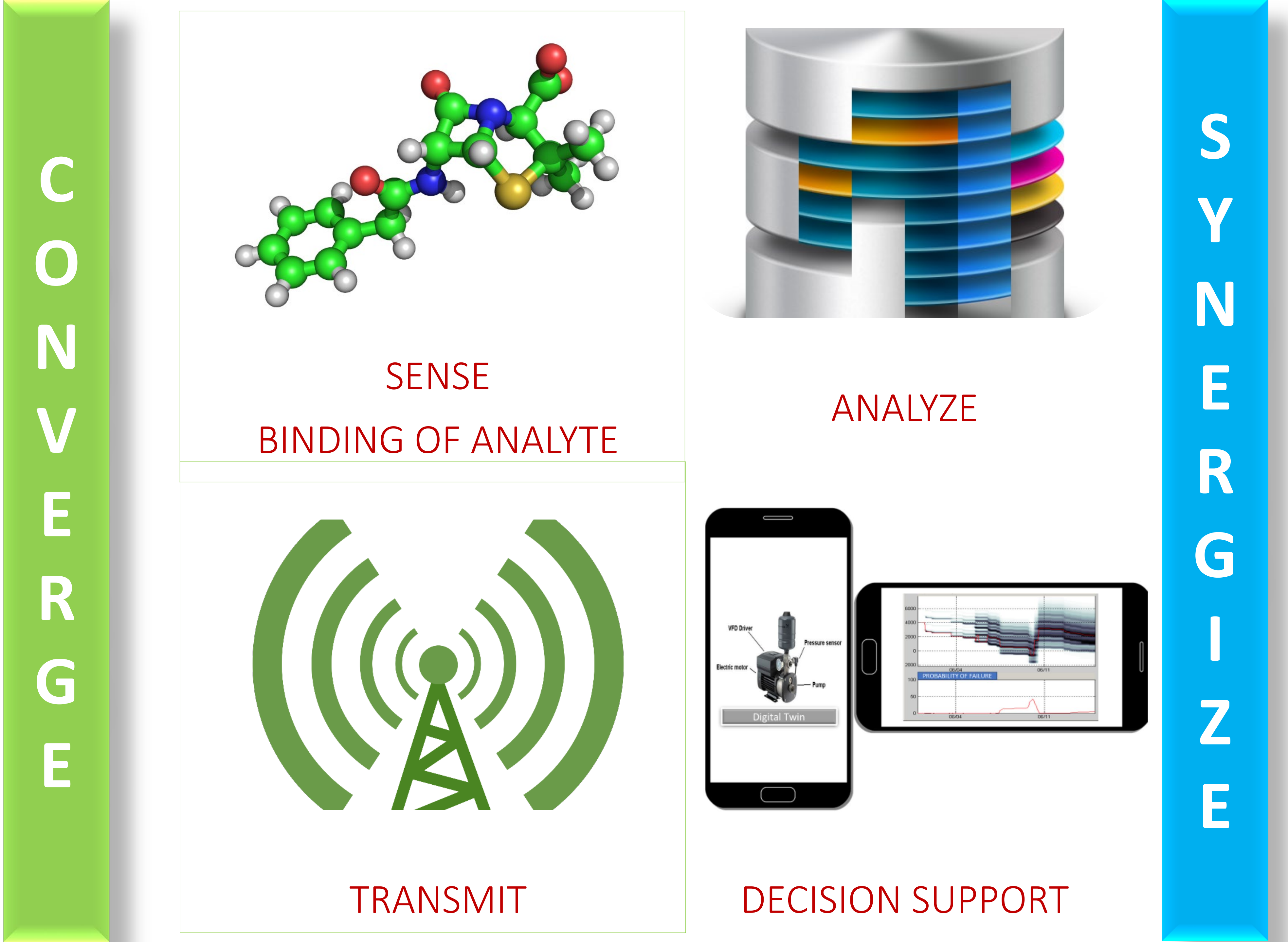
ANALYZE

TRANSMIT



DECISION SUPPORT

SYNERGIZE


CONVERGE
SENSE
BINDING OF ANALYTE
ANALYZE
TRANSMIT
VFD Driver
Pressure sensor
Electric motor
Pump
Digital Twin
PROBABILITY OF FAILURE
DECISION SUPPORT
SYNERGIZE

How does a sensor work? Sensors that bind analytes.

https://onlinelibrary-wiley-com.libproxy.mit.edu/doi/pdf/10.1002/anie.199423751

The Key–Lock Theory and the Induced Fit Theory 1 / 4

There are sensors that may not bind analytes but are activated by reflected radio waves (WiFi, radar, sonar)

REVIEWS

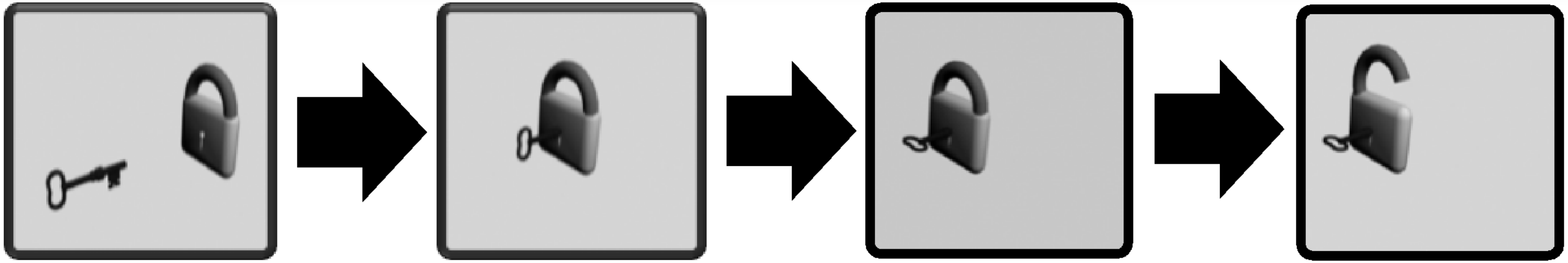
The Key–Lock Theory and the Induced Fit Theory

Daniel E. Koshland, Jr.

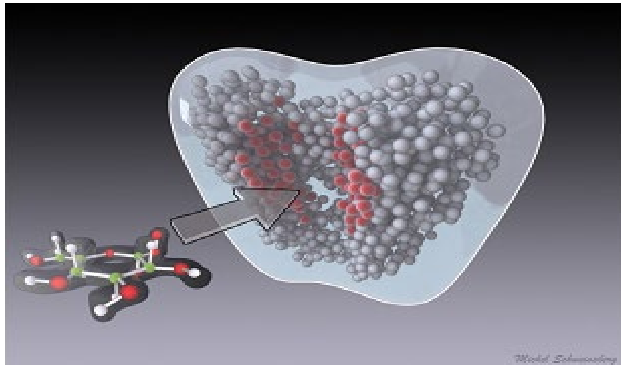
It is a great pleasure for me to contribute to this symposium honoring the great scientist Emil Fischer. My graduate thesis required me to synthesize [1-^{14}C]glucose, which introduced me to the famous Fischer–Kiliani synthesis of glucose and mannose from arabinose and HCN.[1] I was also particularly intrigued with his classic key–lock (or template) theory of enzyme specificity,[2, 3] which like all great theories seemed so obvious once one understood it.

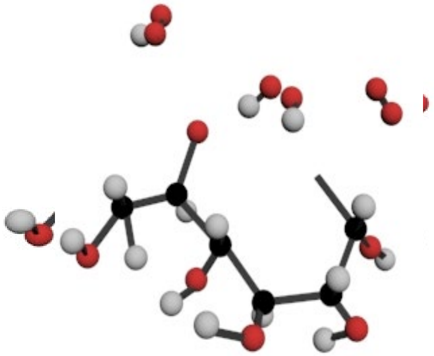
https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.199423751

How does a sensor work? The classical glucometer.



INTERACTION | RECOGNITION | TRANSDUCTION | ACQUISITION





BIORECOGNITION | TRANSDUCTION | ACQUISITION

Step 1) Biorecognition
Step 2) Binding and transduction
Step 3) Acquisition and data analytics

https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.199423751

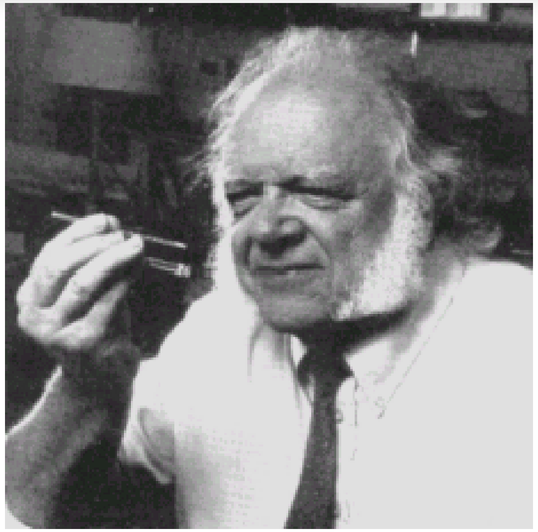
Unchanged: 1962 Classical Chemistry of Clark and Lyons

https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1962.tb13623.x

ELECTRODE SYSTEMS FOR CONTINUOUS MONITORING IN CARDIOVASCULAR SURGERY

Leland C. Clark, Jr., and Champ Lyons
Medical College of Alabama, Birmingham, Ala.

Instruments capable of continuously indicating the chemical composition of blood have proved to be useful in controlling heart-lung machines, in regulating operative and postoperative management of patients, and in teaching and research. At first, such instruments were used with sensors mounted directly in the extracorporeal blood circuit that is used for perfusion of open-heart surgery patients.[1] Later, continuous monitoring of both machine and patients was conducted by means of continuous withdrawal of blood pumped into external cuvettes equipped with appropriate sensors.



Leland C. Clark, Jr. (1959)

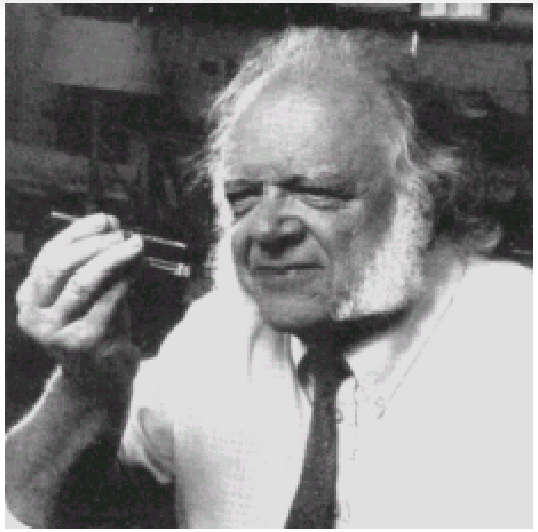
Unchanged: 1962 Classical Chemistry of Clark and Lyons

https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1962.tb13623.x

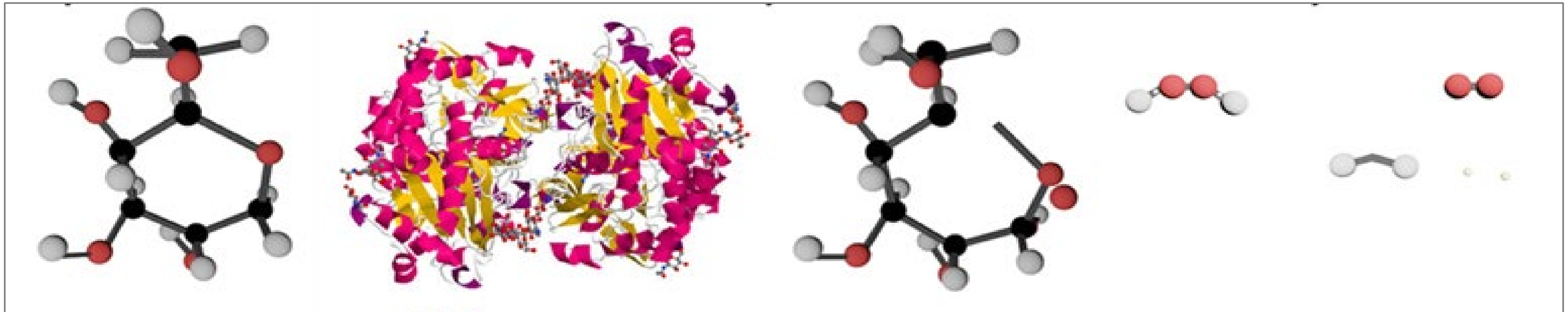
ELECTRODE SYSTEMS FOR CONTINUOUS MONITORING IN CARDIOVASCULAR SURGERY

Leland C. Clark, Jr., and Champ Lyons
Medical College of Alabama, Birmingham, Ala.

Instruments capable of continuously indicating the chemical composition of blood have proved to be useful in controlling heart-lung machines, in regulating operative and postoperative management of patients, and in teaching and research. At first, such instruments were used with sensors mounted directly in the extracorporeal blood circuit that is used for perfusion of open-heart surgery patients.[1] Later, continuous monitoring of both machine and patients was conducted by means of continuous withdrawal of blood pumped into external cuvettes equipped with appropriate sensors.

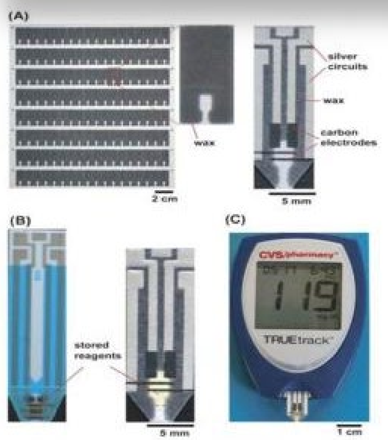


Leland C. Clark, Jr. (1959)



$$\text{Glucose} + O_2 \xrightarrow{GOX} \text{gluconic acid} + \mathrm{H_2O_2} \rightarrow 2H^+ + O_2 + 2e^-$$

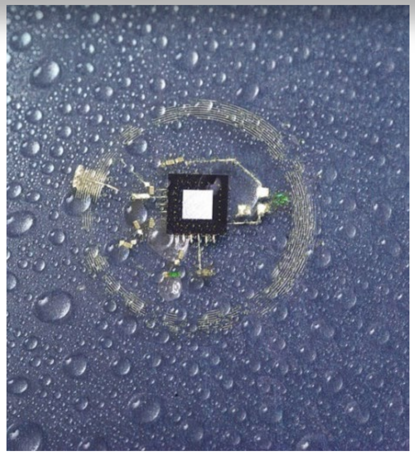
Glucometer – The Evolution of its Form and Function



Paper based (Whitesides)

Noninvasive glycemic monitoring (Wang)

WISP: Wearable Interactive Stamp Platform (MC10)



Resorbable (Rogers)

Contact lens (Google, Inc.)

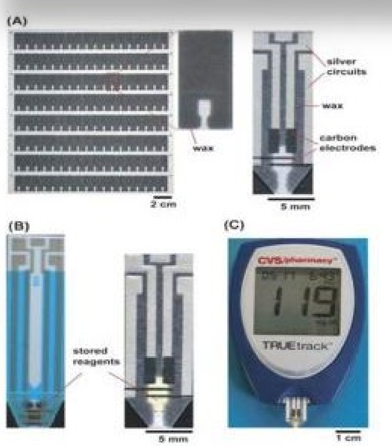
Stretchable fabric (Bhargava-UF)

Function	Form (elements and structure)
What a system does/could do	What a system is/could be
Creates behavior	Is aggregated and decomposed
Is a source of benefit/value	Is a source of costs
Requires form	Enables function

Source: E. Crawley, MIT Course Material

Glucometer – The Evolution of its Form and Function

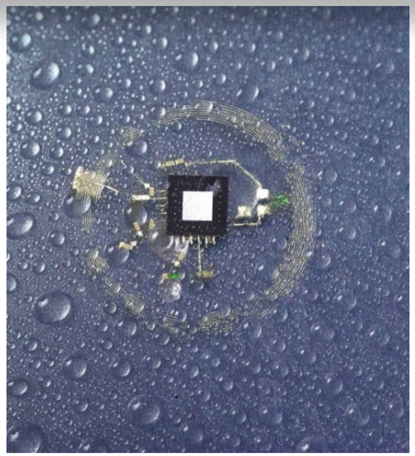
https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200603817



Paper based (Whitesides)

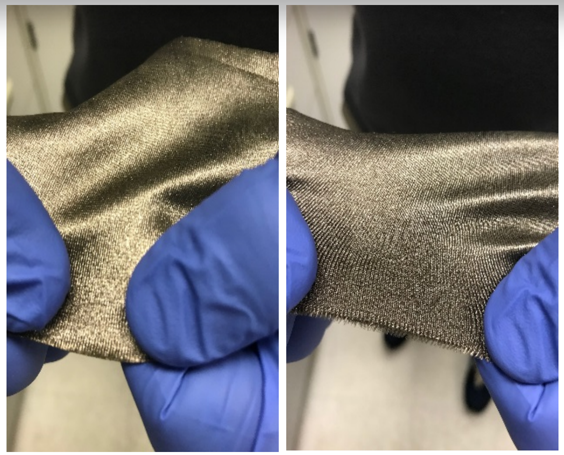
Noninvasive glycemic monitoring (Wang)

WISP: Wearable Interactive Stamp Platform (MC10)



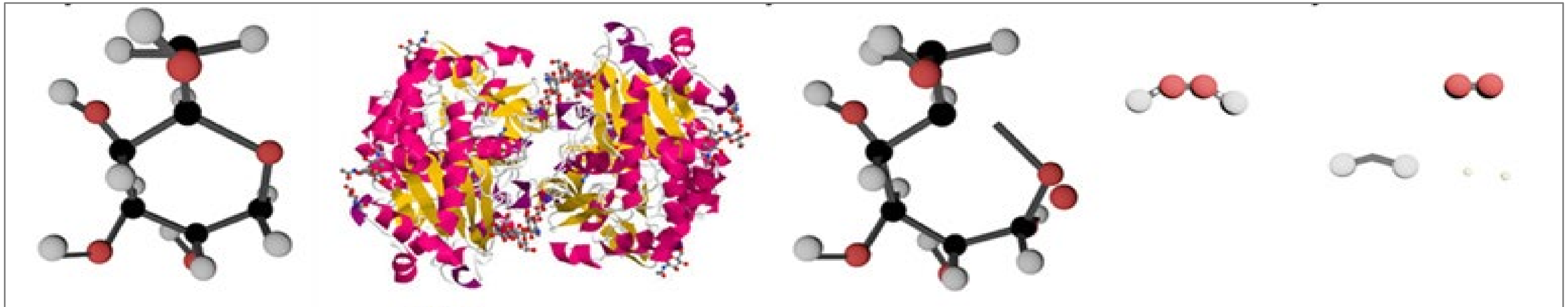
Resorbable (Rogers)

Contact lens (Google, Inc.)



Stretchable fabric (Bhargava-UF)

https://www.cientperiodique.com/article/CPQME-2-1-40.pdf



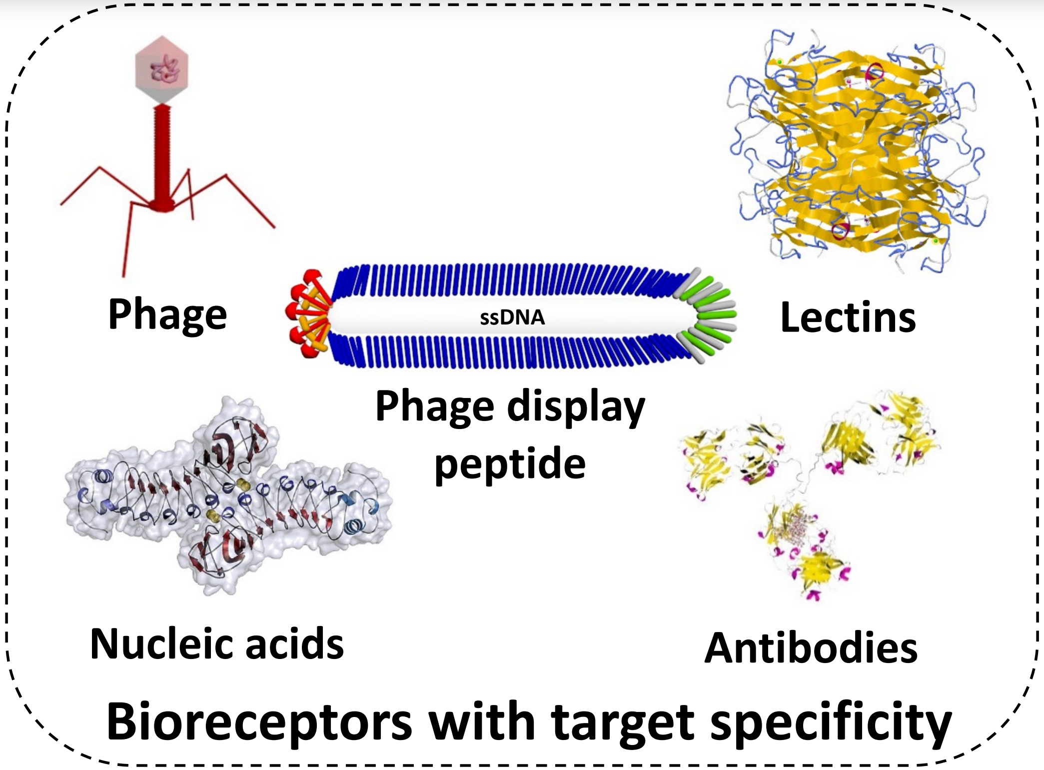
$$\text{Glucose} + O_2 \xrightarrow{GOX} \text{gluconic acid} + H_2O_2 \rightarrow 2H^+ + O_2 + 2e^-$$

The core chemistry developed by Clark and Lyons has not changed since 1962.

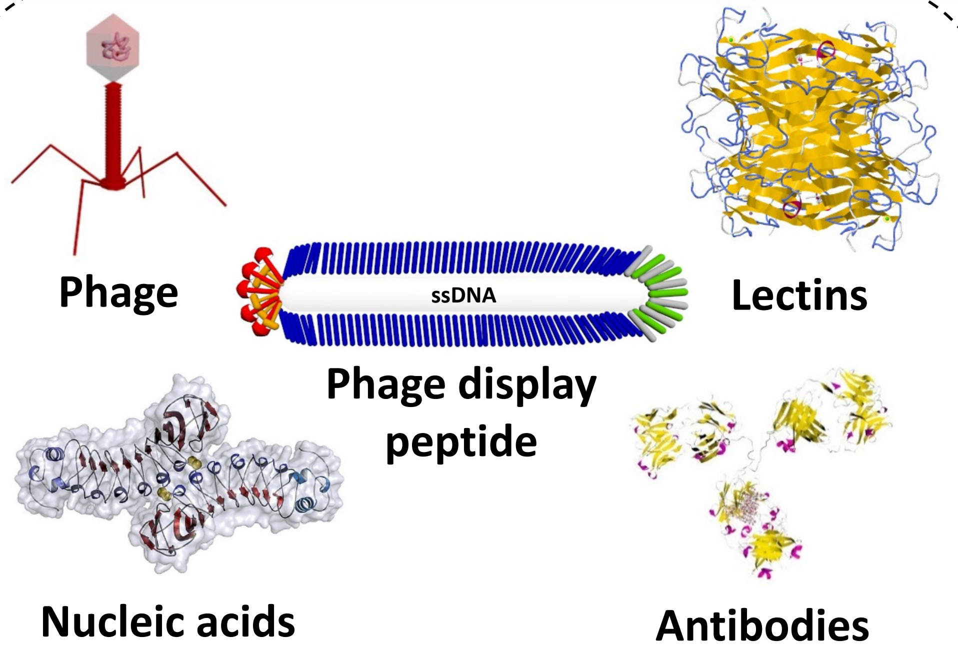
Evolution of Sensors ◆ Nano-bio Sensors

https://emclamor.wixsite.com/mclamorelab

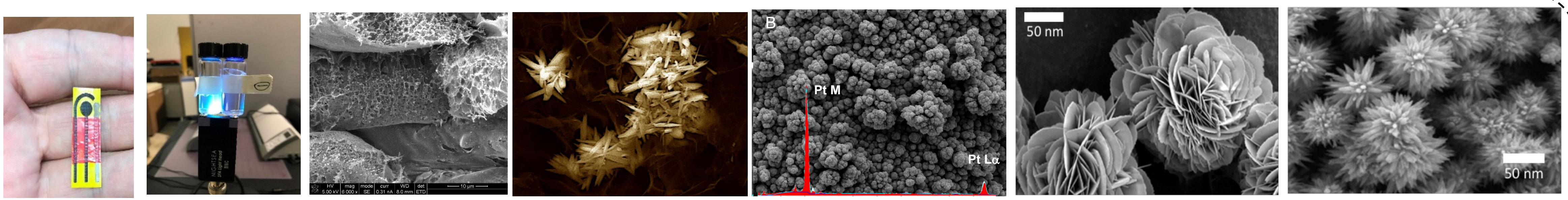
Evolution of Sensors ◆ Nano-bio Sensors



Evolution of Sensors ◆ Nano-bio Sensors

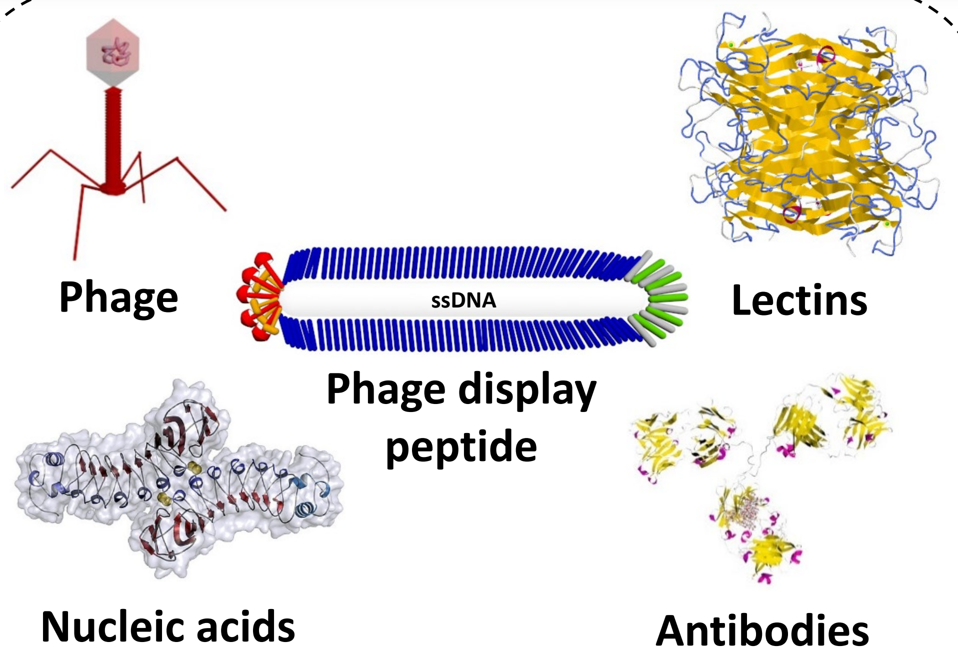


Bioreceptors with target specificity

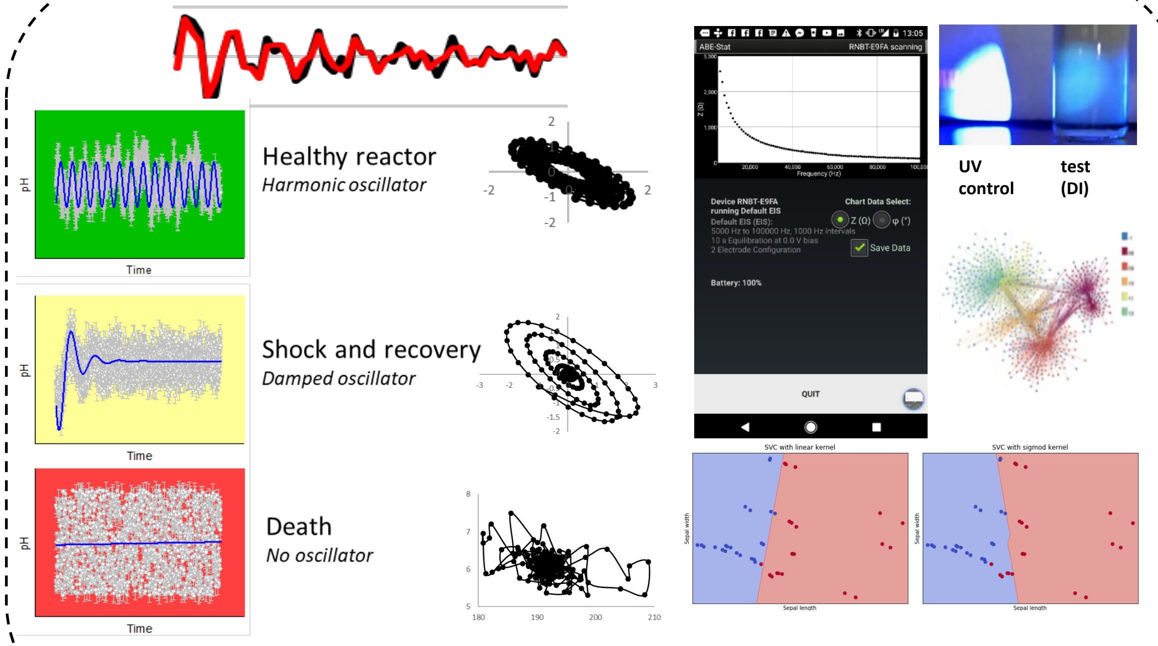


Nanomaterials improve signal transduction

Evolution of Sensors ◆ Nano-bio Sensors

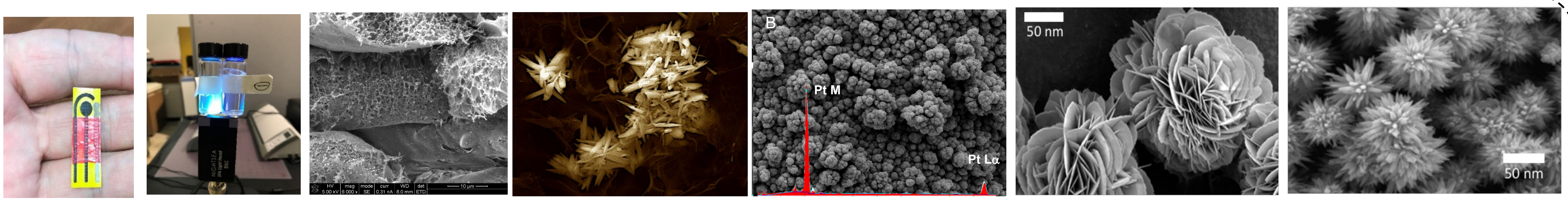


Bioreceptors with target specificity



Signal acquisition and data analytics

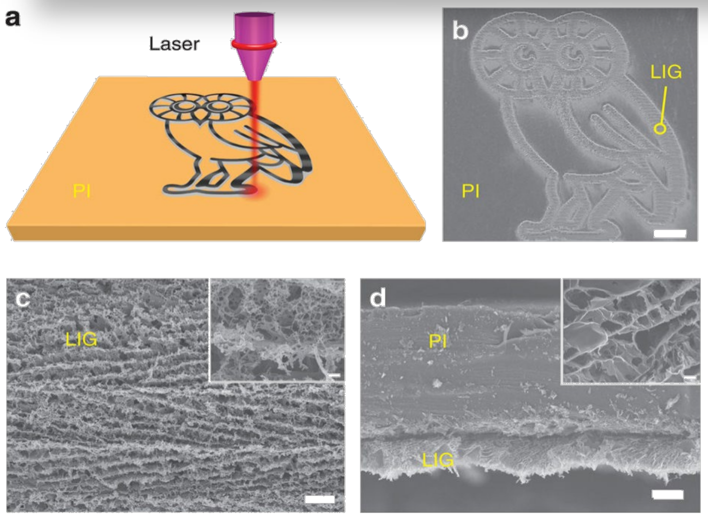
https://emclamor.wixsite.com/mclamorelab



Nanomaterials improve signal transduction

Evolution of Sensors ◆ 3D Printed Nano-bio Sensors

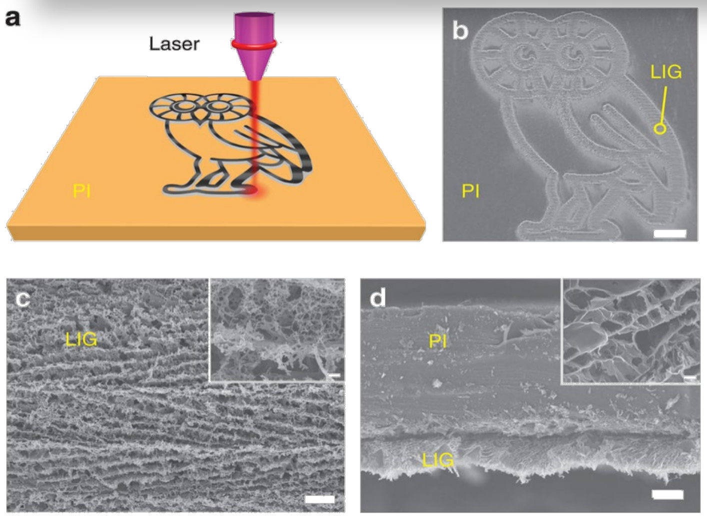
Laser Inscribed Graphene (LIG)



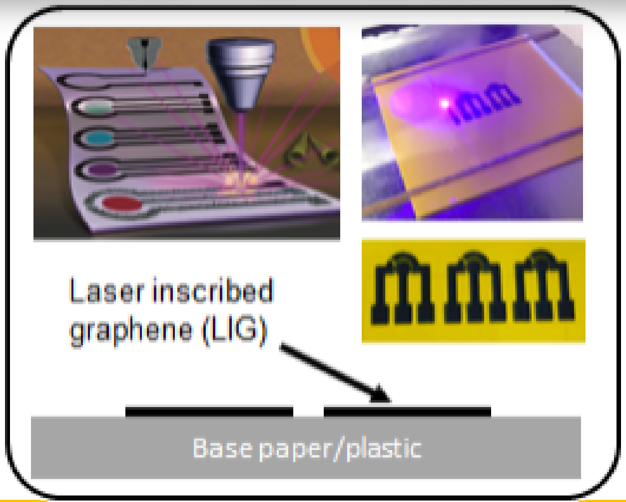
Lin et. al. (2014) develops LIG for supercapacitors

Evolution of Sensors ◆ 3D Printed Nano-bio Sensors

Laser Inscribed Graphene (LIG)



Lin et. al. (2014) develops LIG for supercapacitors

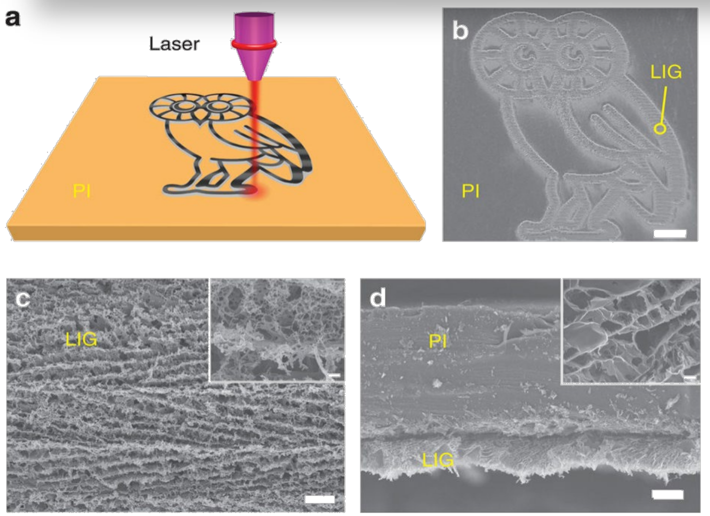


LIG for Sensing

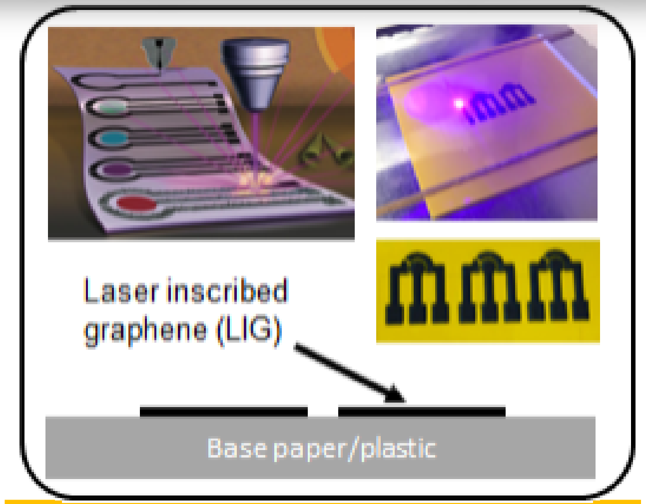
- Sugars, organics
  - Tehrani et. al. (2016); Nayak et. al. (2016); Vanegas et al (2018)
- Ions
  - Garland et al (2019)

Evolution of Sensors ◆ 3D Printed Nano-bio Sensors

Laser Inscribed Graphene (LIG)

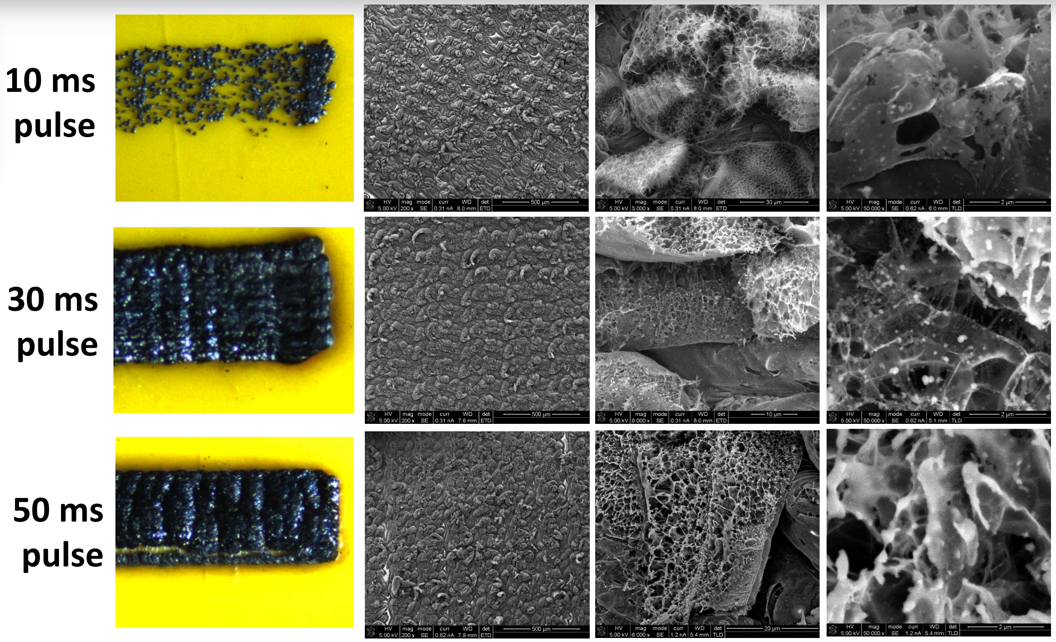


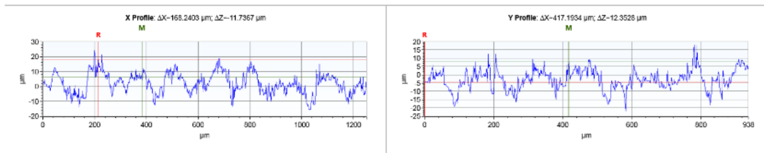
Lin et. al. (2014) develops LIG for supercapacitors

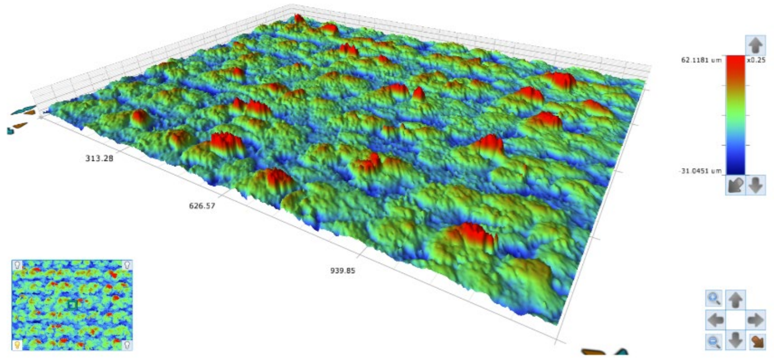


LIG for Sensing

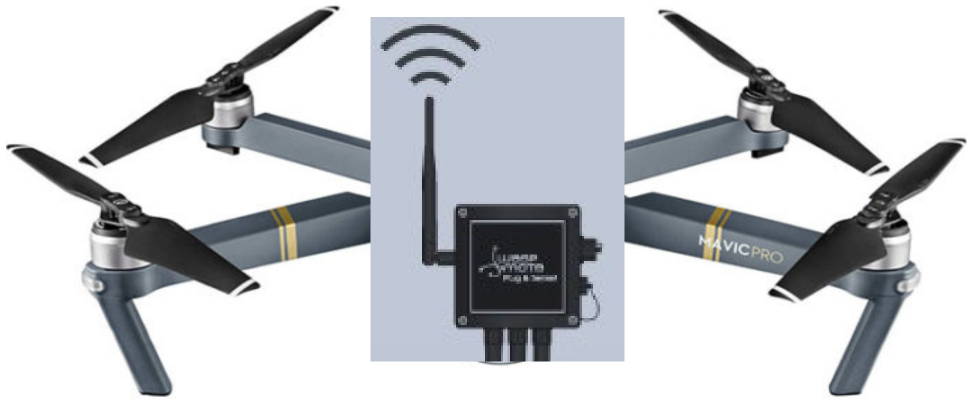
- Sugars, organics
  - Tehrani et. al. (2016); Nayak et. al. (2016); Vanegas et al (2018)
- Ions
  - Garland et al (2019)



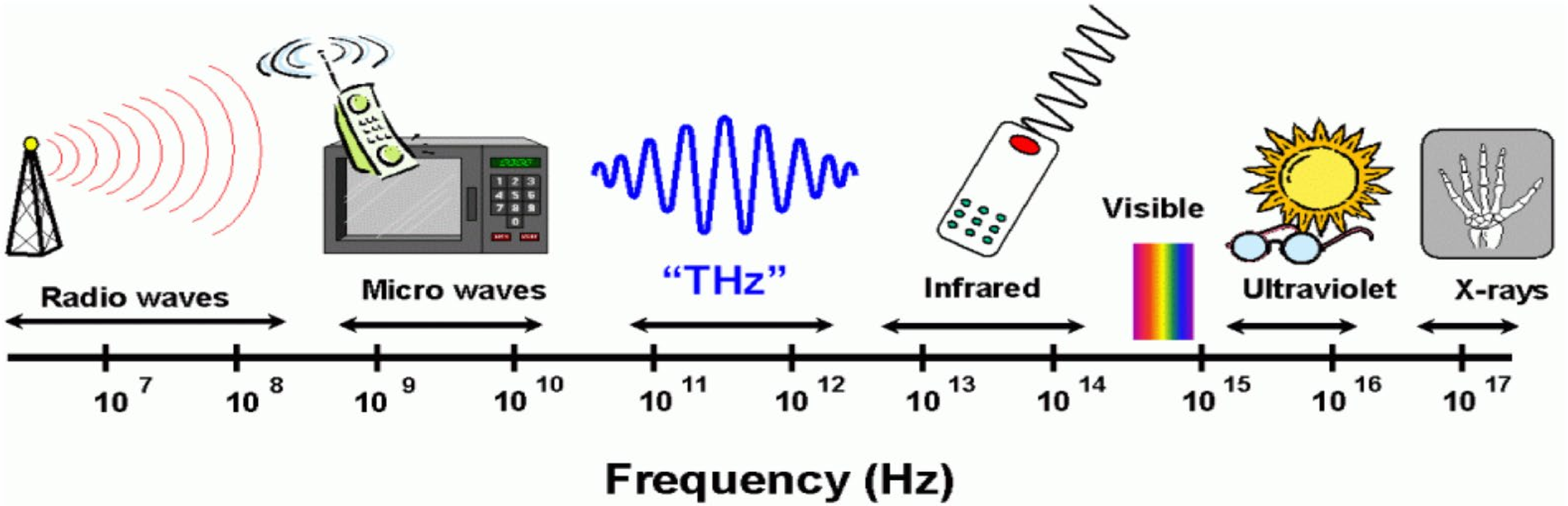




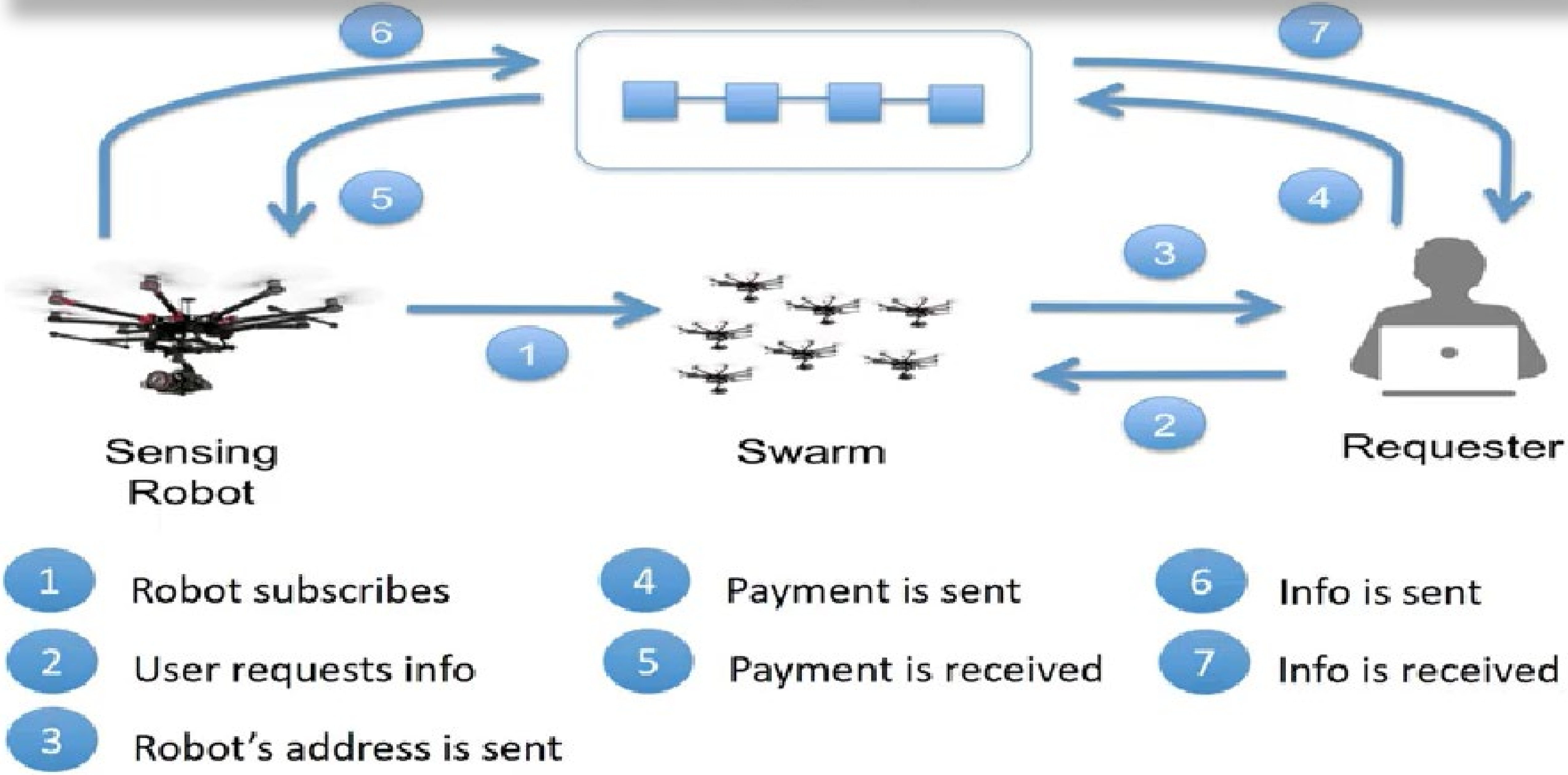
Drones carrying Mobile Sensors acquire Reflected RF Signals



Terahertz (THz) spectrum, occupying frequency range between 0.3 and 3 THz, has potential for transformative applications in communication, sensing, spectroscopy, and imaging due to its desirable properties such as non-ionizing photon energy, penetration capability through optically opaque materials, unique spectral signatures for macro-molecules and chemicals. https://www.nature.com/articles/s41467-019-09868-6



Rent-a-Robot drone-swarm "plug & play" remote sensors to collect desired data

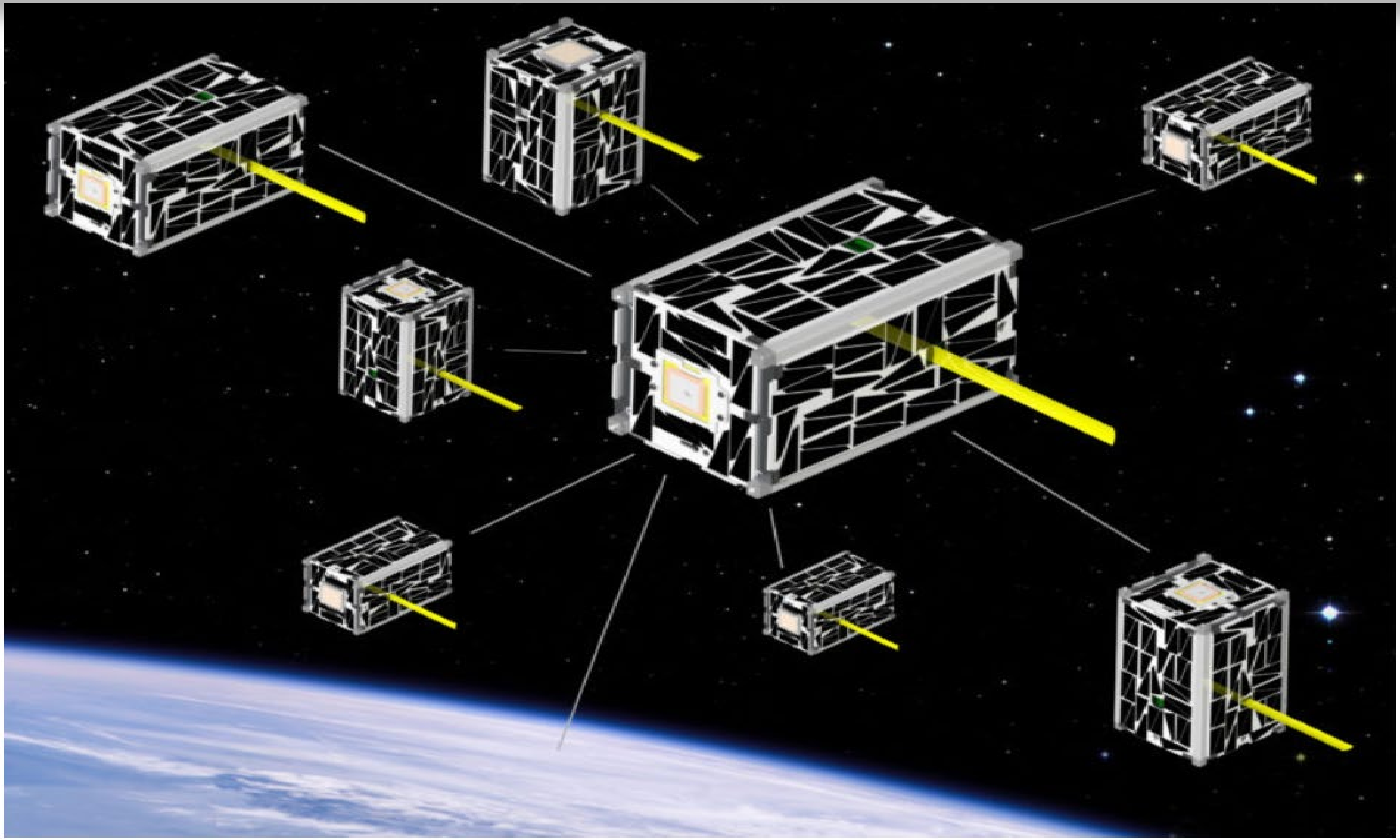


Mobile Sensor Swarms deployed on Drones-on-Demand

Swarm Intelligence, Swarm Robotics, Mobile Robotics

https://dl.acm.org/citation.cfm?id=37406; https://link.springer.com/chapter/10.1007%2F978-3-642-58069-7_38; https://vlsicad.ucsd.edu/Publications/Journals/j26.pdf

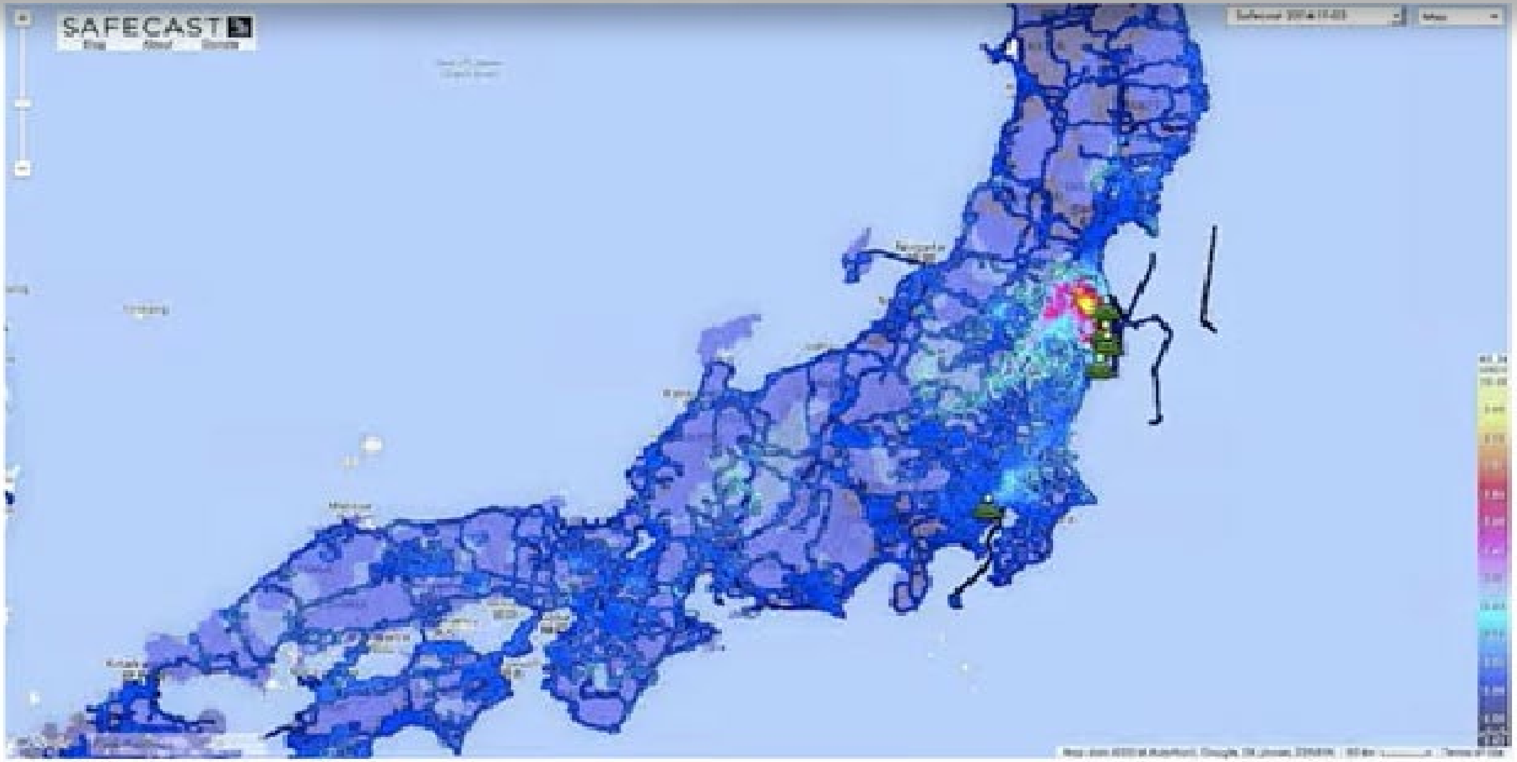
Sensor Swarm



NASA and Lockheed Martin have been studying how small satellites could be knit together into a distributed swarm. (NASA Illustration)

Crowd-sourced radiation sensor data after the explosion Fukushima Daiichi Nuclear Power Plant in Ōkuma (2013)

Safecast Nano - Mobile Radiation Monitoring
Wireless Geiger Counter, GPS with Bluetooth



SAFECAST

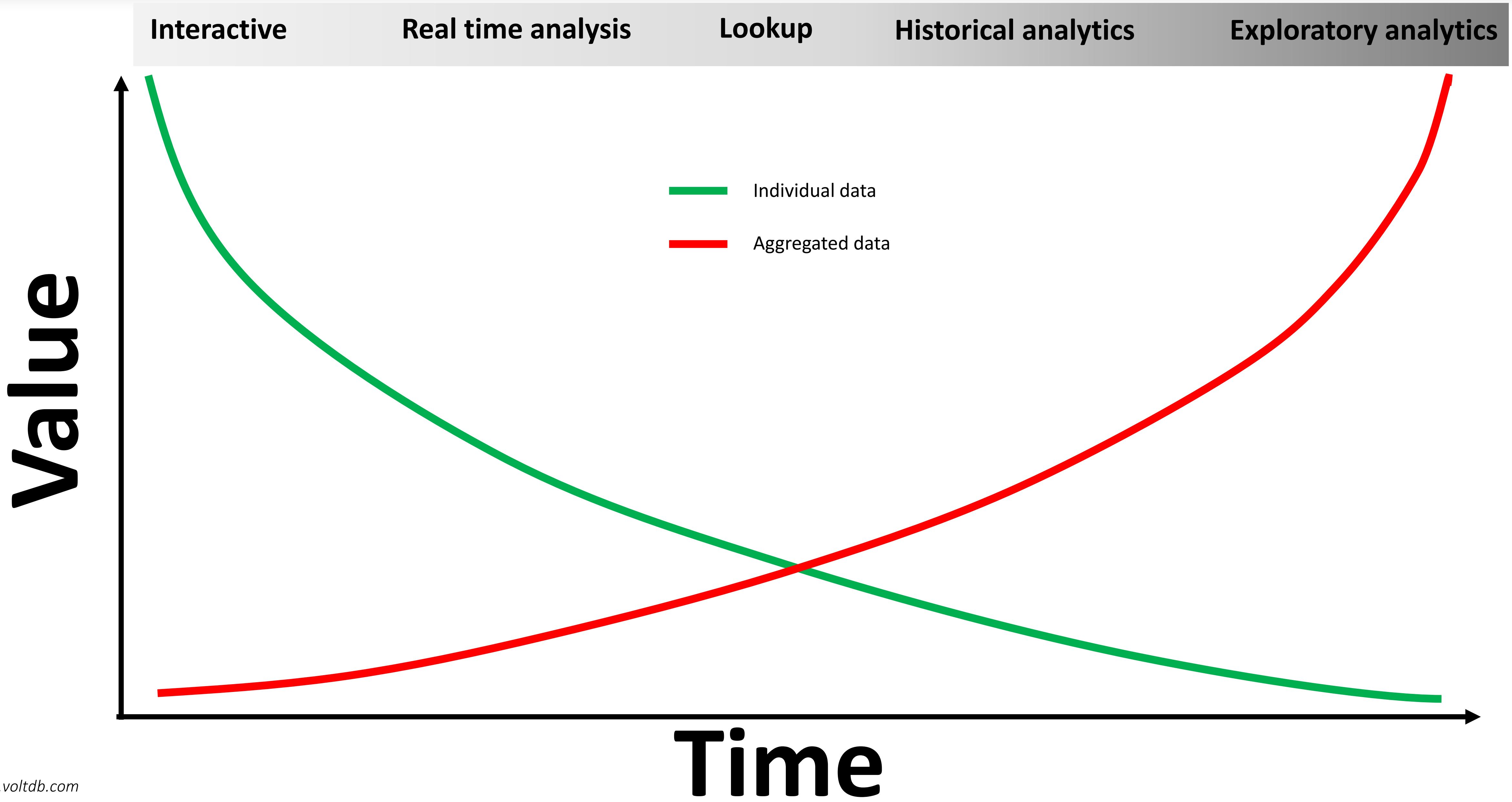
mit media lab

http://bit.ly/EDUARDO-CASTELLO

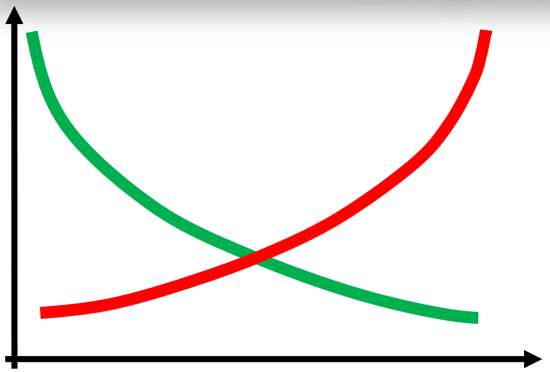
Source: Google Images Labeled for Reuse

DATA from SENSORS

Is data perishable?

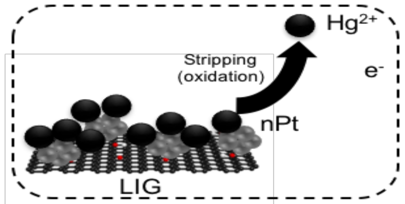


Varying Time-Sensitivity of Data from Sensors

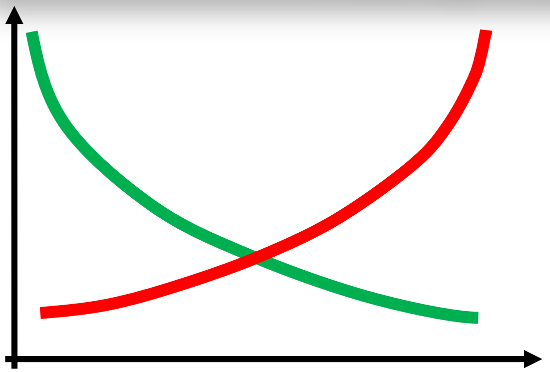


Persistent

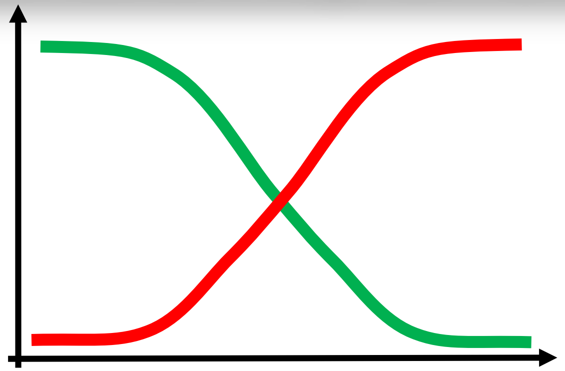
- Soil health
    - Erosion, degradation
- Agrochemical runoff
    - Nutrients, agrochemicals
- Land use change
    - Deforestation, mining



Varying Time-Sensitivity of Data from Sensors

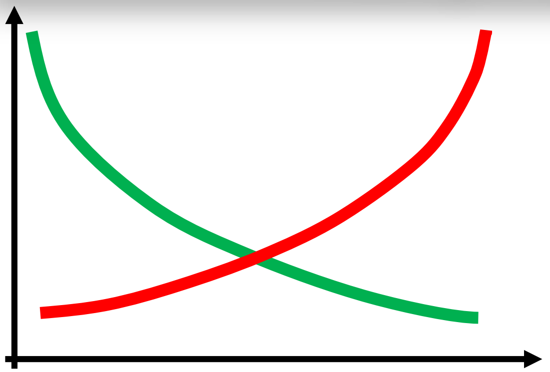


Persistent

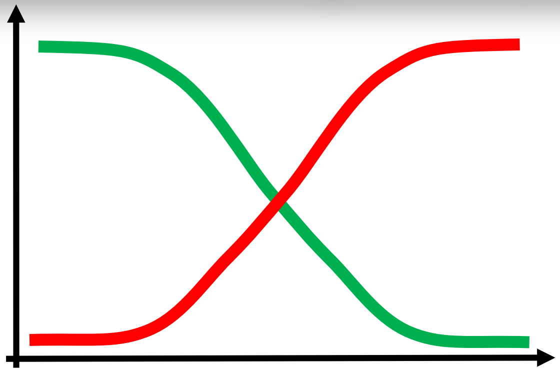


Urgent

Varying Time-Sensitivity of Data from Sensors

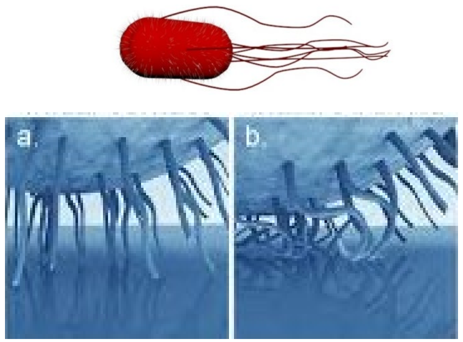


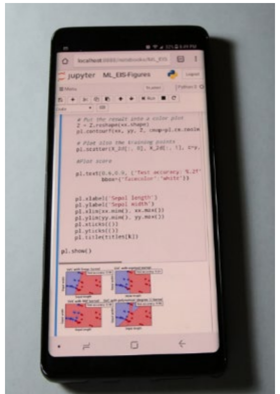
Persistent



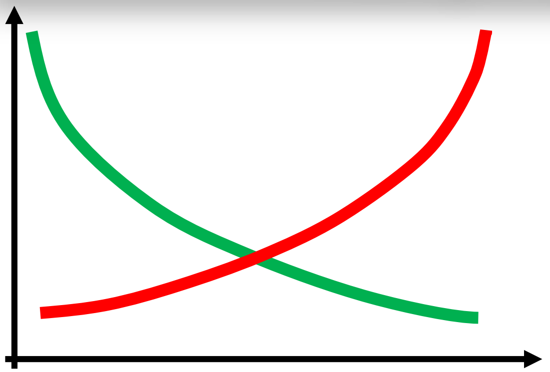
Urgent

- Water scarcity
  - Quantity, quality
- Climate change
- Solid waste/wastewater
  - Pathogens, heavy metals

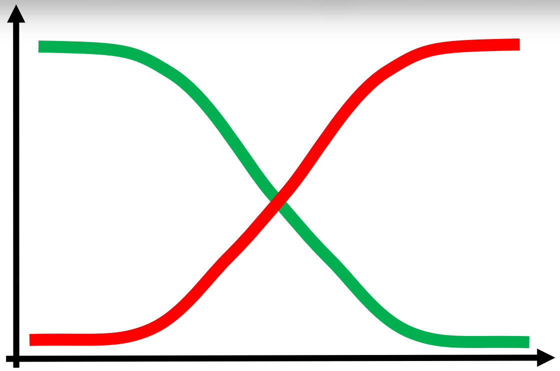




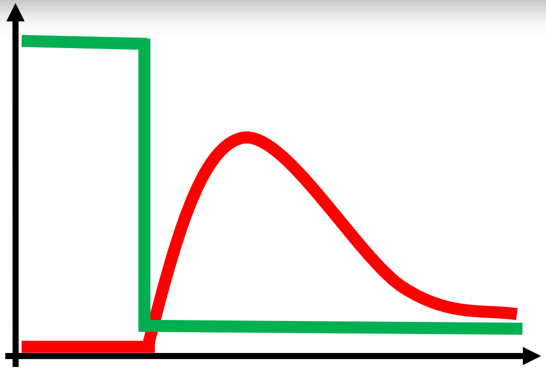
Varying Time-Sensitivity of Data from Sensors



Persistent

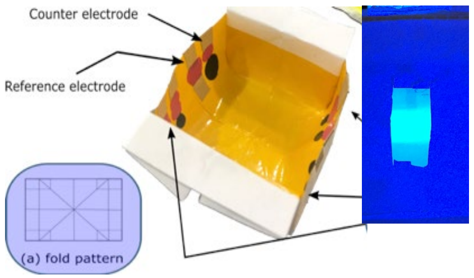


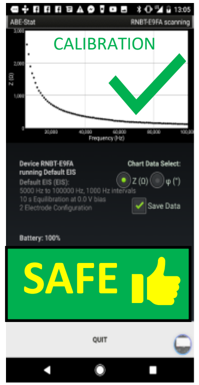
Urgent



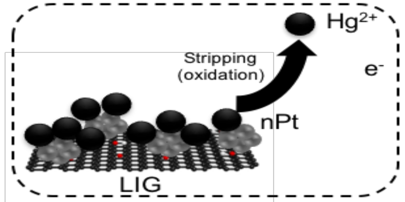
Triage

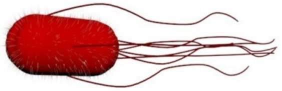
- Natural disaster
- Attack on infrastructure

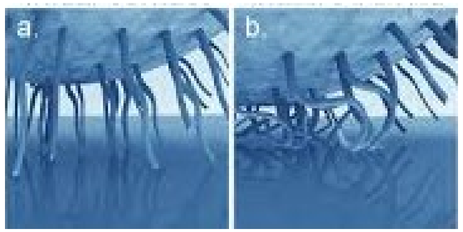


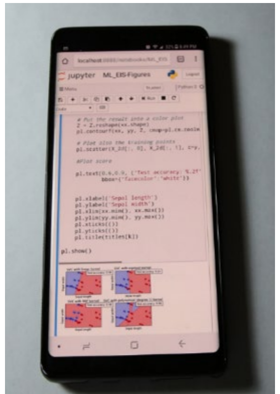


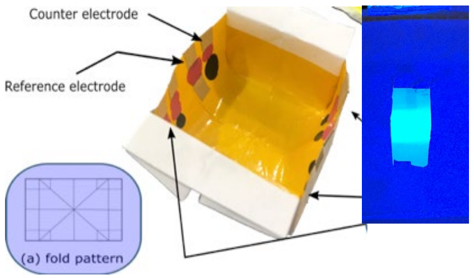
Data from the Perspective of Sensors

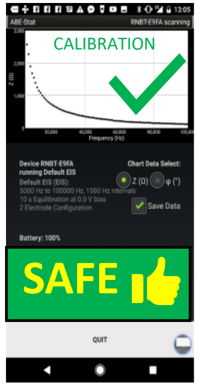




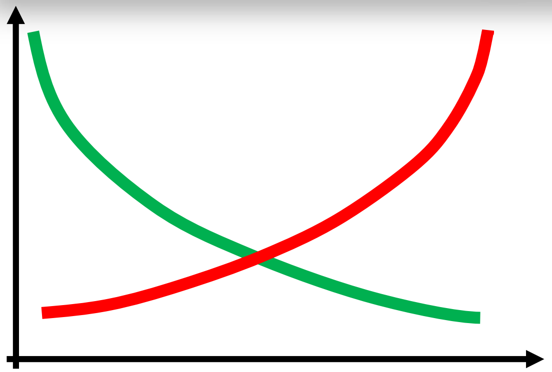




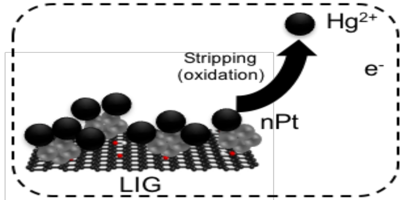


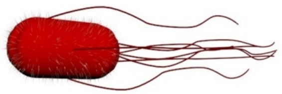


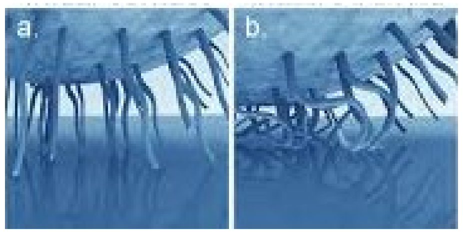
Data from the Perspective of Sensors

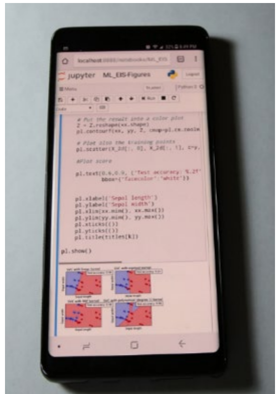


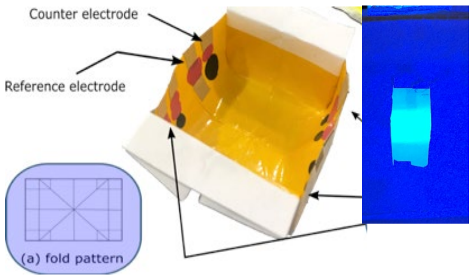
Persistent

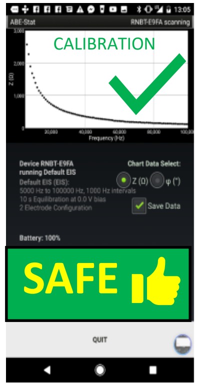




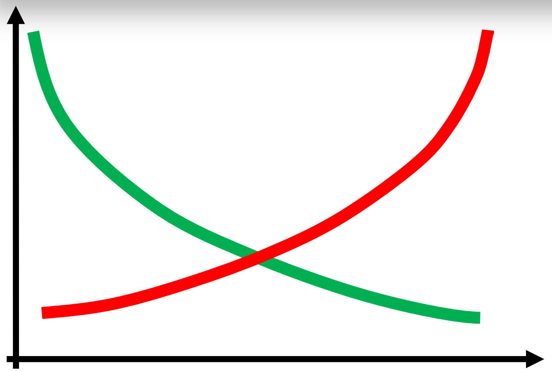




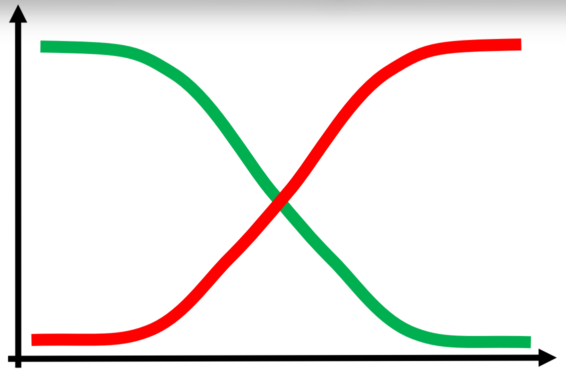




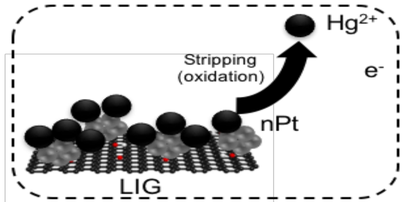
Data from the Perspective of Sensors

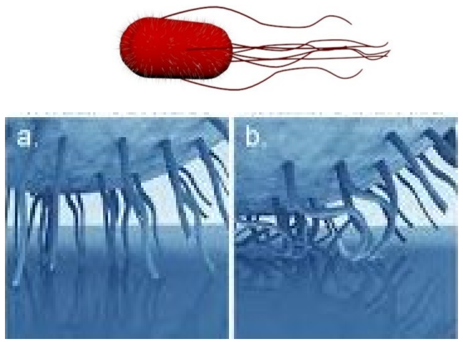


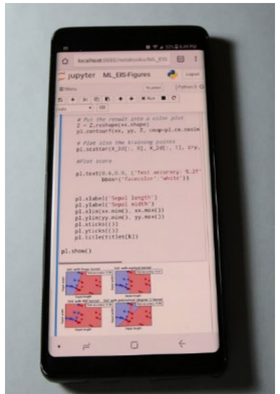
Persistent

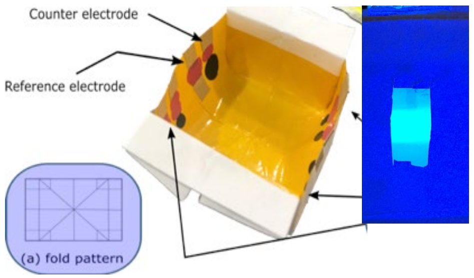


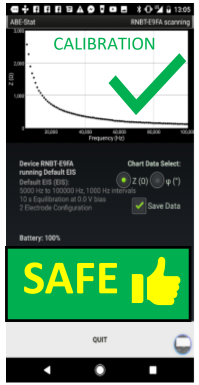
Urgent



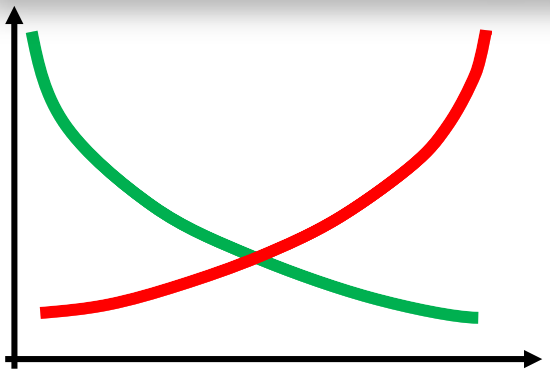




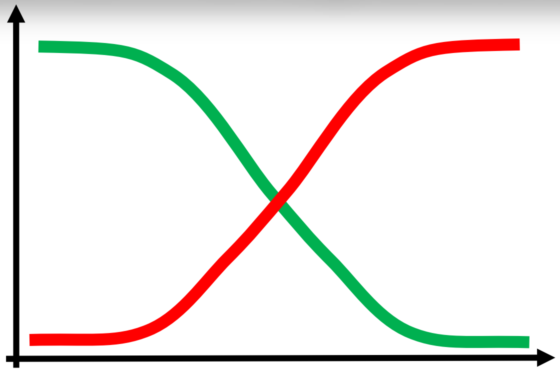




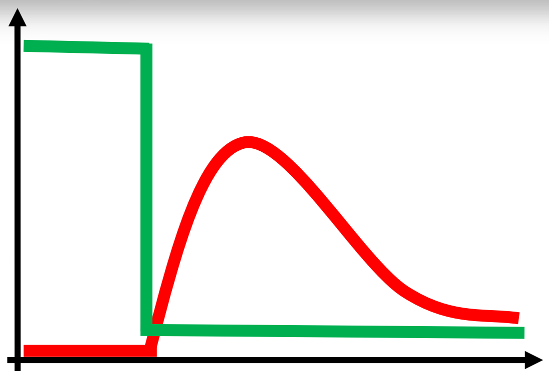
Data from the Perspective of Sensors



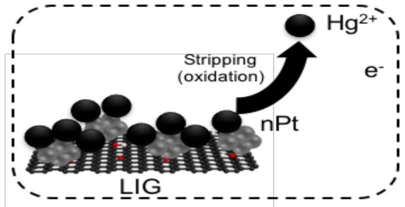
Persistent

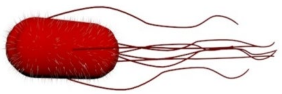


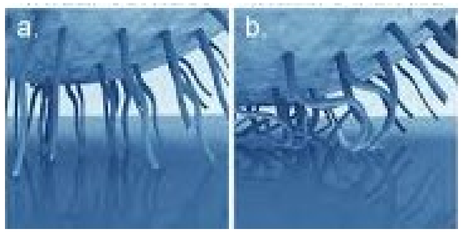
Urgent

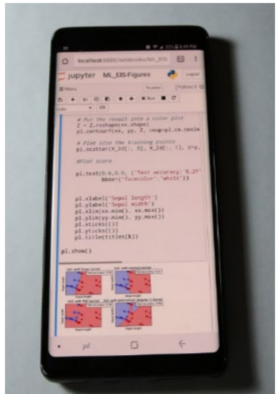


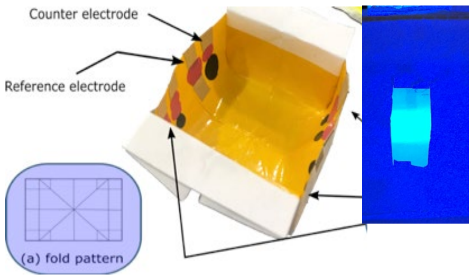
Triage

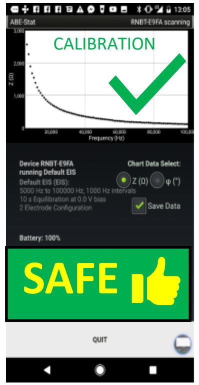




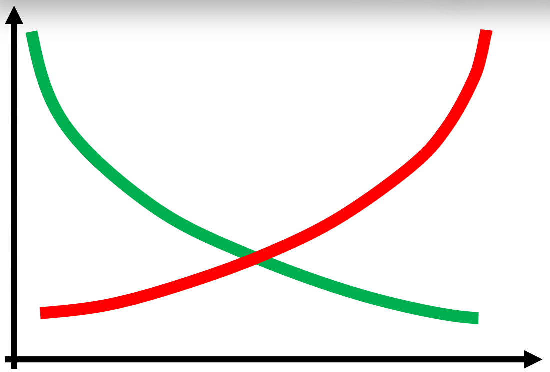






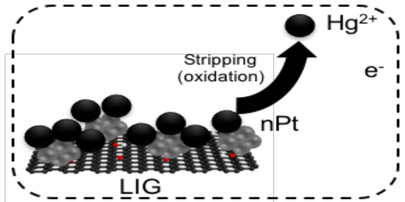


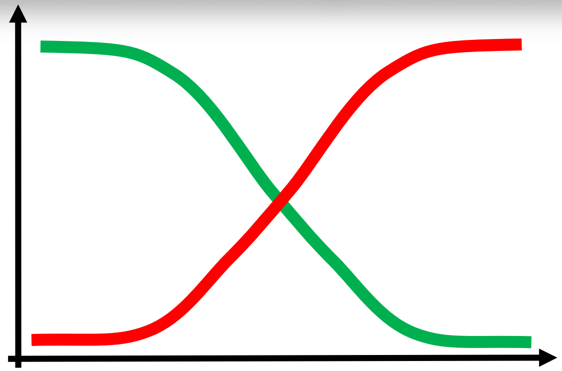
Varying Time-Sensitivity of Data from Sensors



Persistent

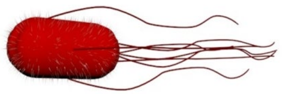
- Soil health
    - Erosion, degradation
- Agrochemical runoff
    - Nutrients, agrochemicals
- Land use change
    - Deforestation, mining

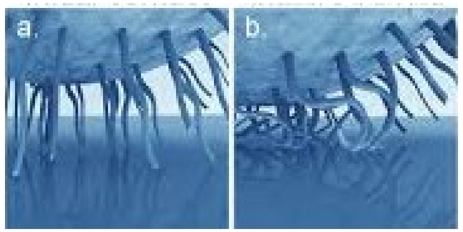


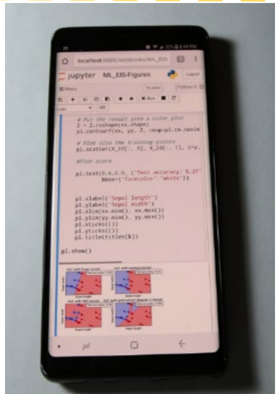


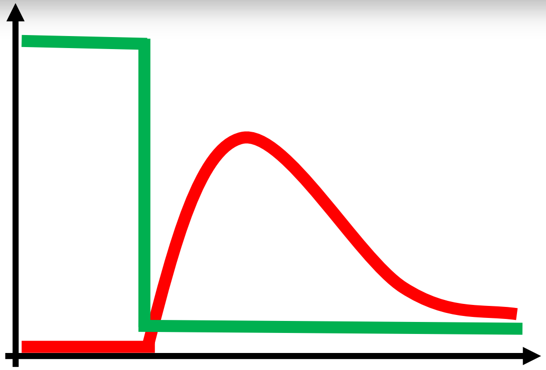
Urgent

- Water scarcity
    - Quantity, quality
- Climate change
- Solid waste/wastewater
    - Pathogens, heavy metals



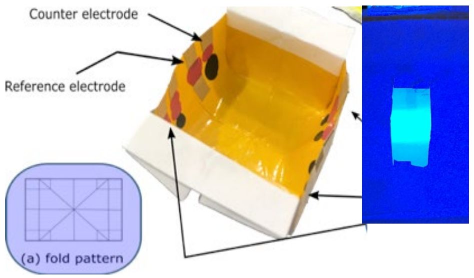


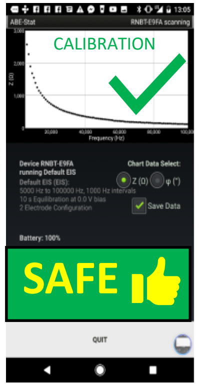




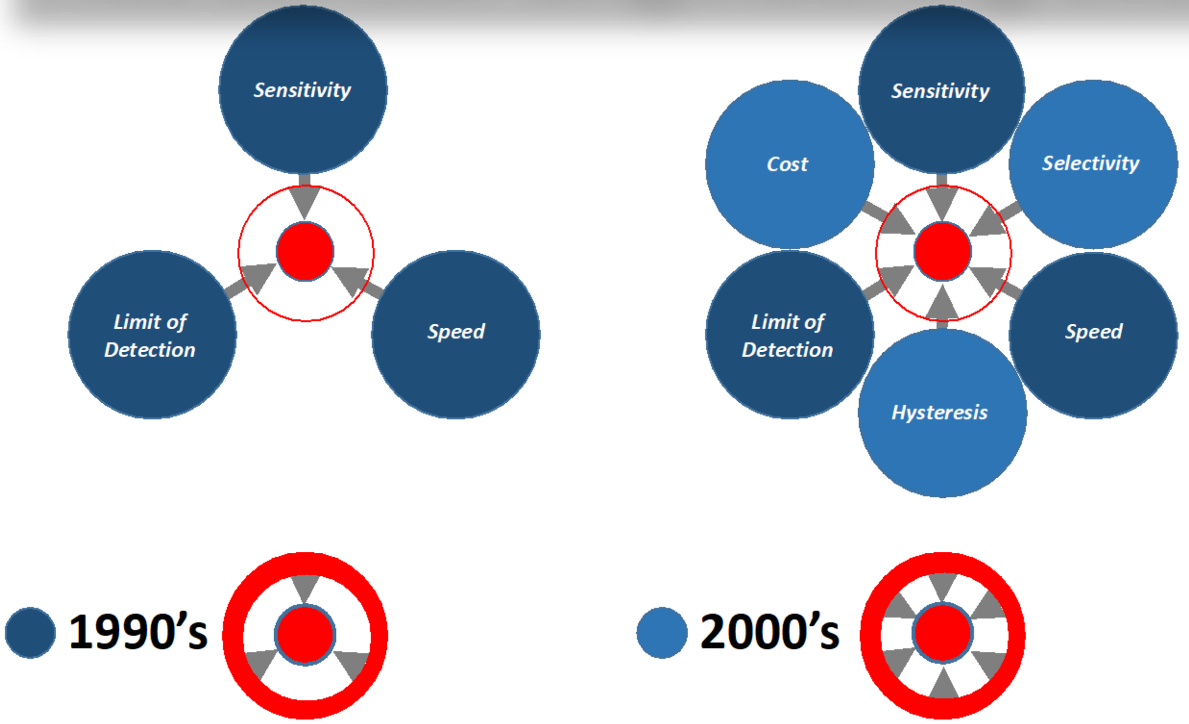
Triage

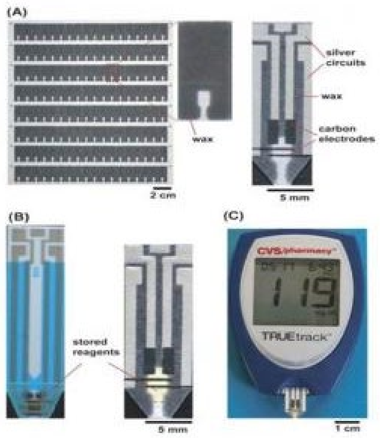
- Natural disaster
- Attack on infrastructure

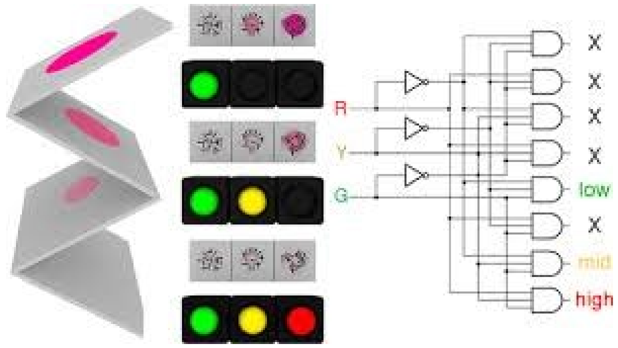




Has sensor engineering evolved with digital transformation

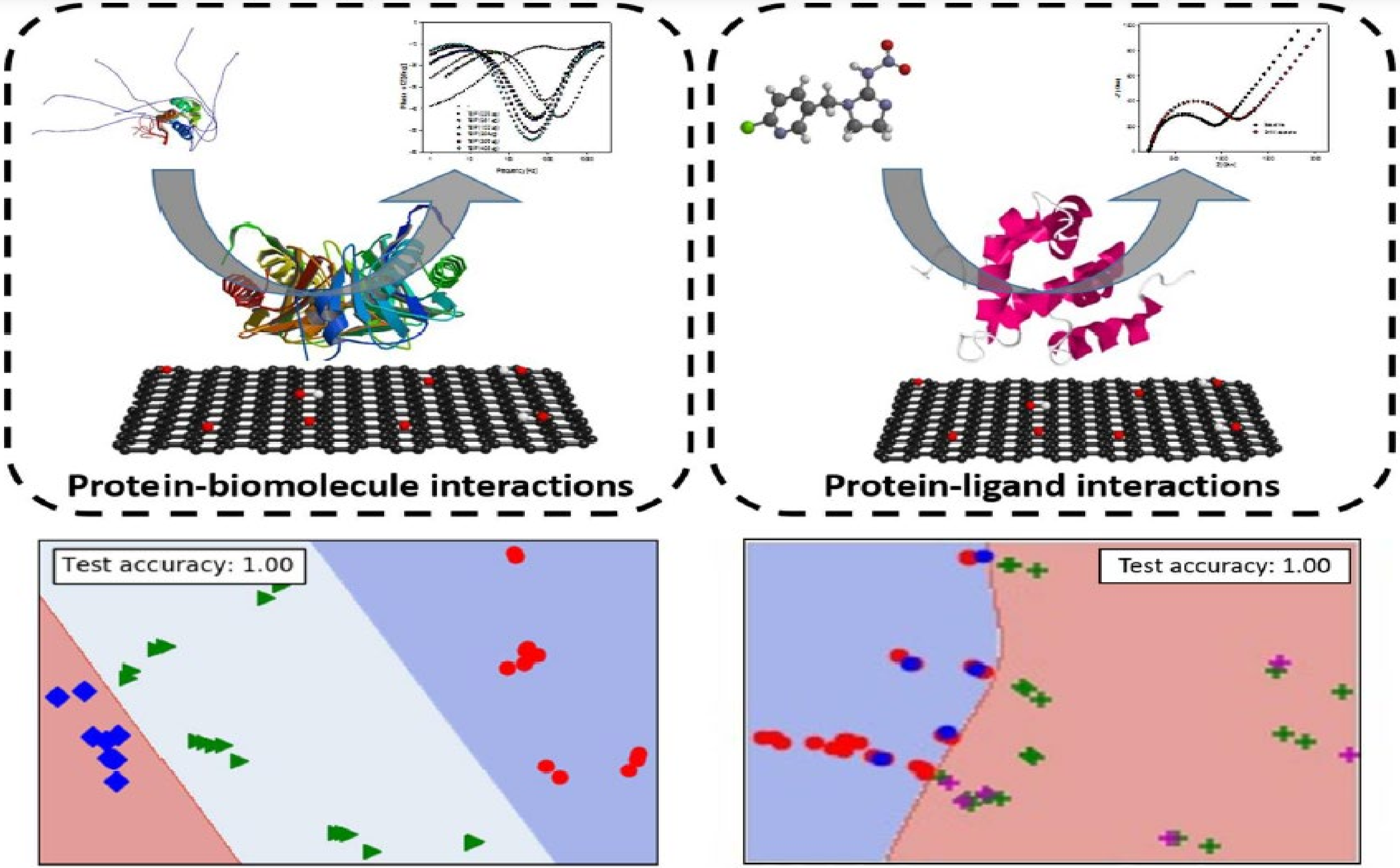






Original innovation

Water Polluted by Mercury (Hg)



Professor Eric Scott McLamore, University of Florida ● www.ncbi.nlm.nih.gov/pubmed/29629449

Microbial Contamination of Water

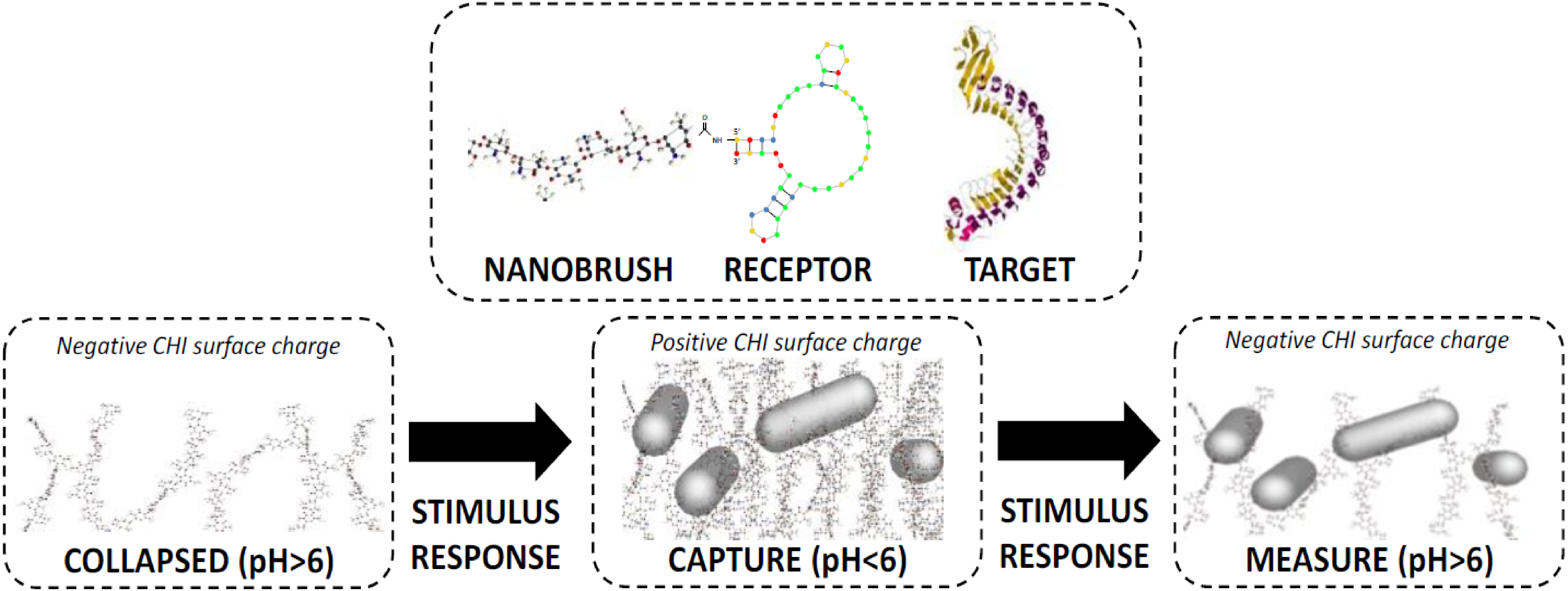


Image courtesy of Hills et al (2018), Analyst, 143(7): 1650-1661.

Professor Eric Scott McLamore, University of Florida ● www.ncbi.nlm.nih.gov/pubmed/29541704

Water Contamination Data Analysis

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.

Professor Eric McLamore, University of Florida ● www.ncbi.nlm.nih.gov/pubmed/29629449

What is the value of this data analysis to the user ?

What is the value of this data analysis to the user ?

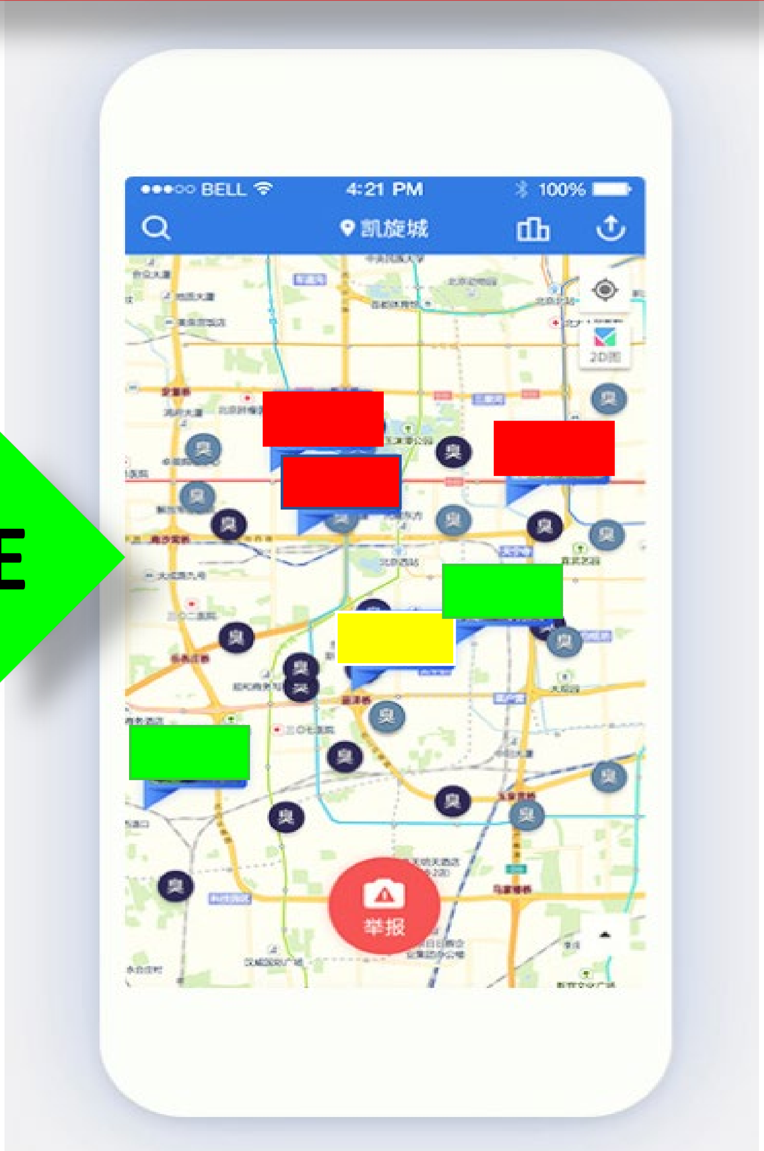
0

Z E R O

Professor Eric McLamore, University of Florida ● www.ncbi.nlm.nih.gov/pubmed/29629449

IoT-by-design: Data Analytics of Value to End-User

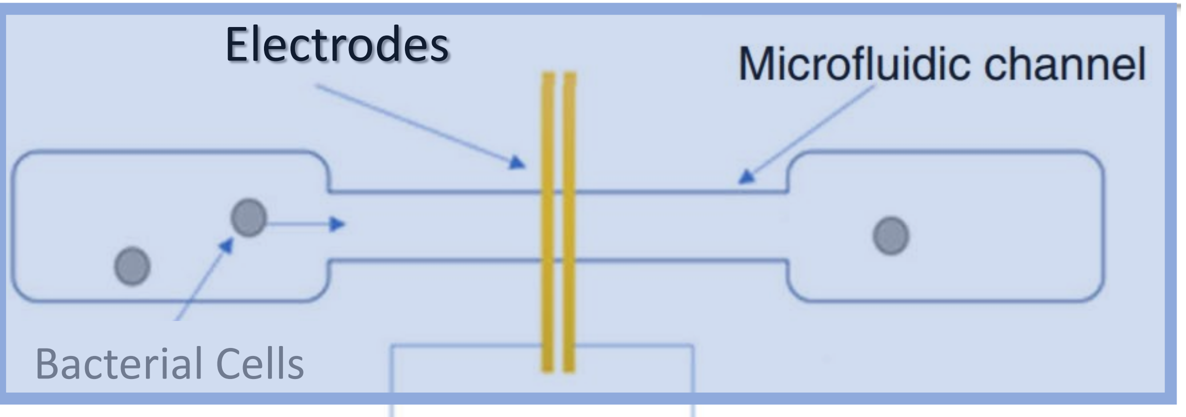
VALUE



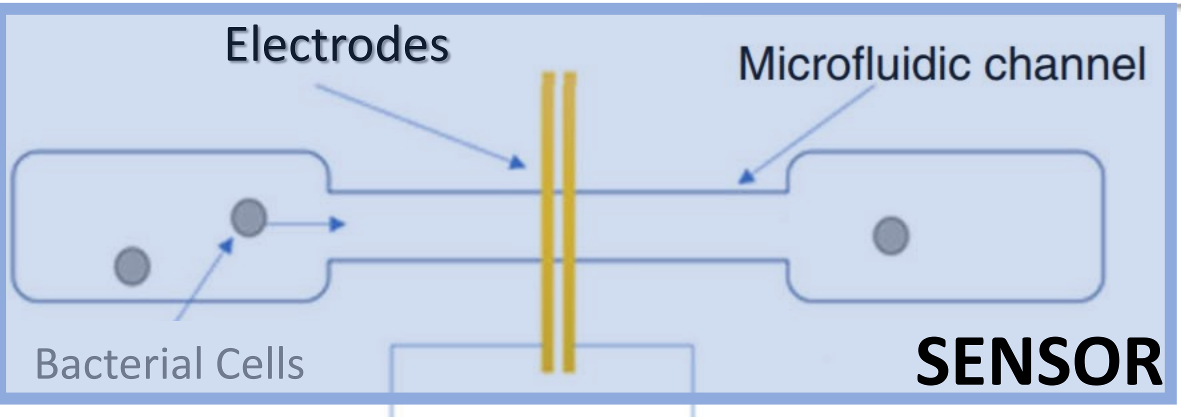
Professor Eric McLamore, University of Florida - www.ncbi.nlm.nih.gov/pubmed/29629449

Bacterial Cells

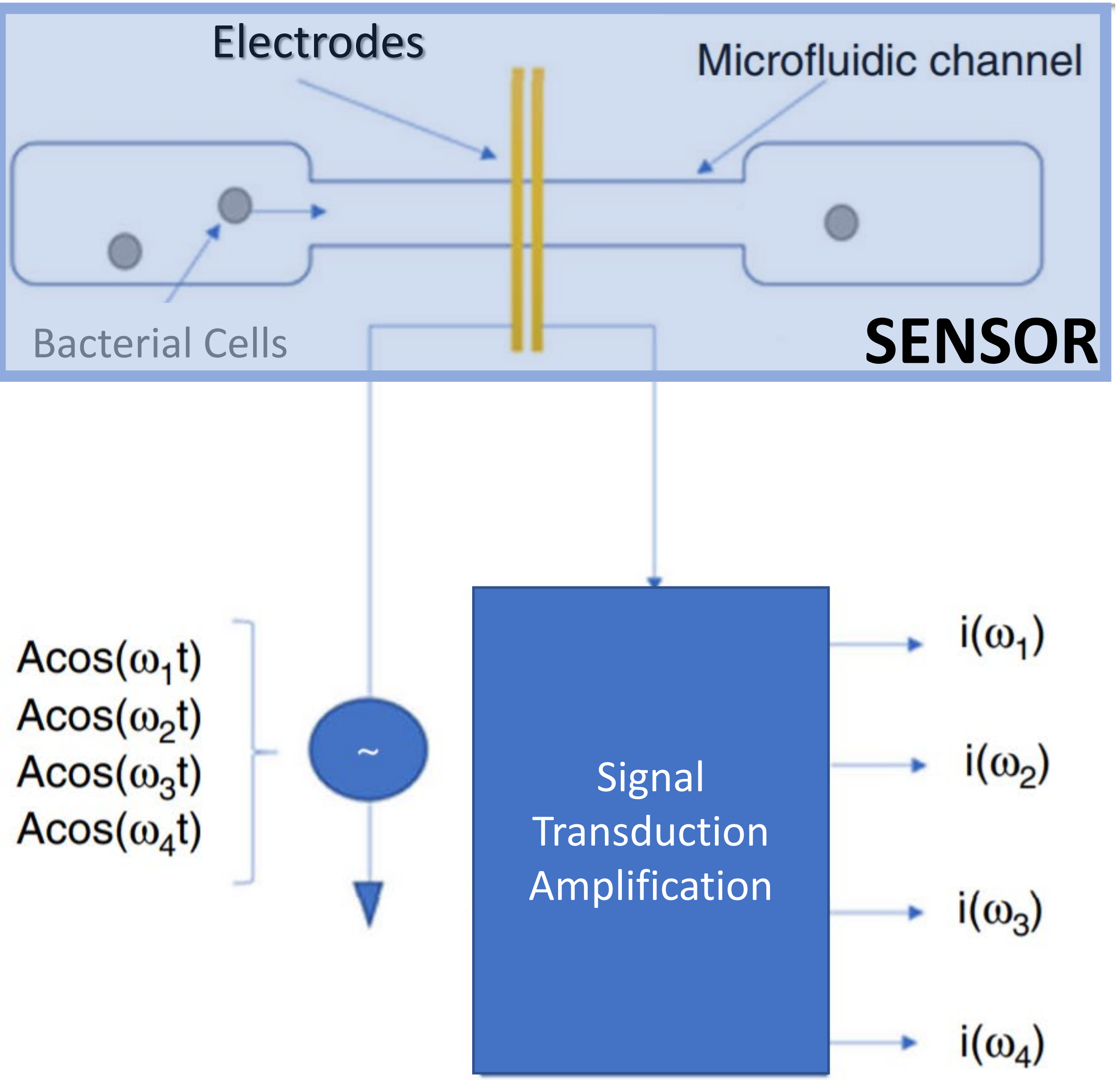
Near Real-Time Analytics: Data-Informed Services at the Edge



Near Real-Time Analytics: Data-Informed Services at the Edge

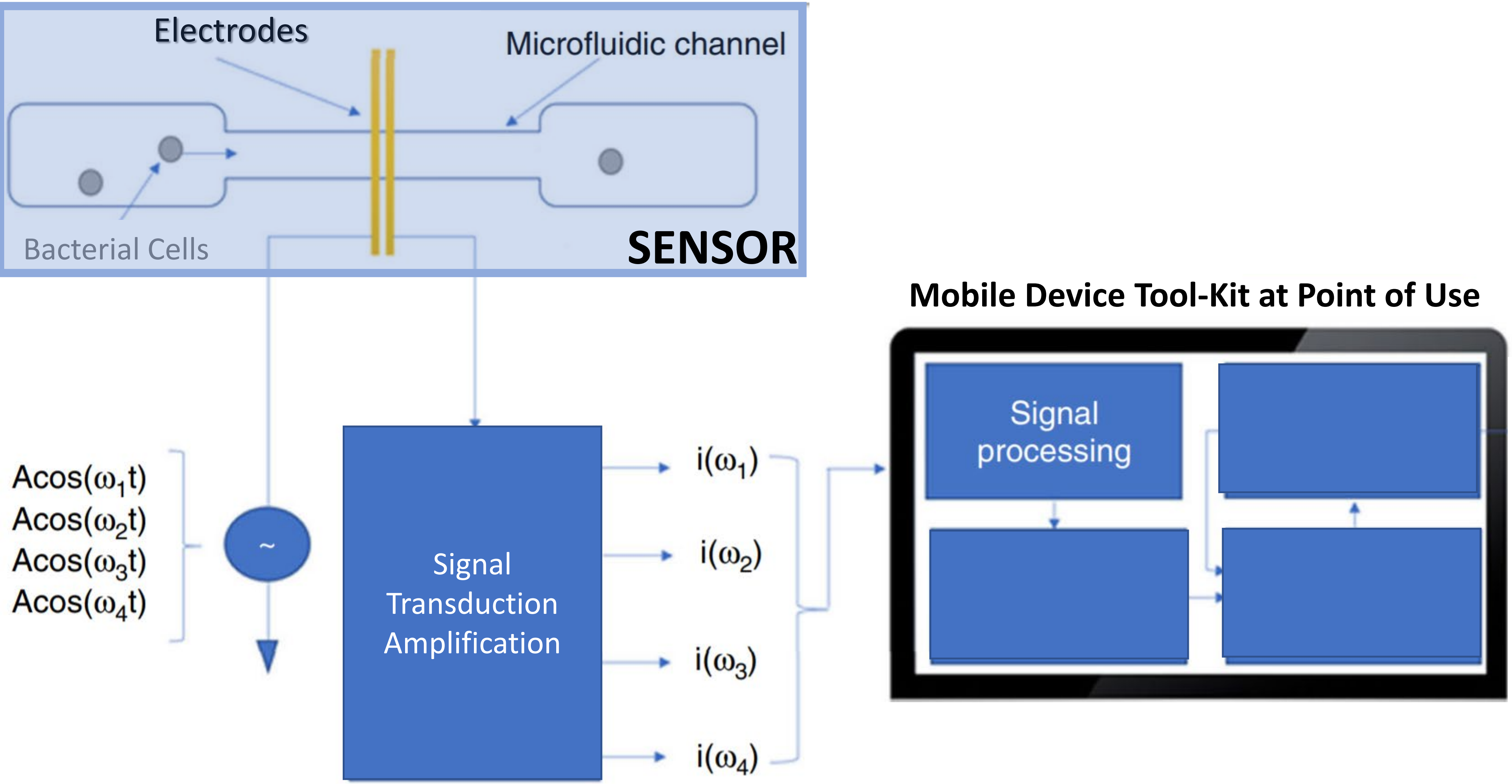


Near Real-Time Analytics: Data-Informed Services at the Edge



Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

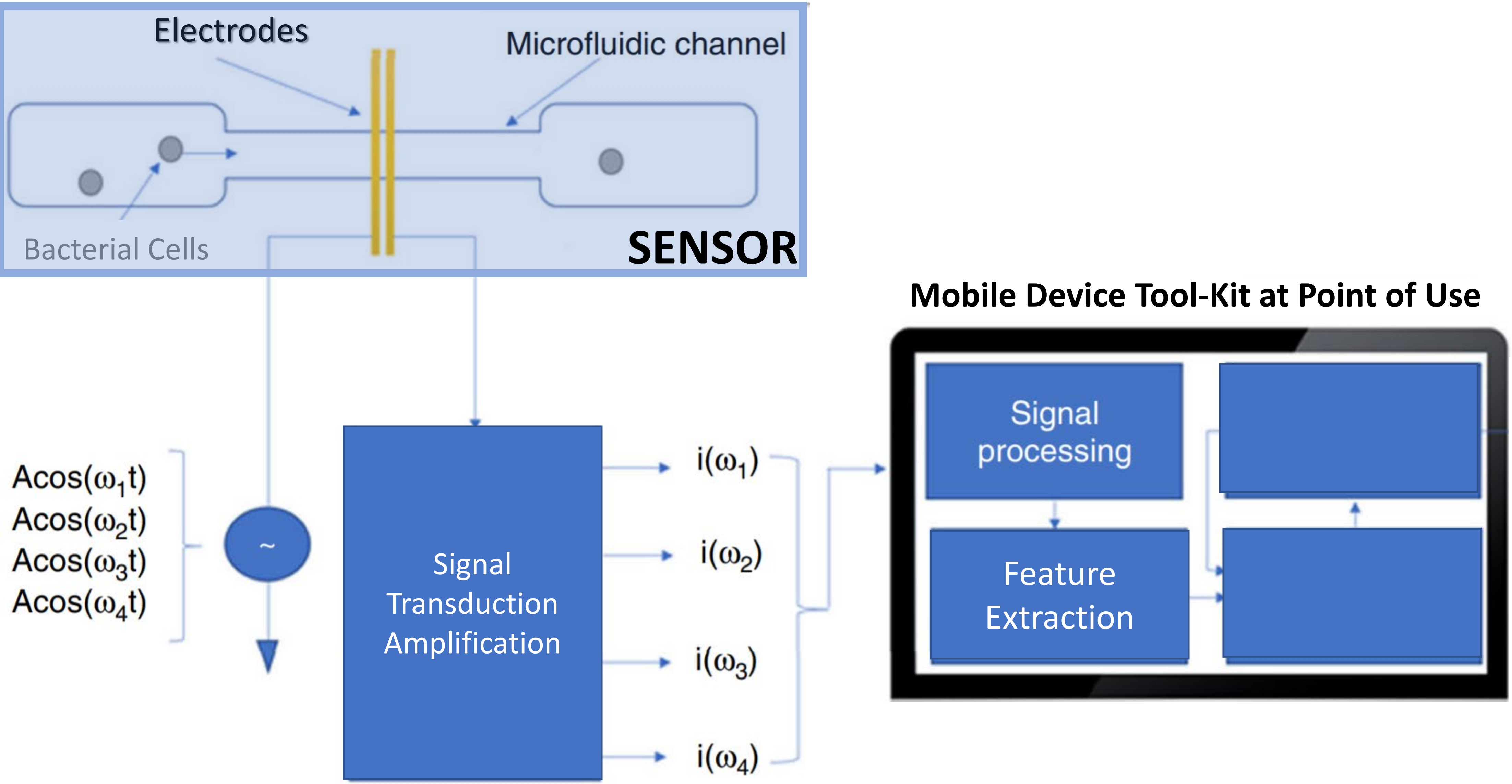
Near Real-Time Analytics: Data-Informed Services at the Edge



Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34

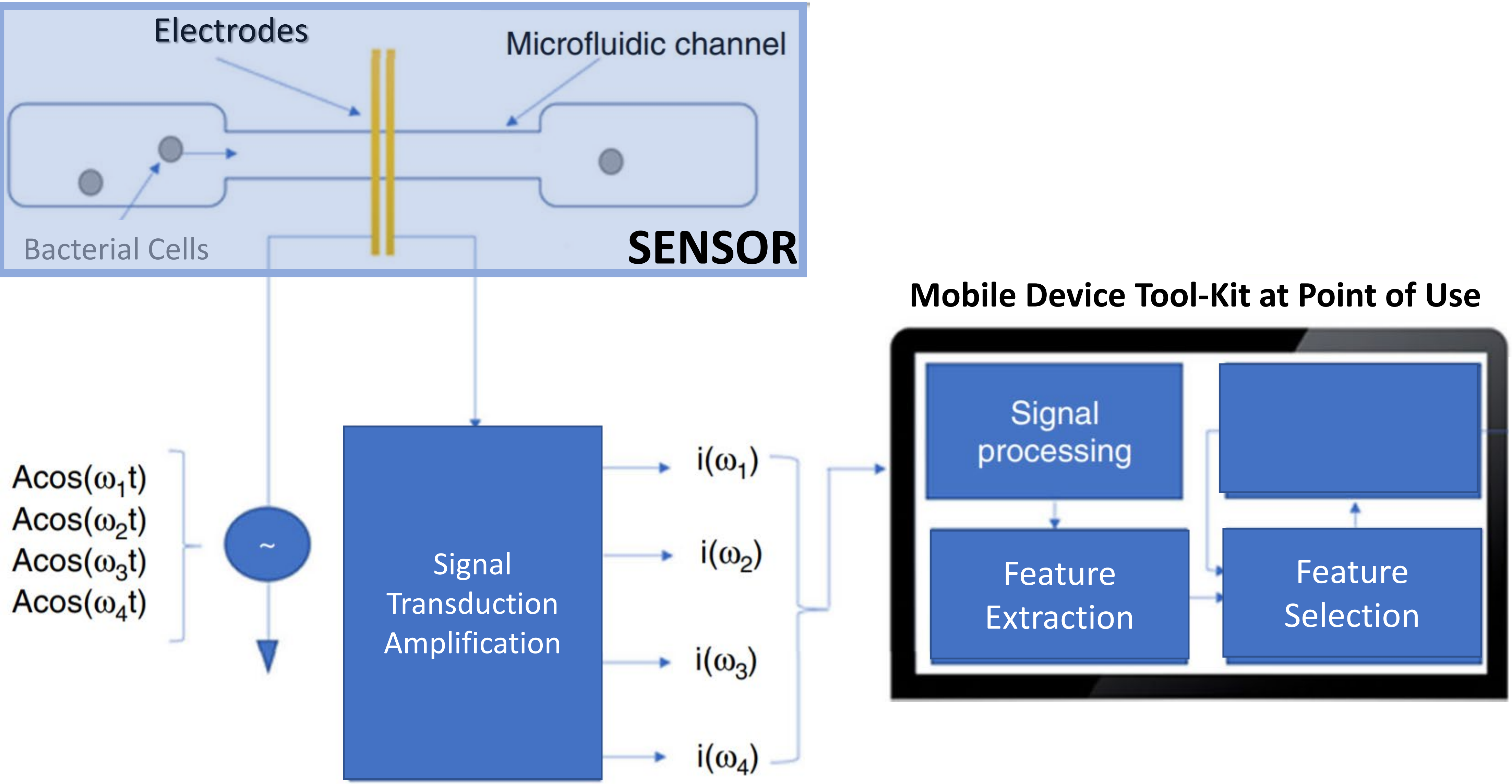
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

Near Real-Time Analytics: Data-Informed Services at the Edge



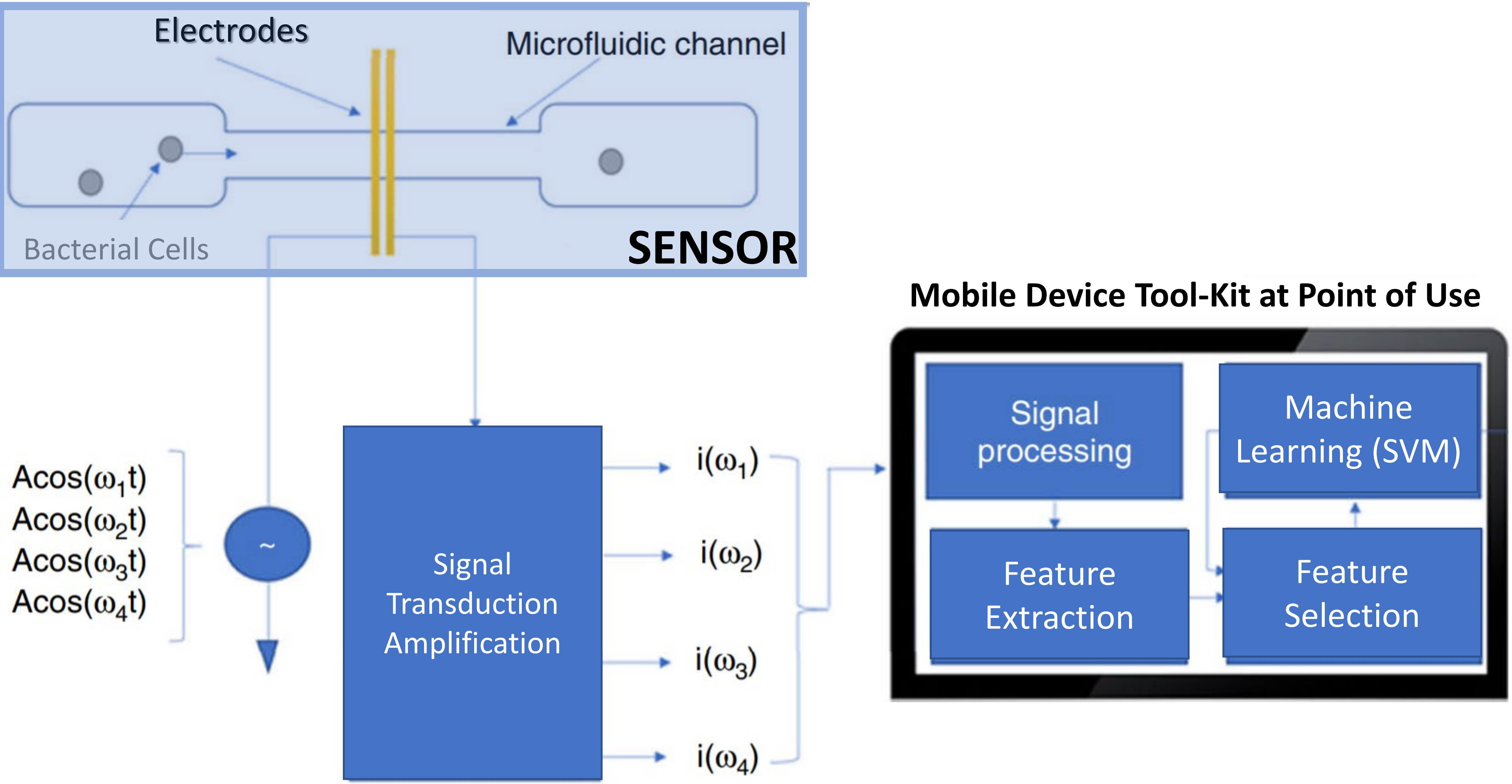
Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

Near Real-Time Analytics: Data-Informed Services at the Edge



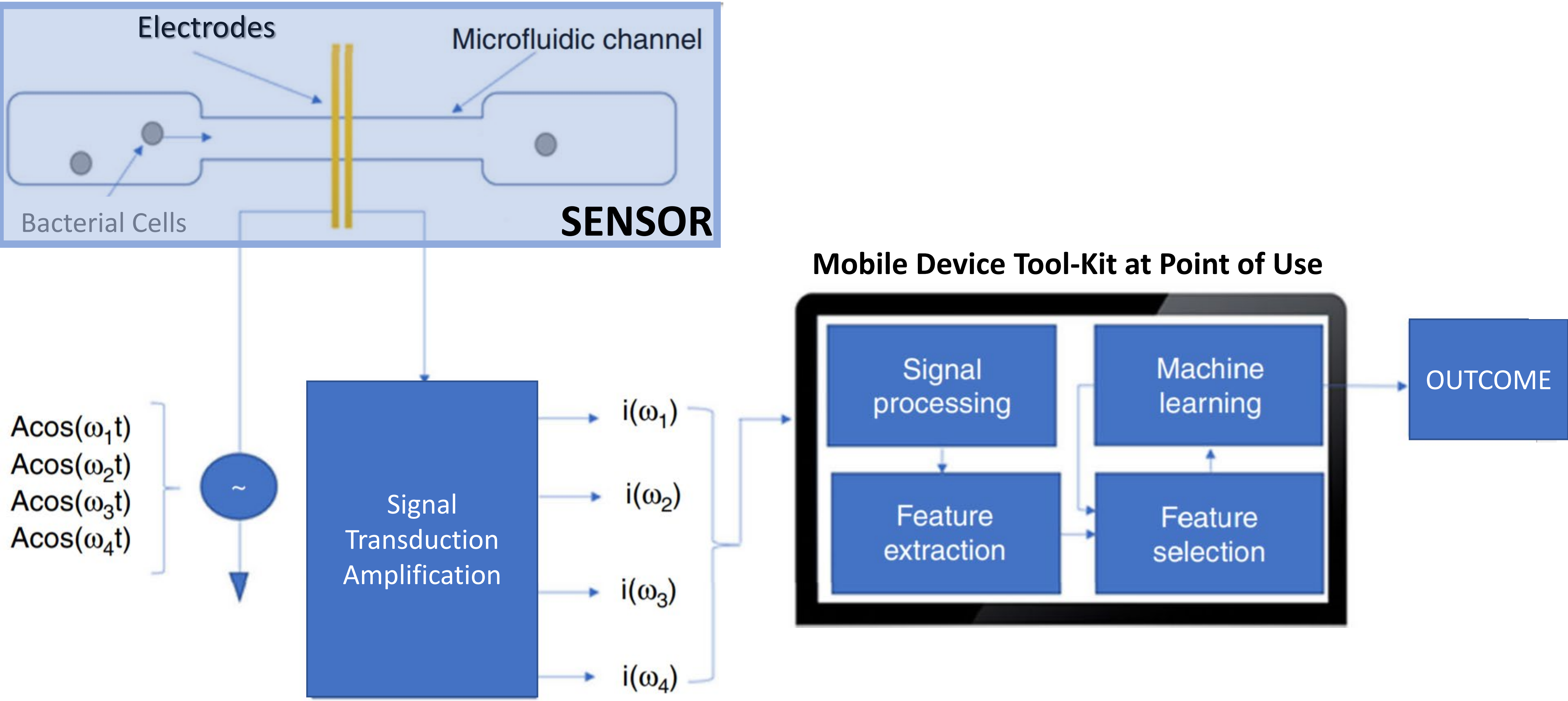
Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

Near Real-Time Analytics: Data-Informed Services at the Edge



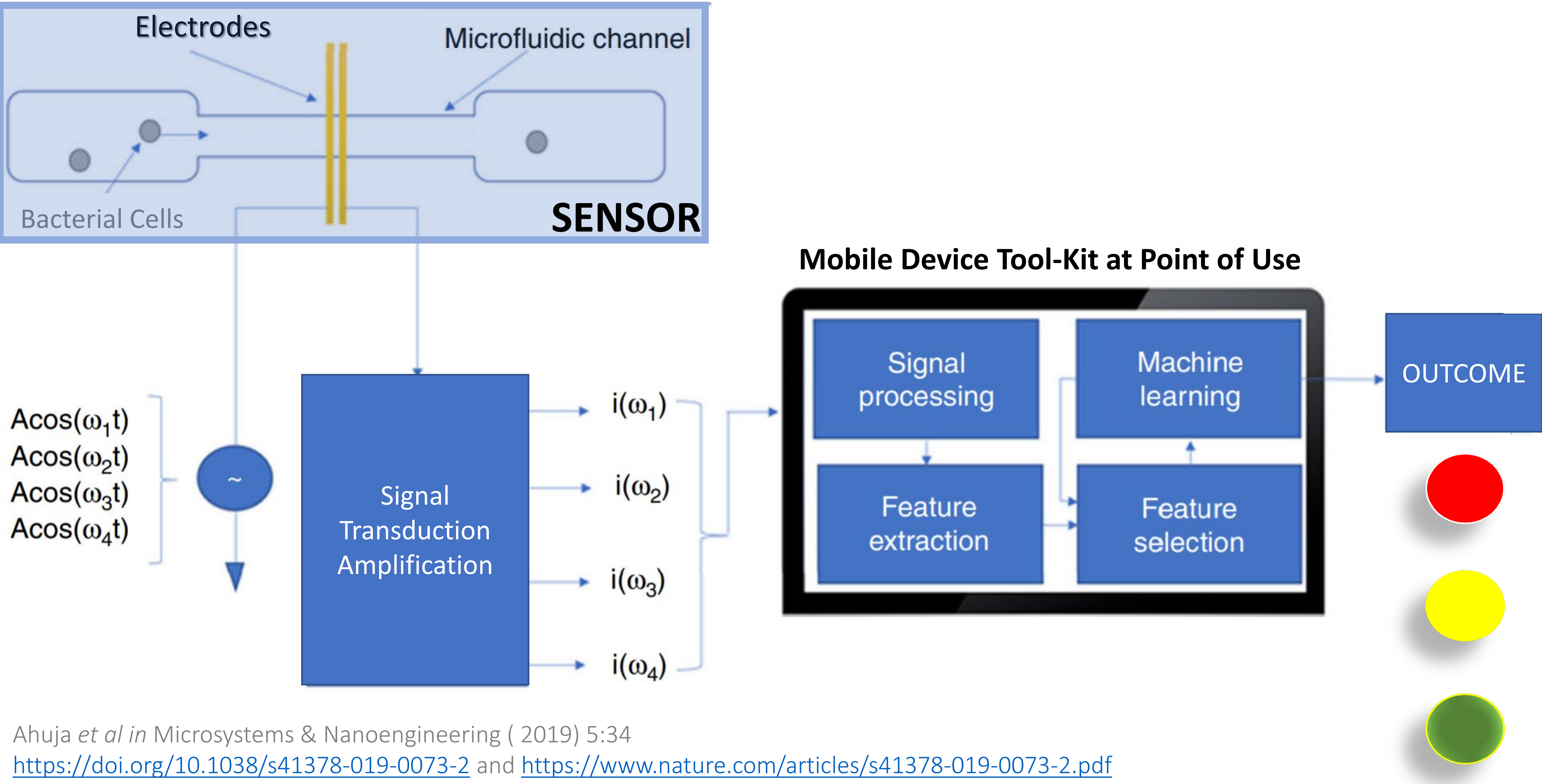
Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

Near Real-Time Analytics: Data-Informed Services at the Edge



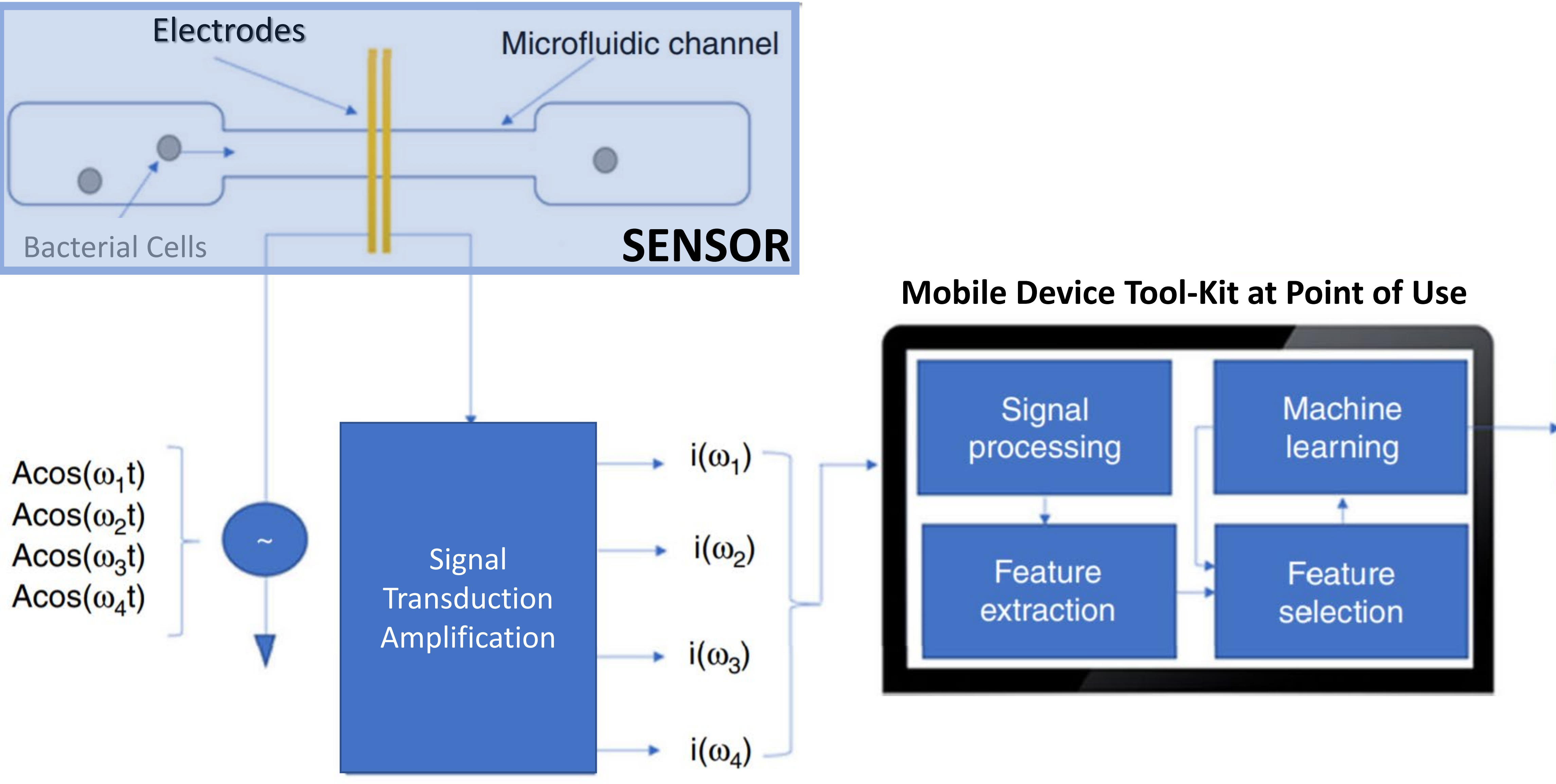
Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

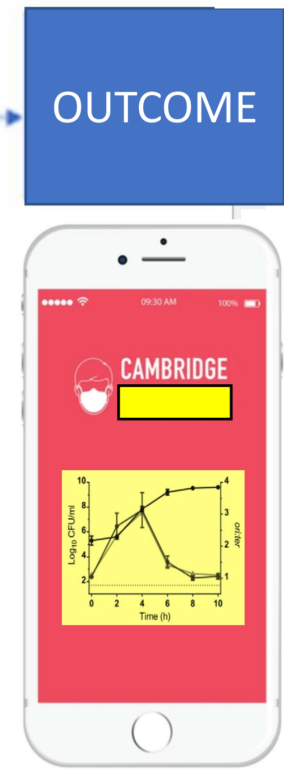
Near Real-Time Analytics: Data-Informed Services at the Edge



Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

Near Real-Time Analytics: Data-Informed Services at the Edge

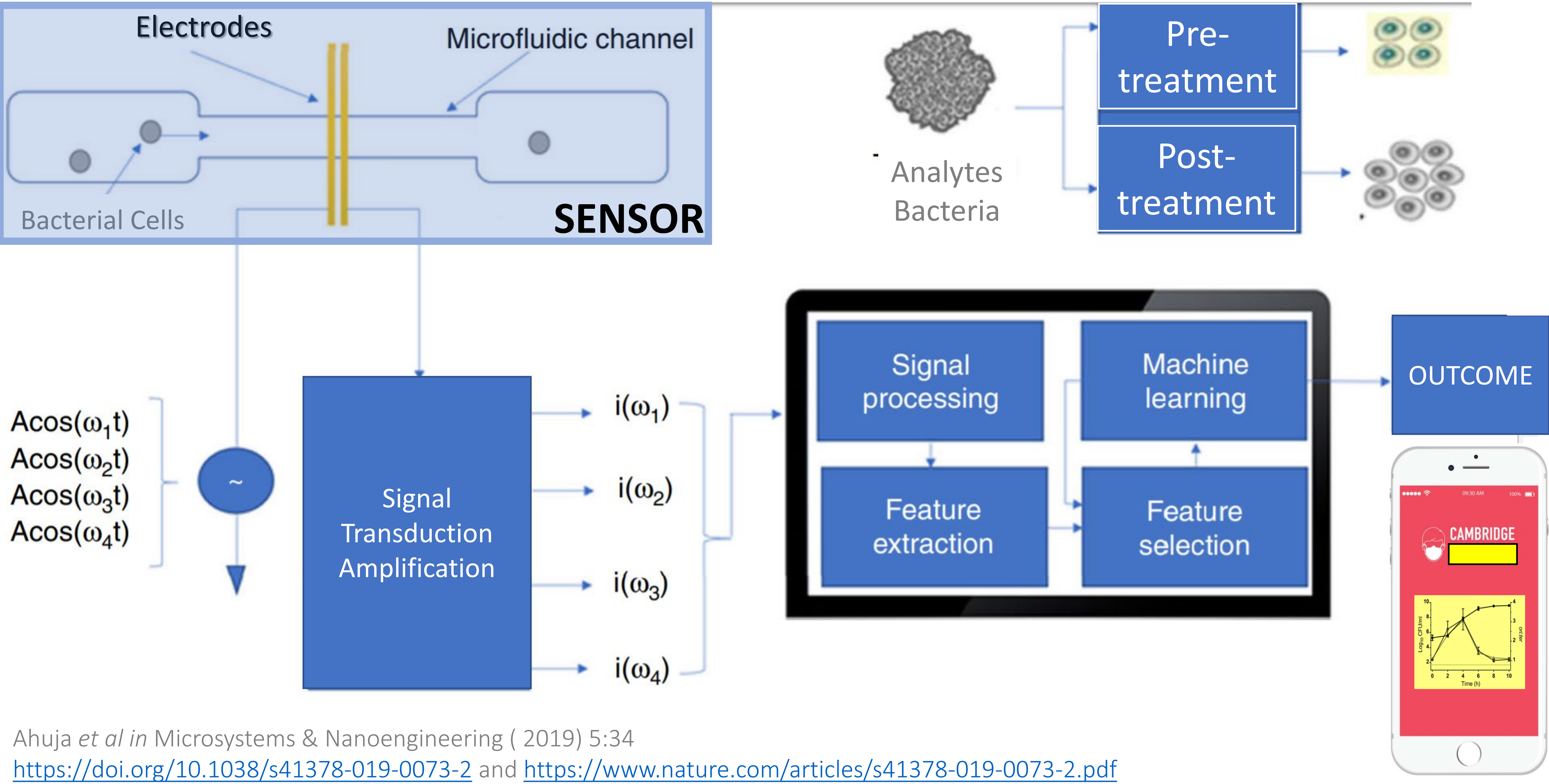




Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34

https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

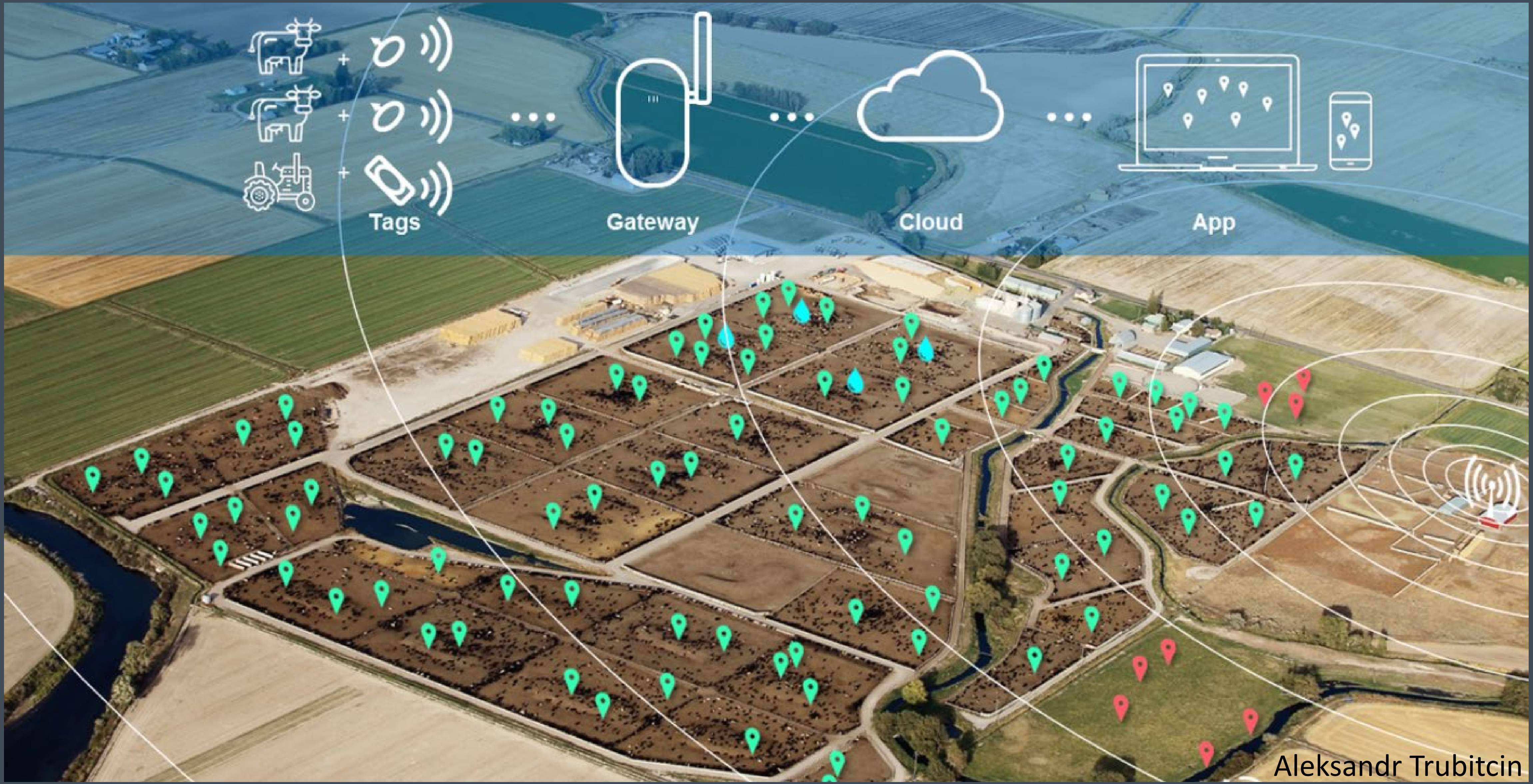
Near Real-Time Analytics: Data-Informed Services at the Edge



Ahuja *et al in* Microsystems & Nanoengineering (2019) 5:34
https://doi.org/10.1038/s41378-019-0073-2 and https://www.nature.com/articles/s41378-019-0073-2.pdf

IoT-by-design: Data Analytics of Value to End-User

See SIGNALS - https://dspace.mit.edu/handle/1721.1/111021



LoRaWAN Cattle Ear Tags: 200 acre feed-lot for 10,000 cows

IoT-BY-DESIGN: PAY A PENNY PER UNIT (PAPPU) PARADIGM ?

See SIGNALS - https://dspace.mit.edu/handle/1721.1/111021

Fill in the details of your deployment.

Install Address

421 N 3200 E, Lewisville, ID 8343

Install Environment: Rural

Gateway Height: 50 ft

Select Gateway: Field 64c

SF 7 | SF 8 | SF 9 | SF 10

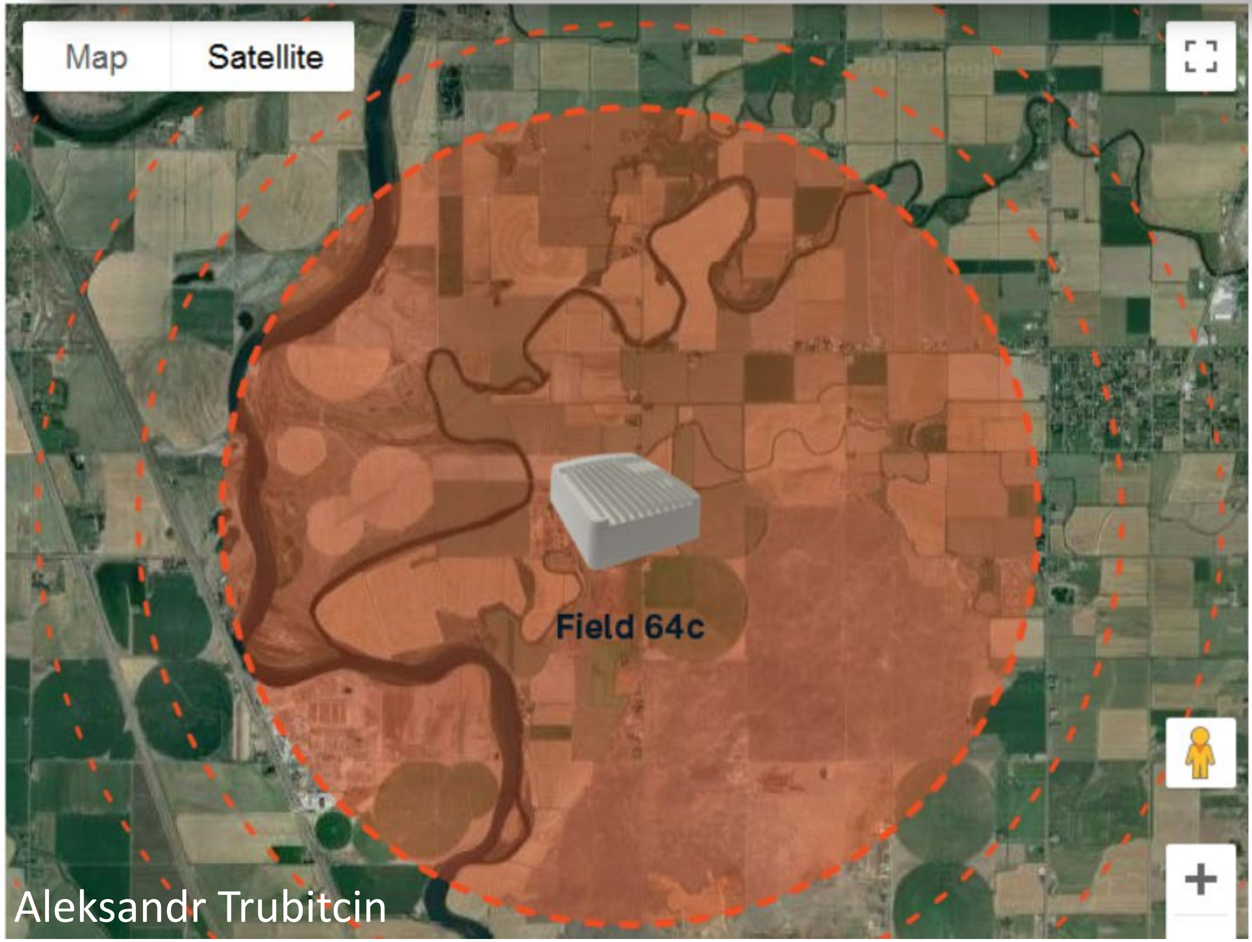
Sensor Placement: Outdoor

Sensor Height: 1 ft

Sensor Tx Power: 14 dBm

Estimated Gateway Range
3.5 miles

CONNECTIVITY costs $0.00137 per cow per day



Field 64c gateway with 64 channels of LoRaWAN connectivity and Ethernet/cellular backhaul manufactured by Tektelic, Canada.

MachineQ prices:

Gateway Field 64c $2800 (CAPEX)

Software License $4979 pa (OPEX)

Cell $119 per gateway pa (OPEX)

Connectivity fee per animal (10,000)

US $0.50 per year (50 cents pa)

0.001 cents

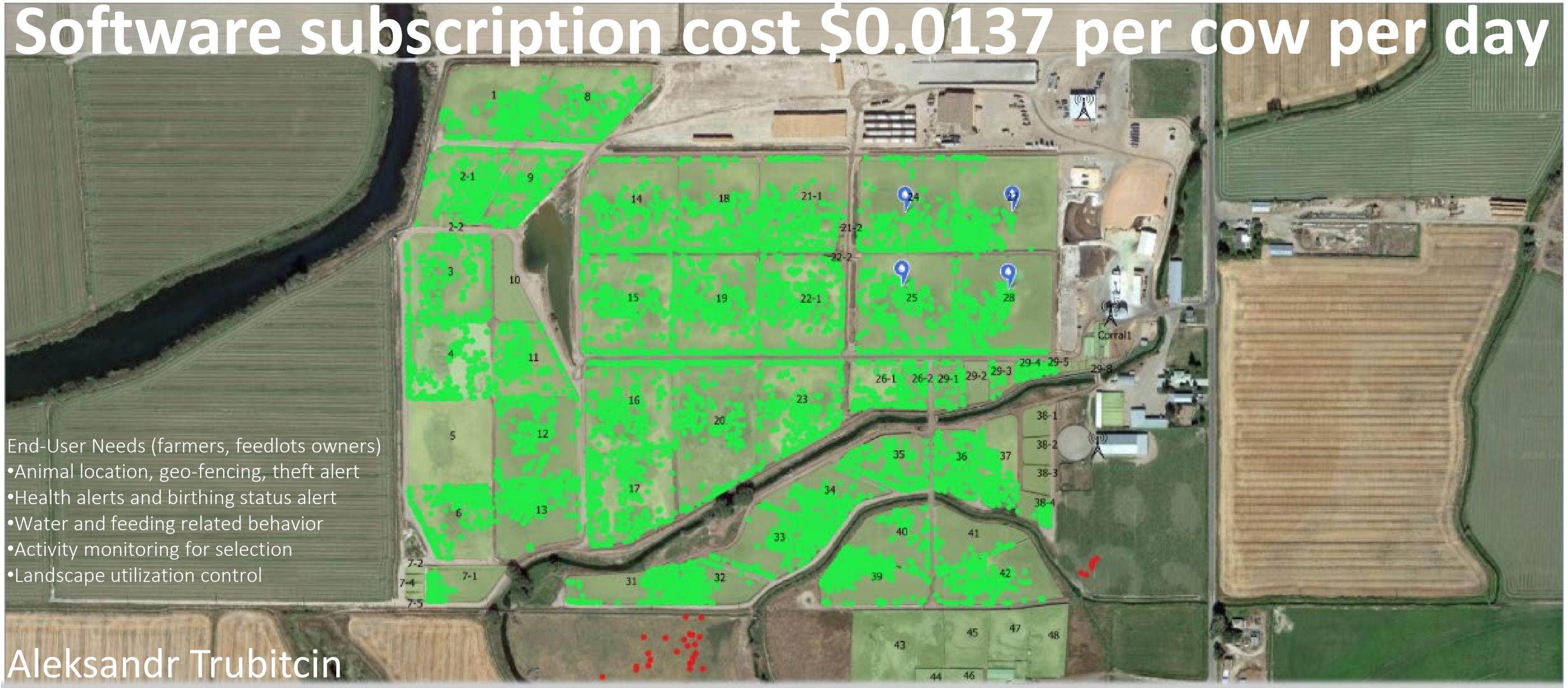
LoRaWAN Cattle Ear Tags: 200 acre feed-lot for 10,000 cows

IoT SYSTEMS: HOW NANO-FEES MAY GENERATE MEGA-MILLIONS

See SIGNALS - https://dspace.mit.edu/handle/1721.1/111021

The user will also need access to the cattle tracking and monitoring web application. For example, the cost of an annual subscription to the Cattle Tags Technologies app will be $5 per animal.

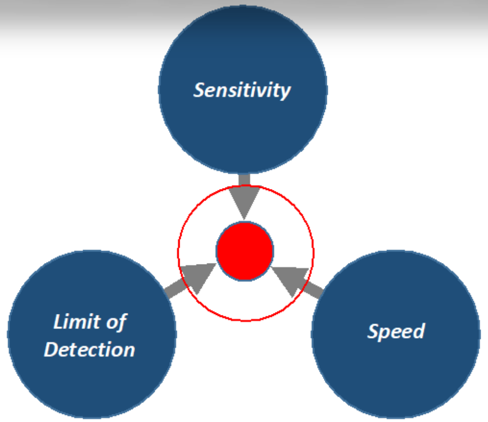
0.12 cents

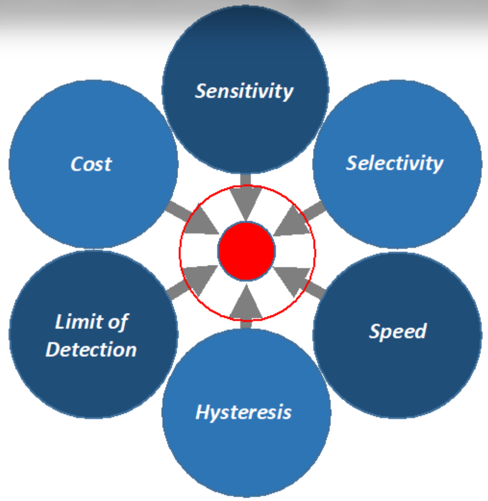


LoRaWAN ear tag from Cattle Tags Technologies starts from $39. Tags have embedded GPS receiver, accelerometer, temperature sensor and replaceable battery. Operator reads RFID-tag with Bluetooth reader (ID sent to ERP system). Installation of activated LoRaWAN ear tag follows.
alex.trubitcin@gmail.com

Digital Transformation (Connectivity+App+Tag) Cost $0.12 per cow per day

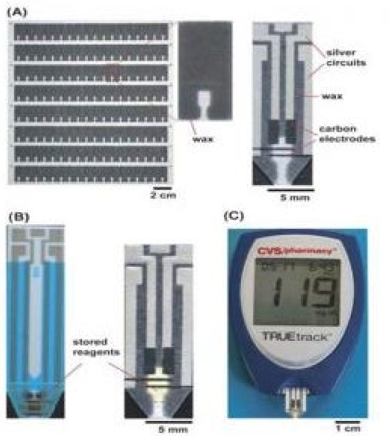
Has sensor engineering evolved with digital transformation

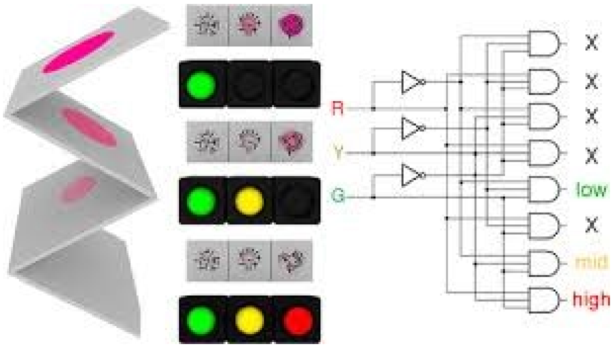




1990's

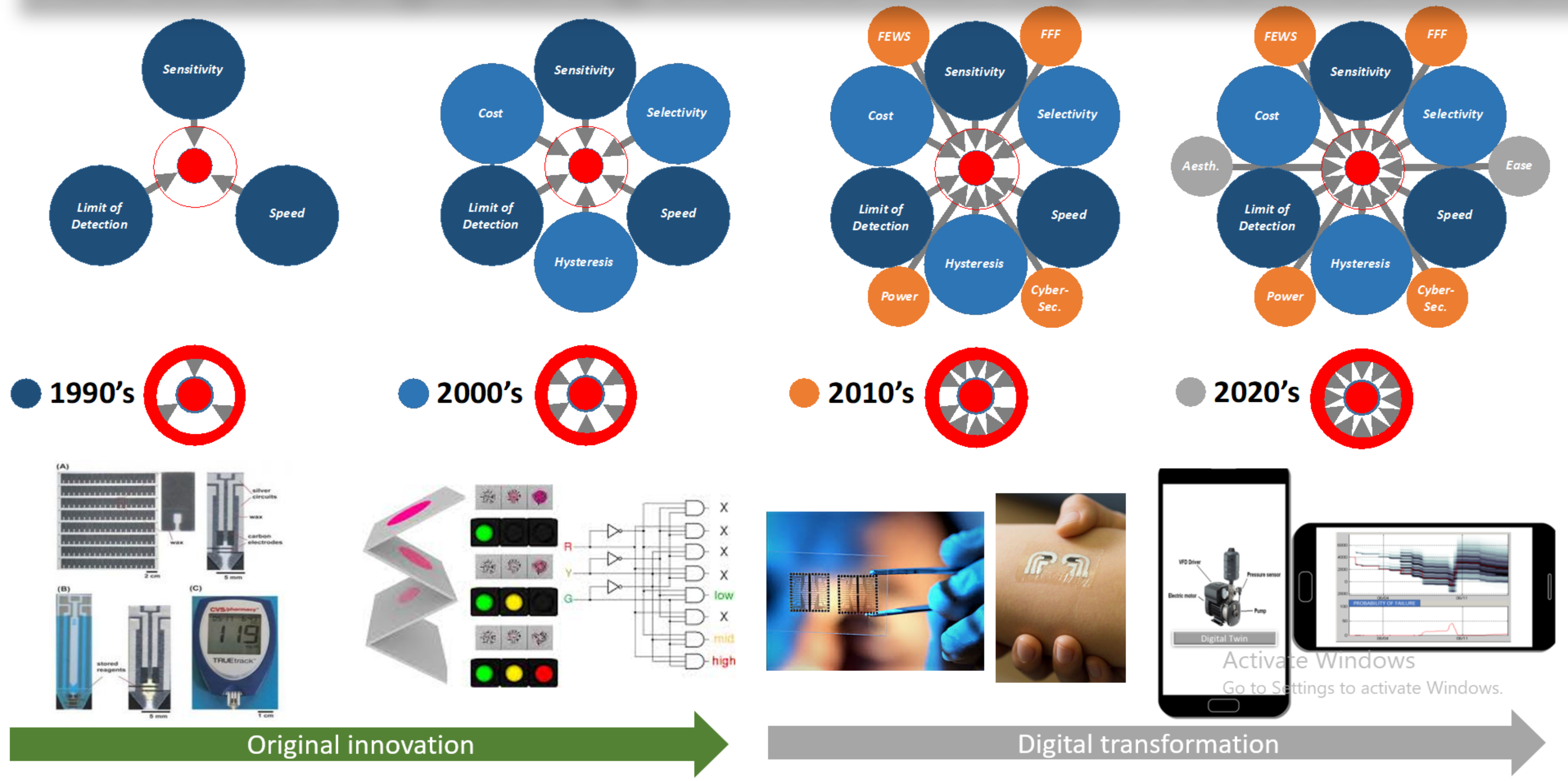
2000's





Original innovation

Has sensor engineering evolved with digital transformation



Sensor engineering has evolved with digital transformation

SENSE, ANALYZE, RESPONSE SYSTEMS – SARS

Convergence of

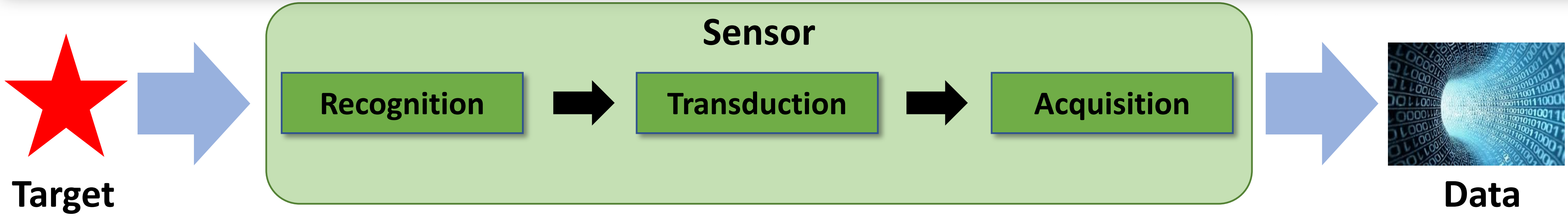
DATA from SENSORS

Role of Sensors and Sensor Data in Decision as a Service

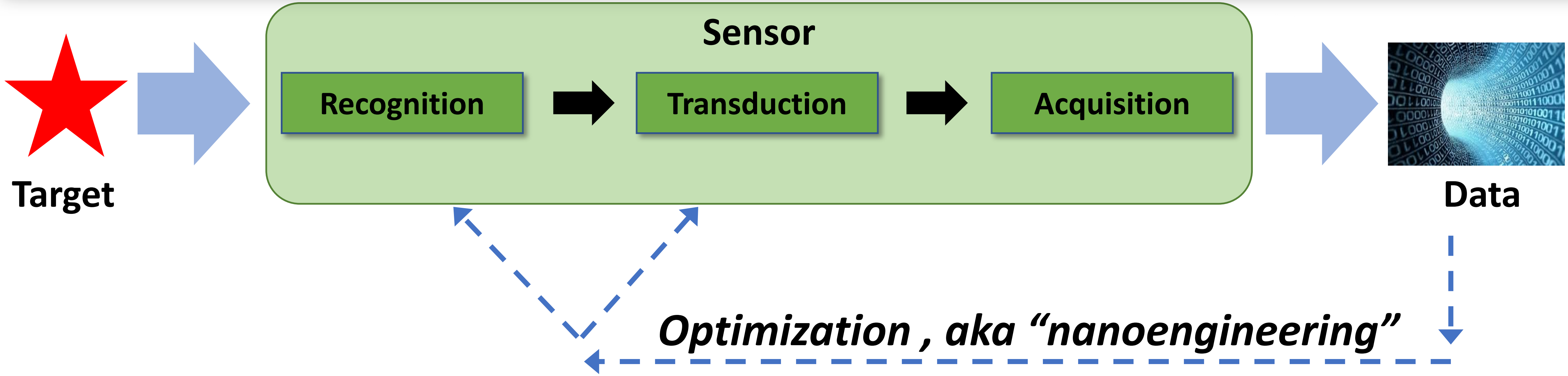
Target

Sensor

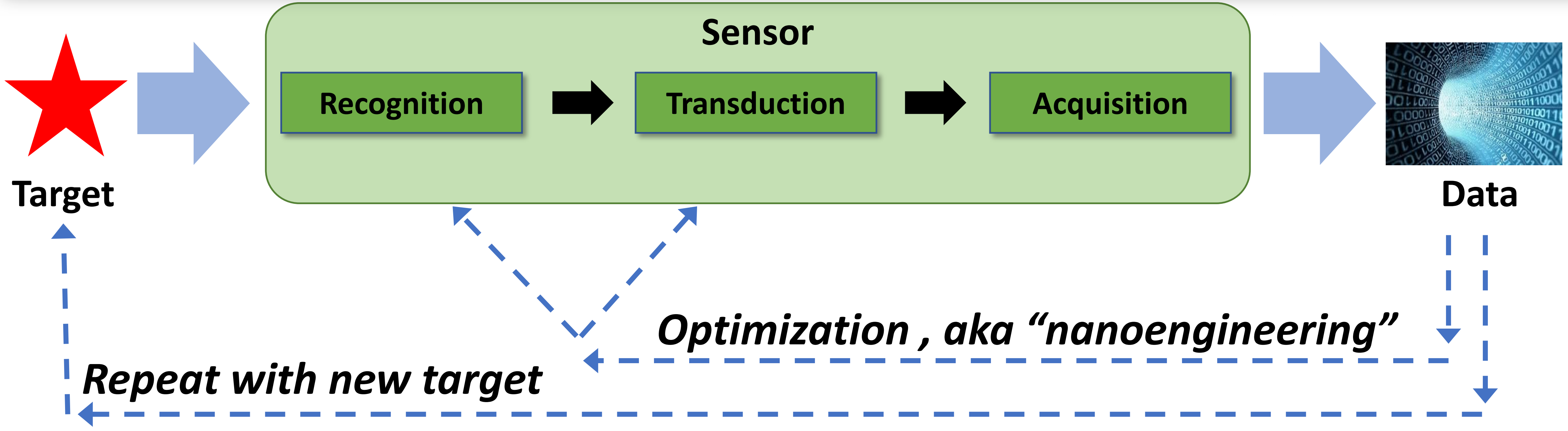
Role of Sensors and Sensor Data in Decision as a Service



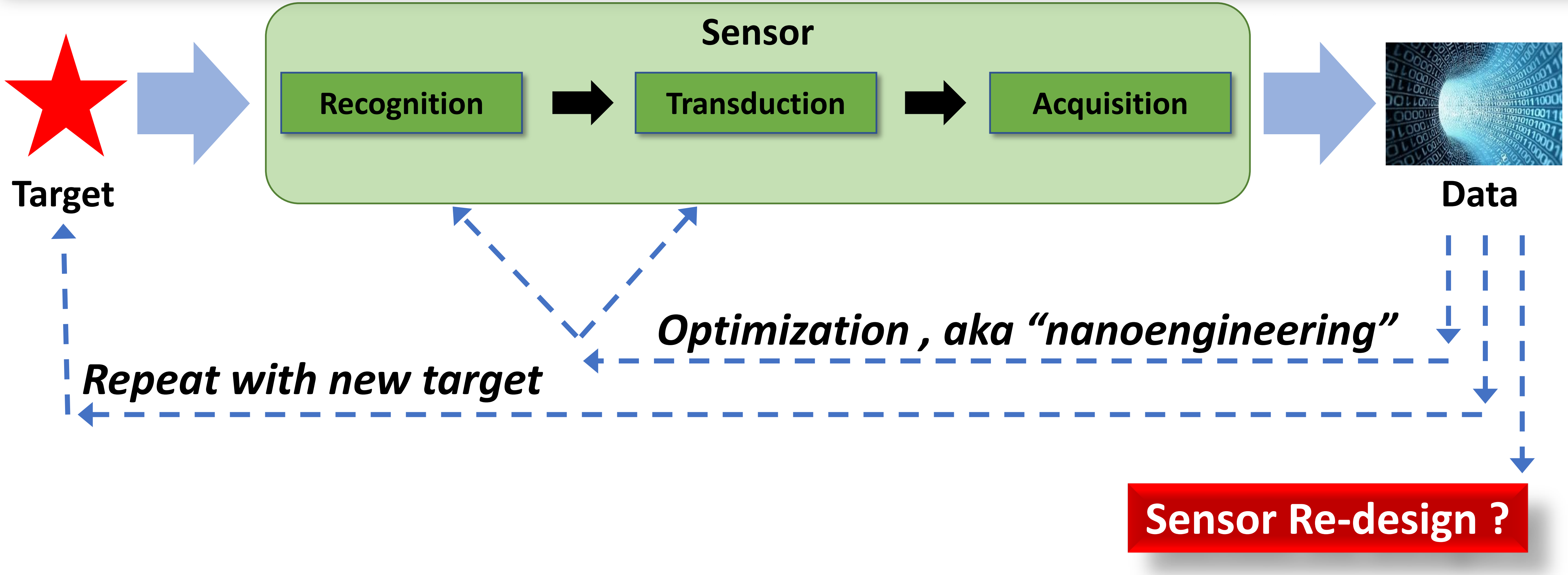
Role of Sensors and Sensor Data in Decision as a Service



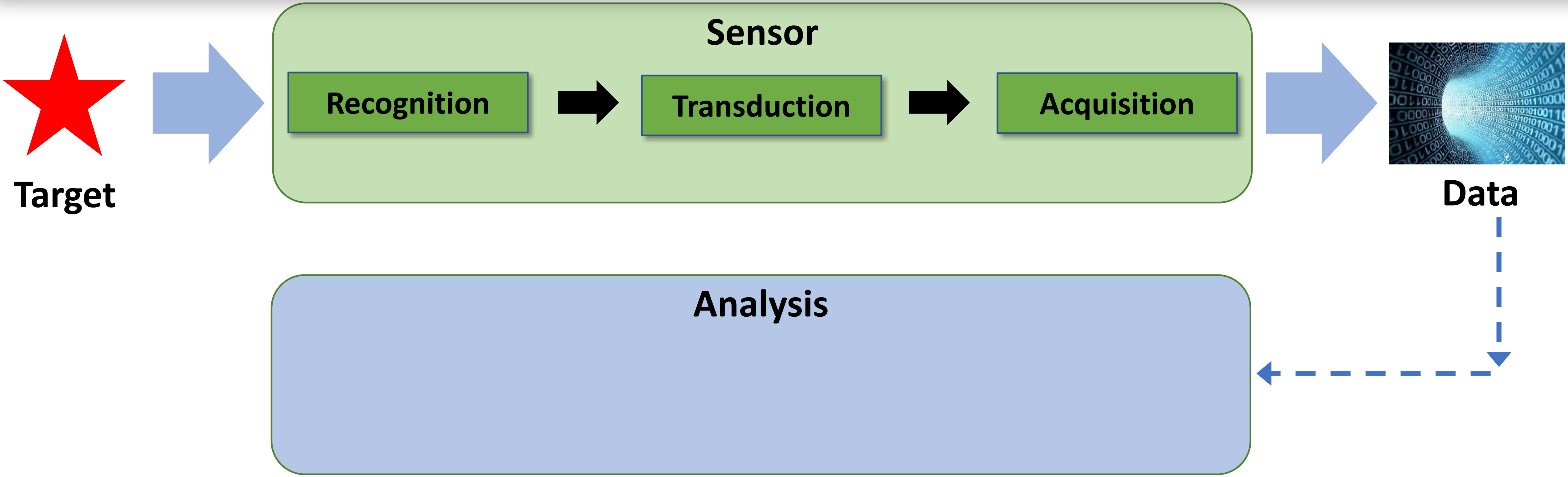
Role of Sensors and Sensor Data in Decision as a Service



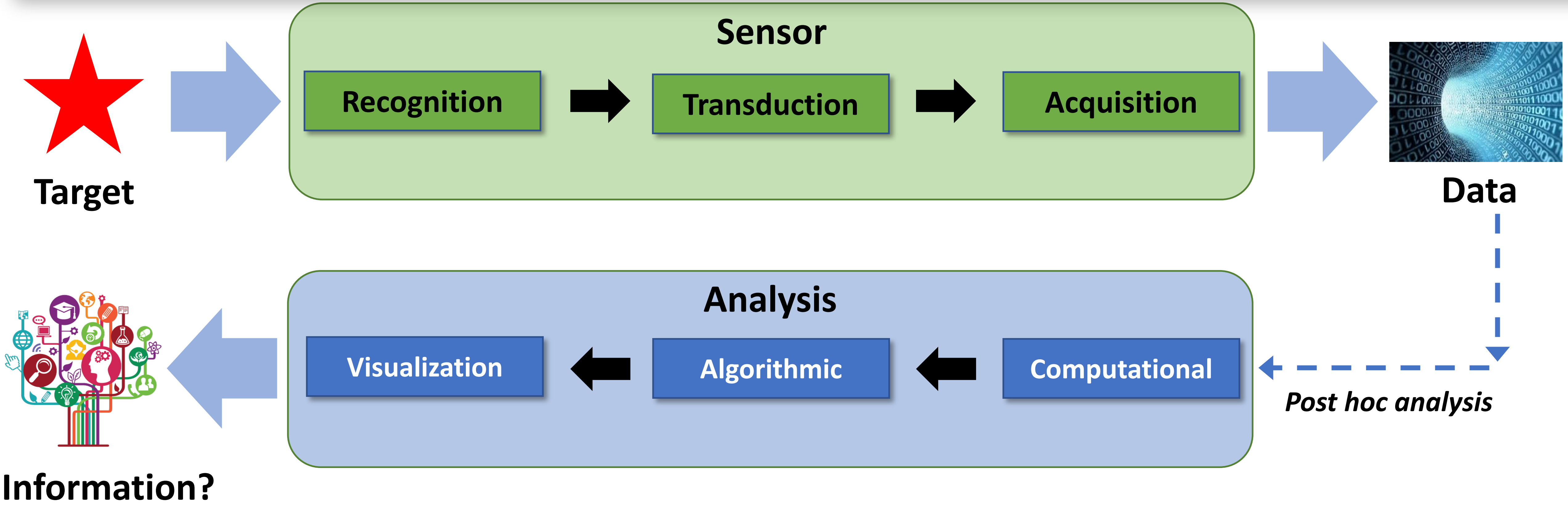
Role of Sensors and Sensor Data in Decision as a Service



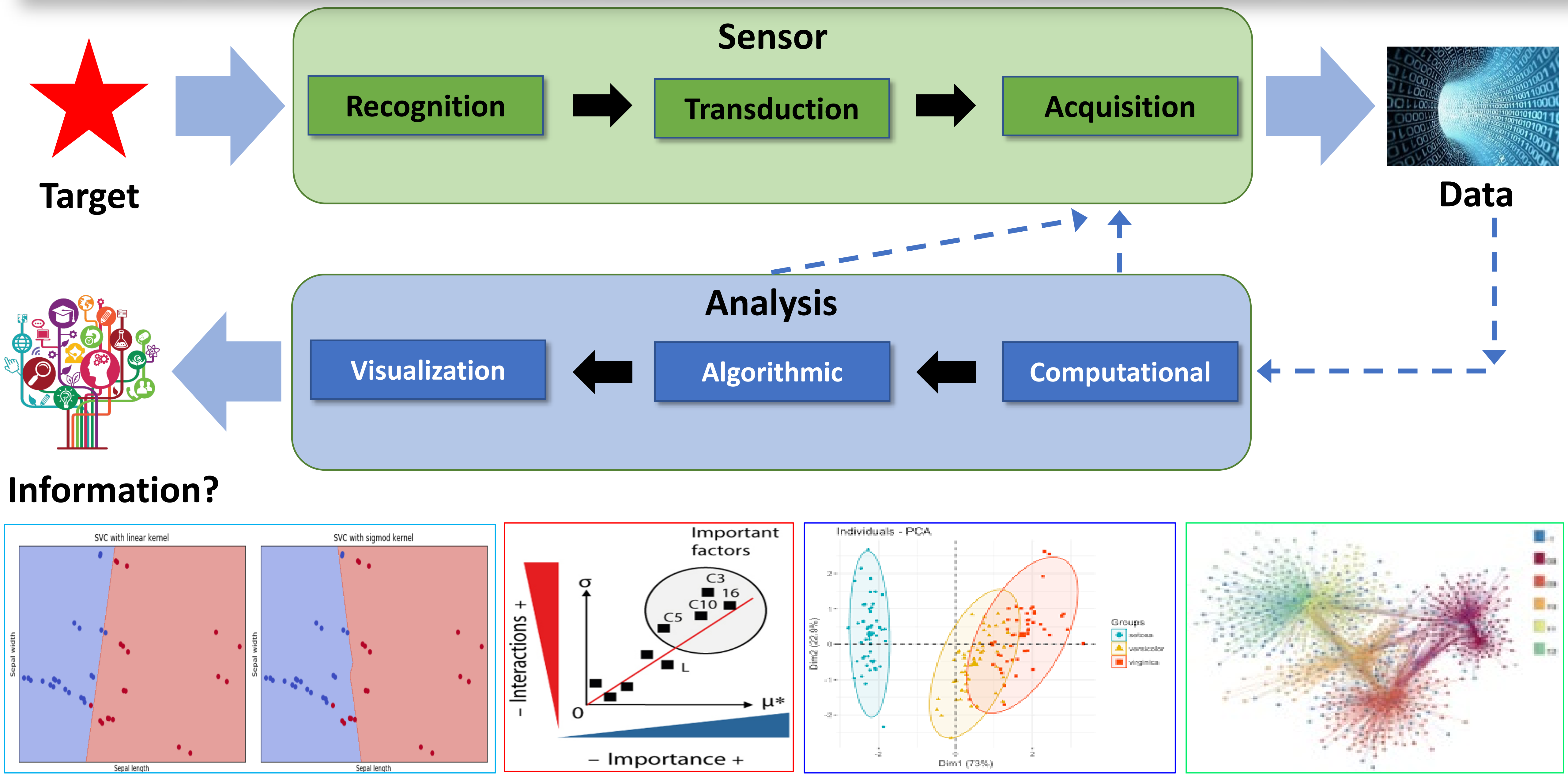
Role of Sensors and Sensor Data in Decision as a Service



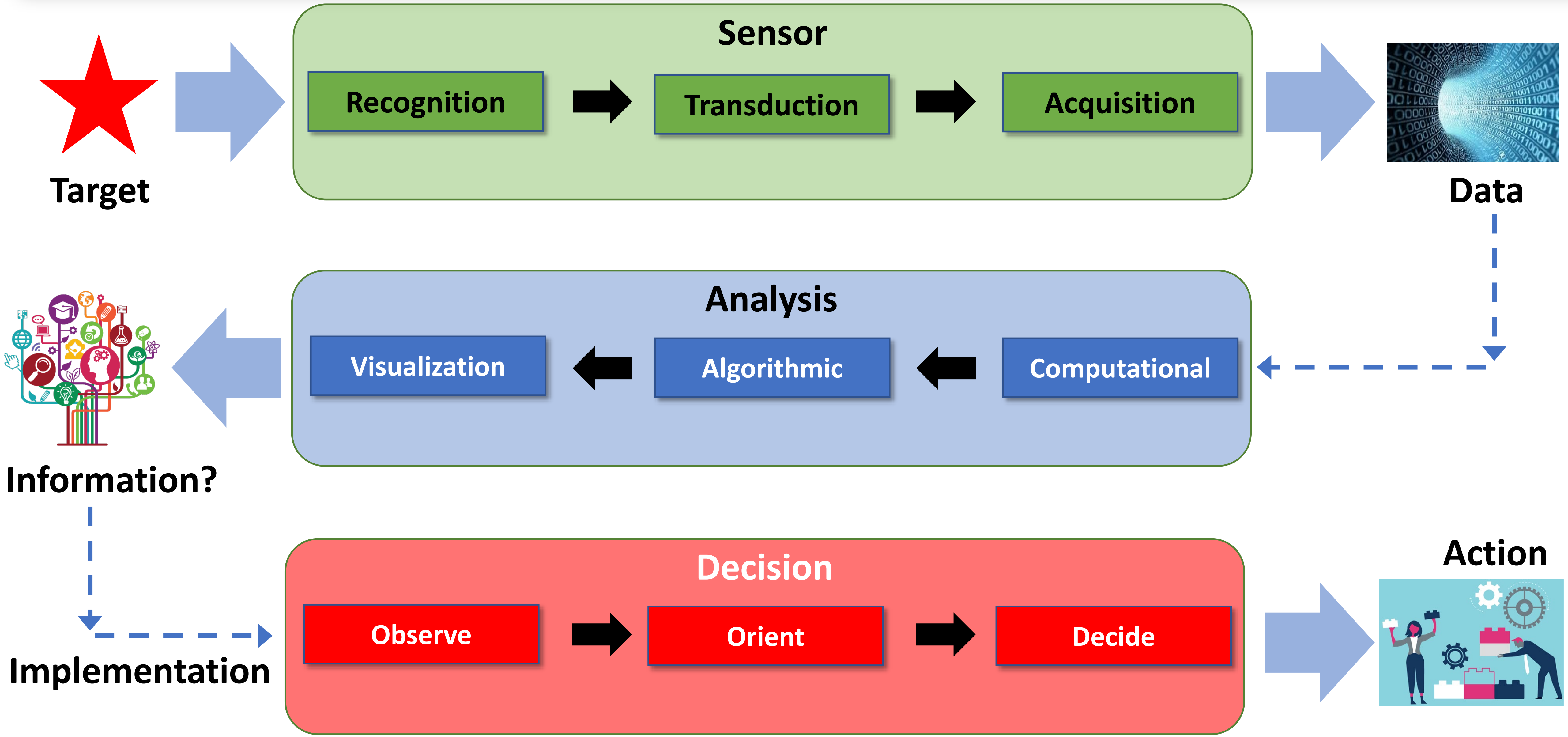
Role of Sensors and Sensor Data in Decision as a Service



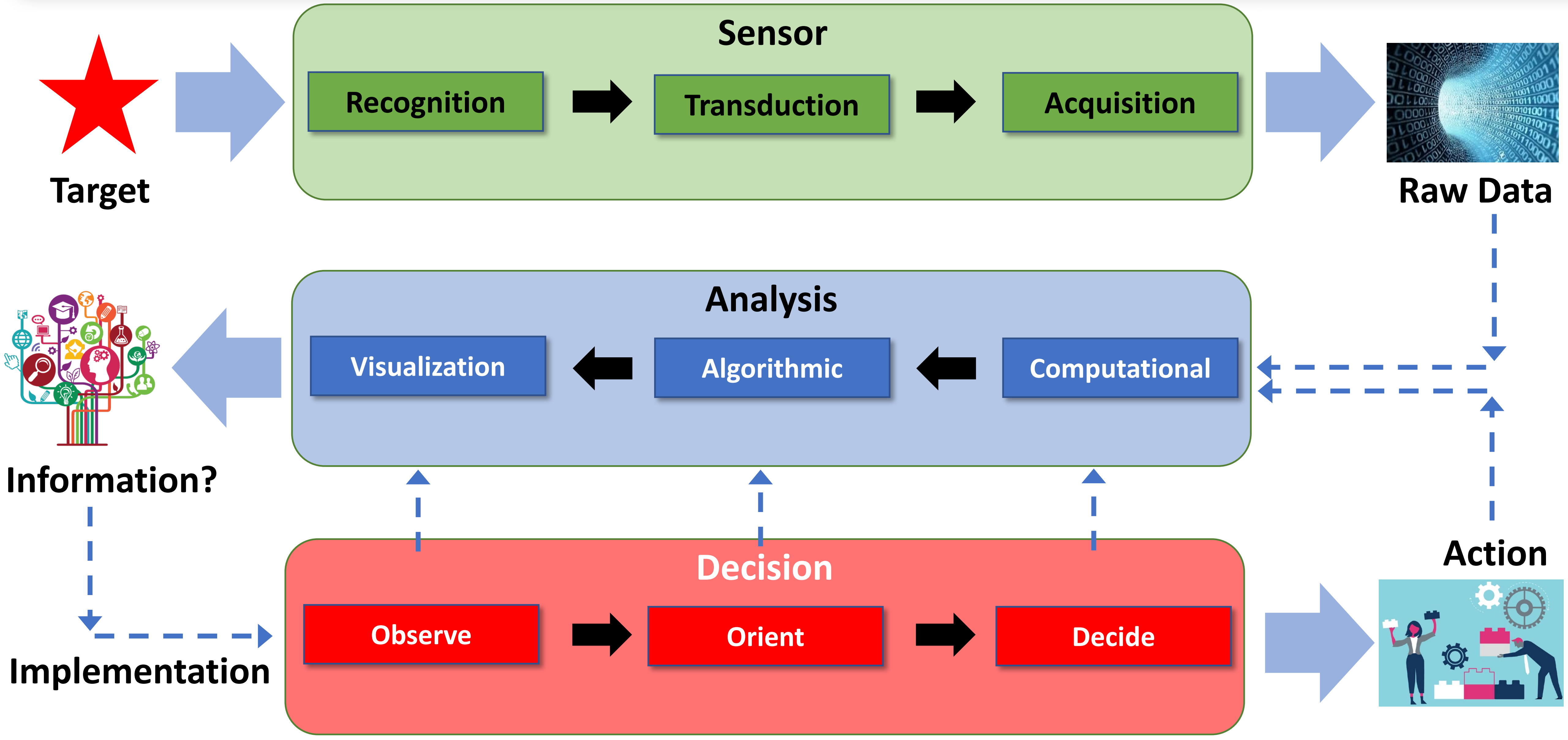
Role of Sensors and Sensor Data in Decision as a Service



Role of Sensors and Sensor Data in Decision as a Service



Data-Informed Decision as a Service

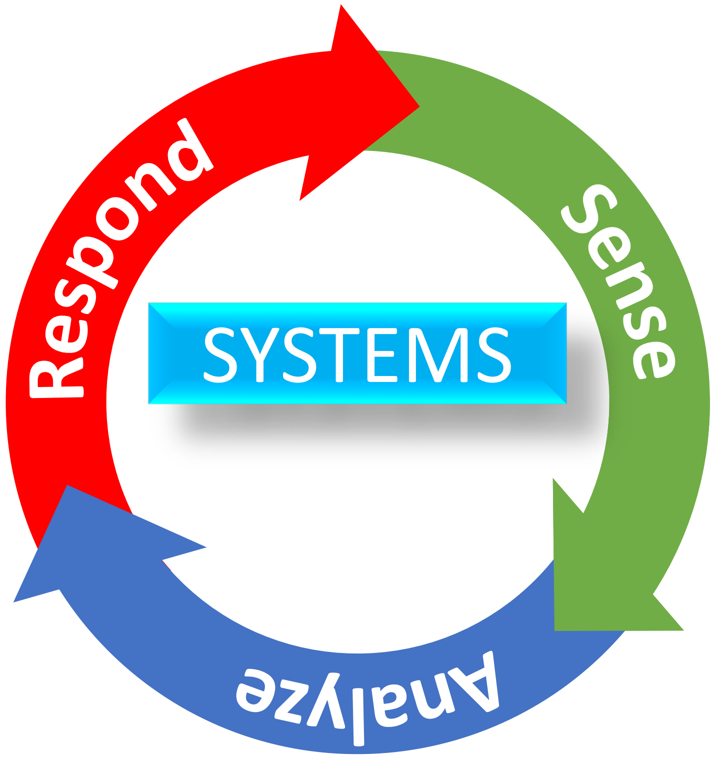


Role of Sensors and Sensor Data in Decision as a Service

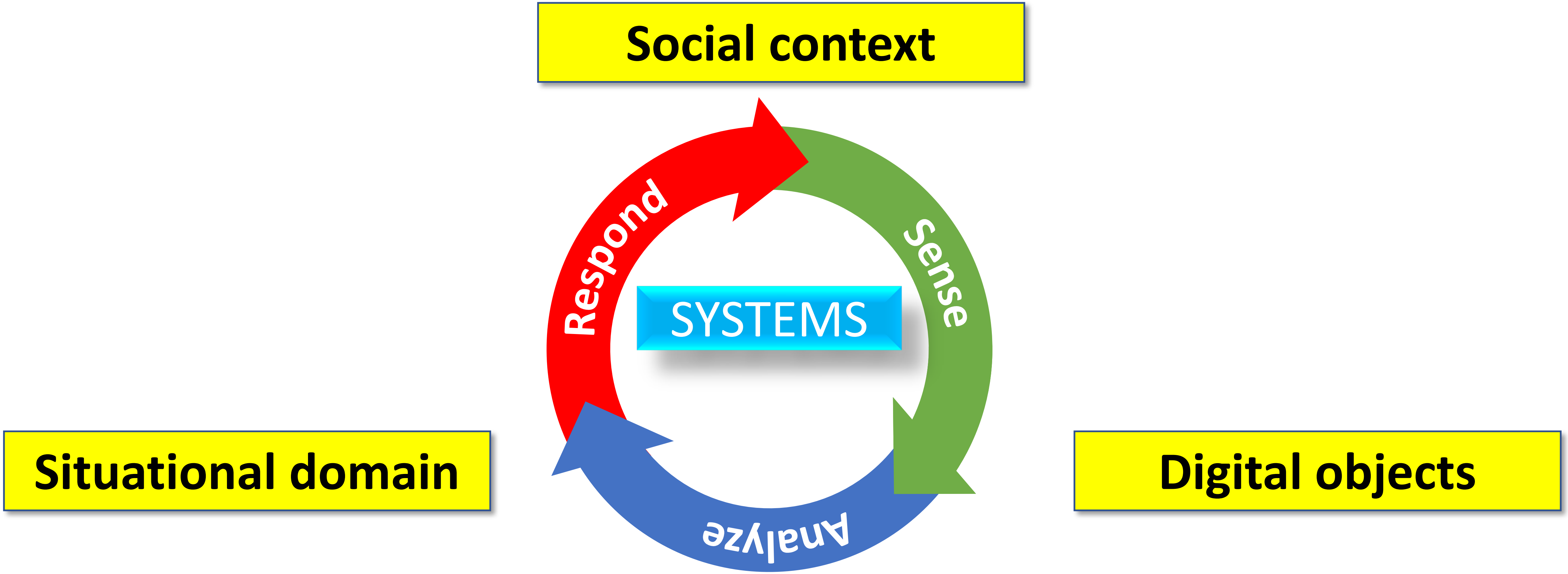
DIDA'S

Data-Informed Decision as a Service

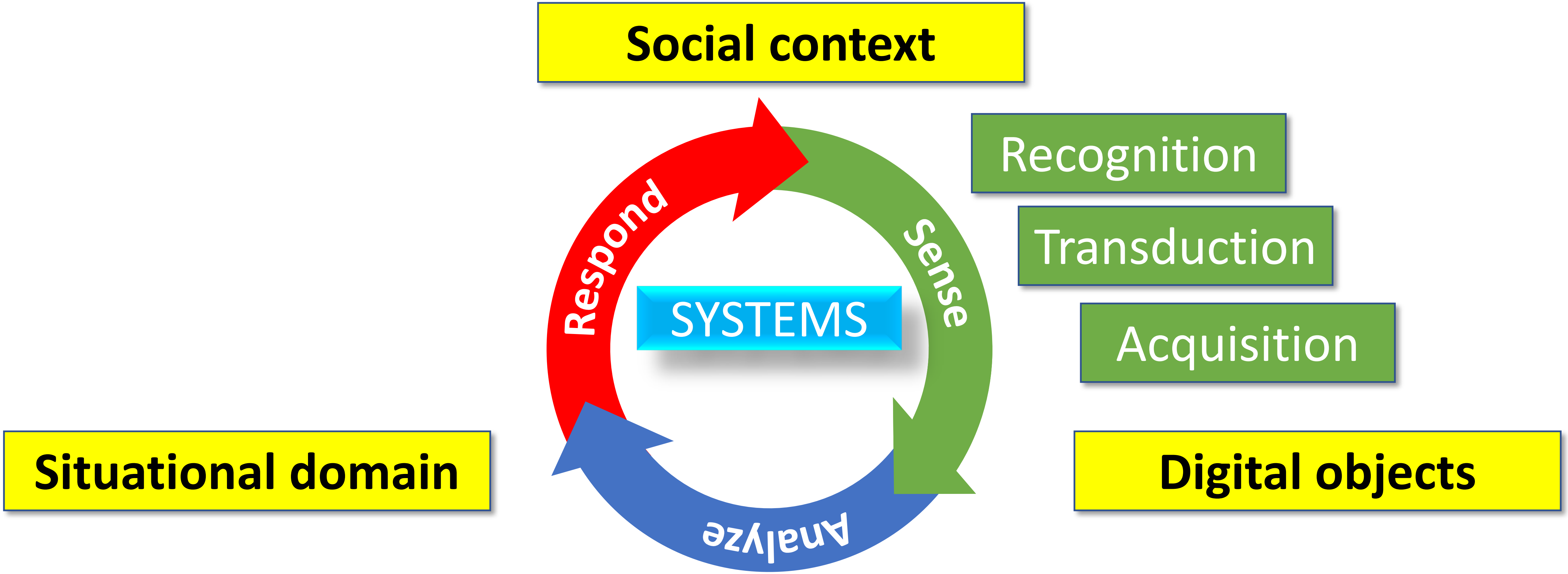
DIDA'S includes Sense, Analyze, Response, Systems (SARS)



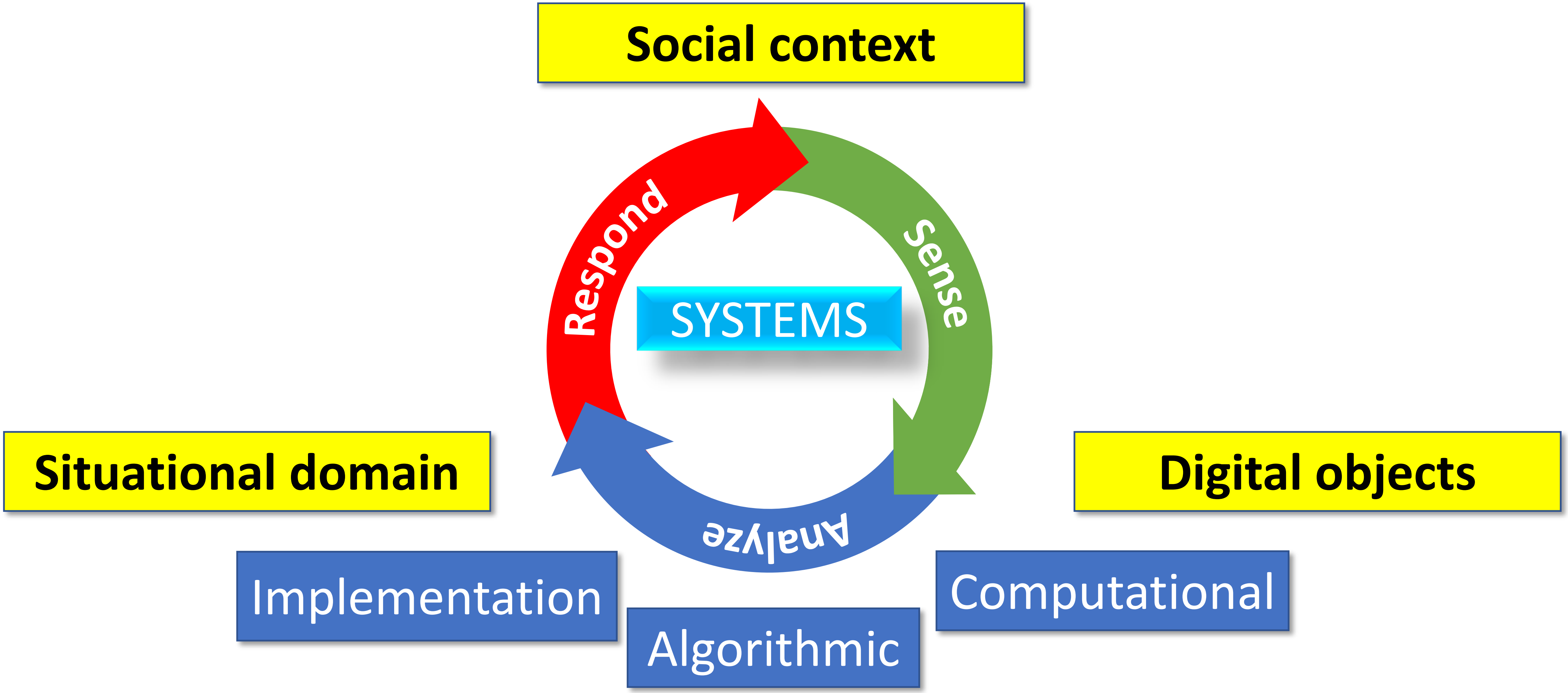
DIDA'S includes SARS context, objects and domains



Granularity of the Data-Informed Decision as a Service



Granularity of the Data-Informed Decision as a Service



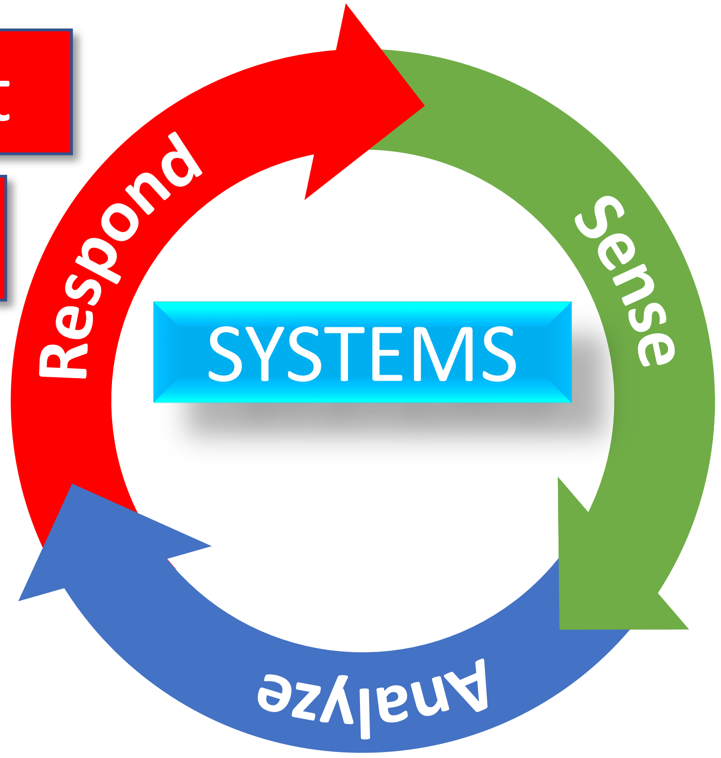
Granularity of the Data-Informed Decision as a Service

Social context

Decide-Act

Orient

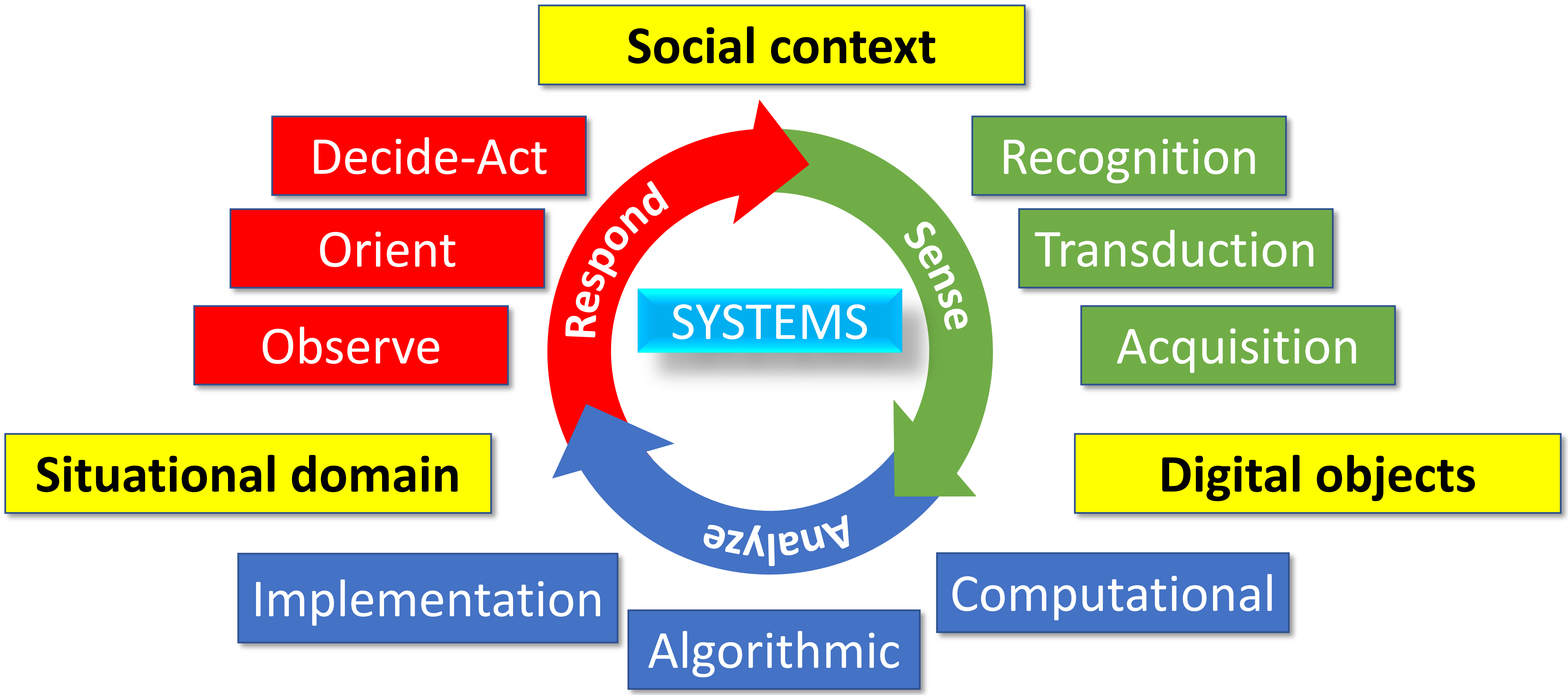
Observe



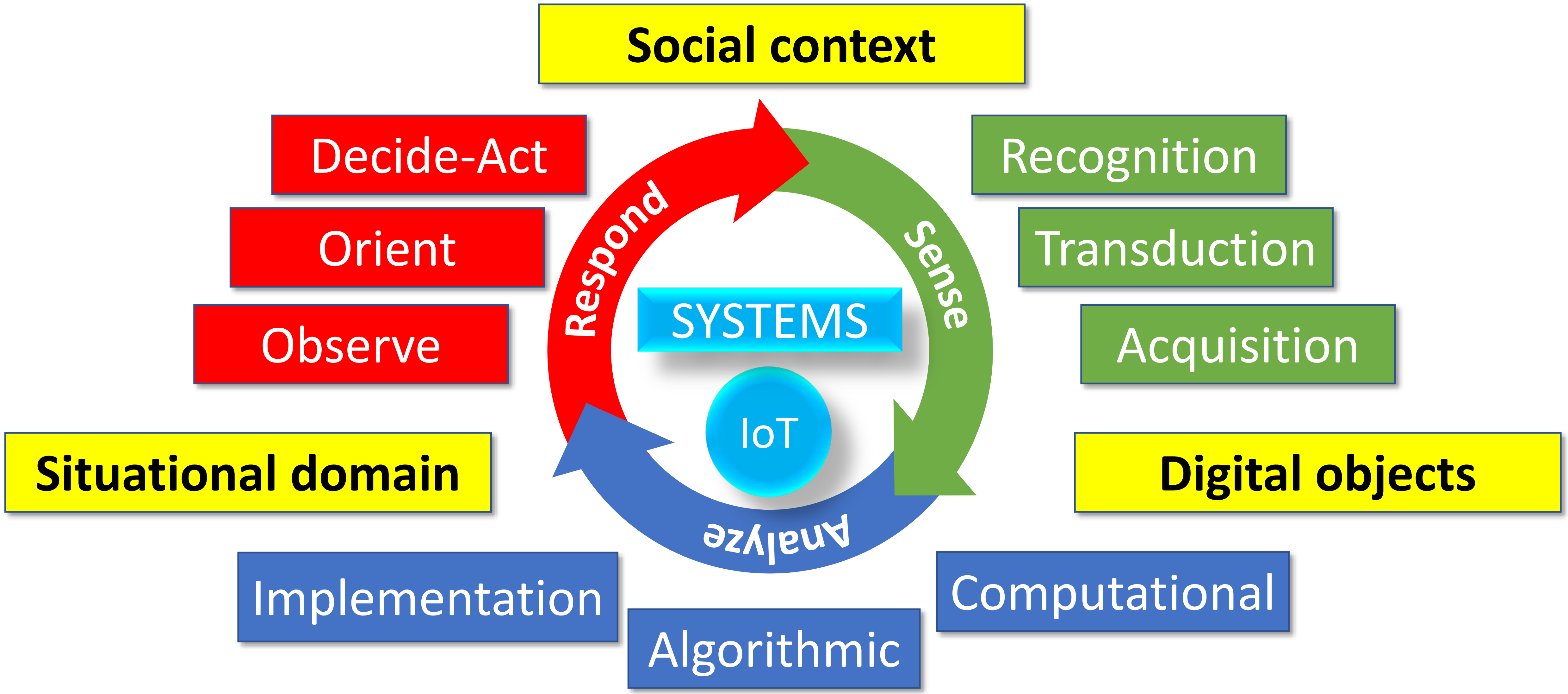
Situational domain

Digital objects

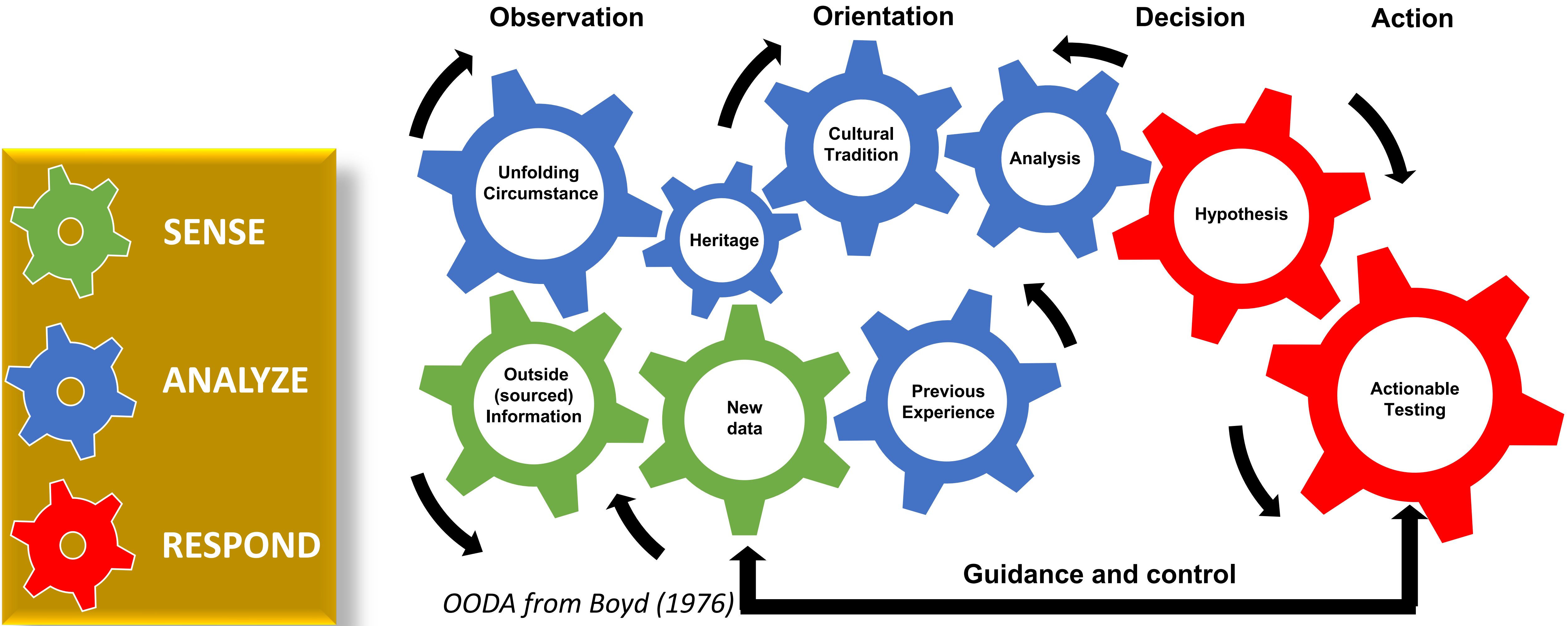
DIDA'S : Data-Informed Decision as a Service with SARS



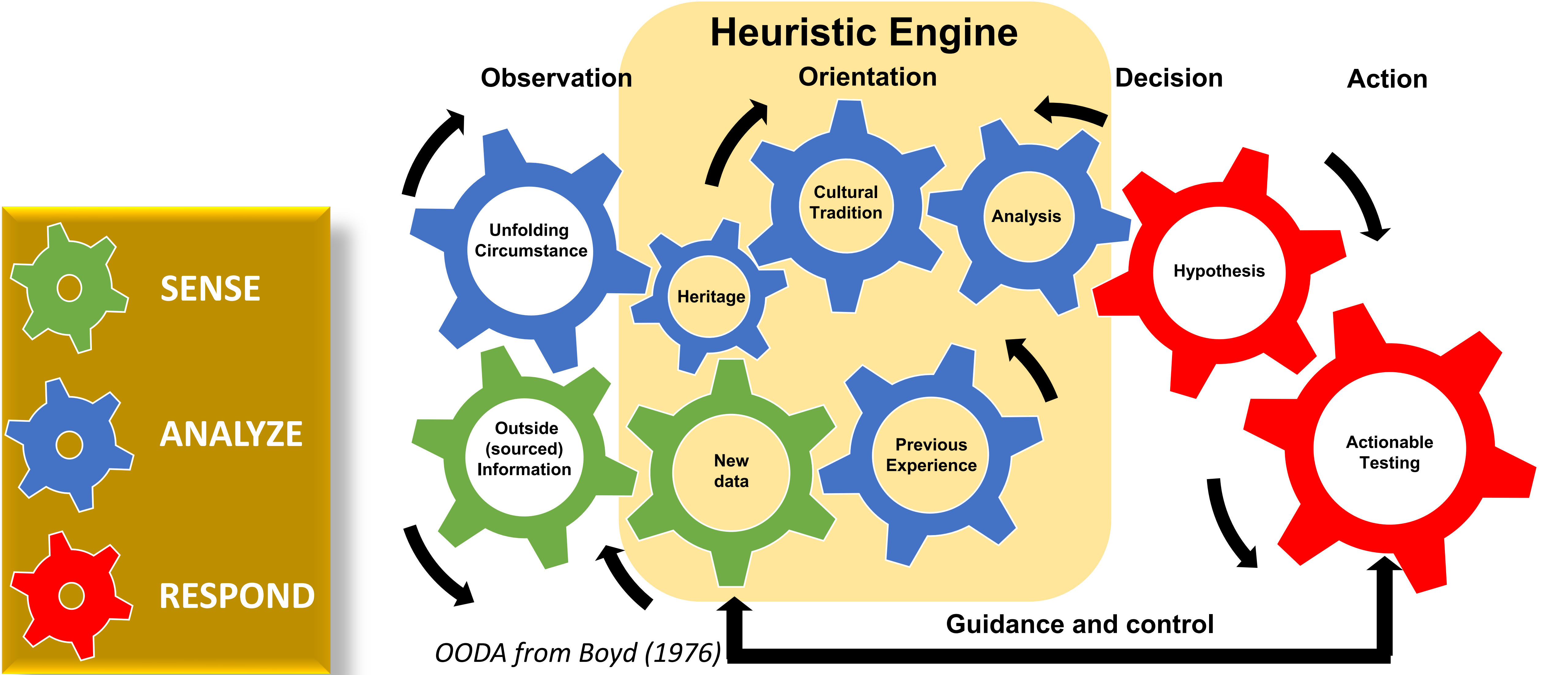
DIDA'S : Data-Informed Decision as a Service



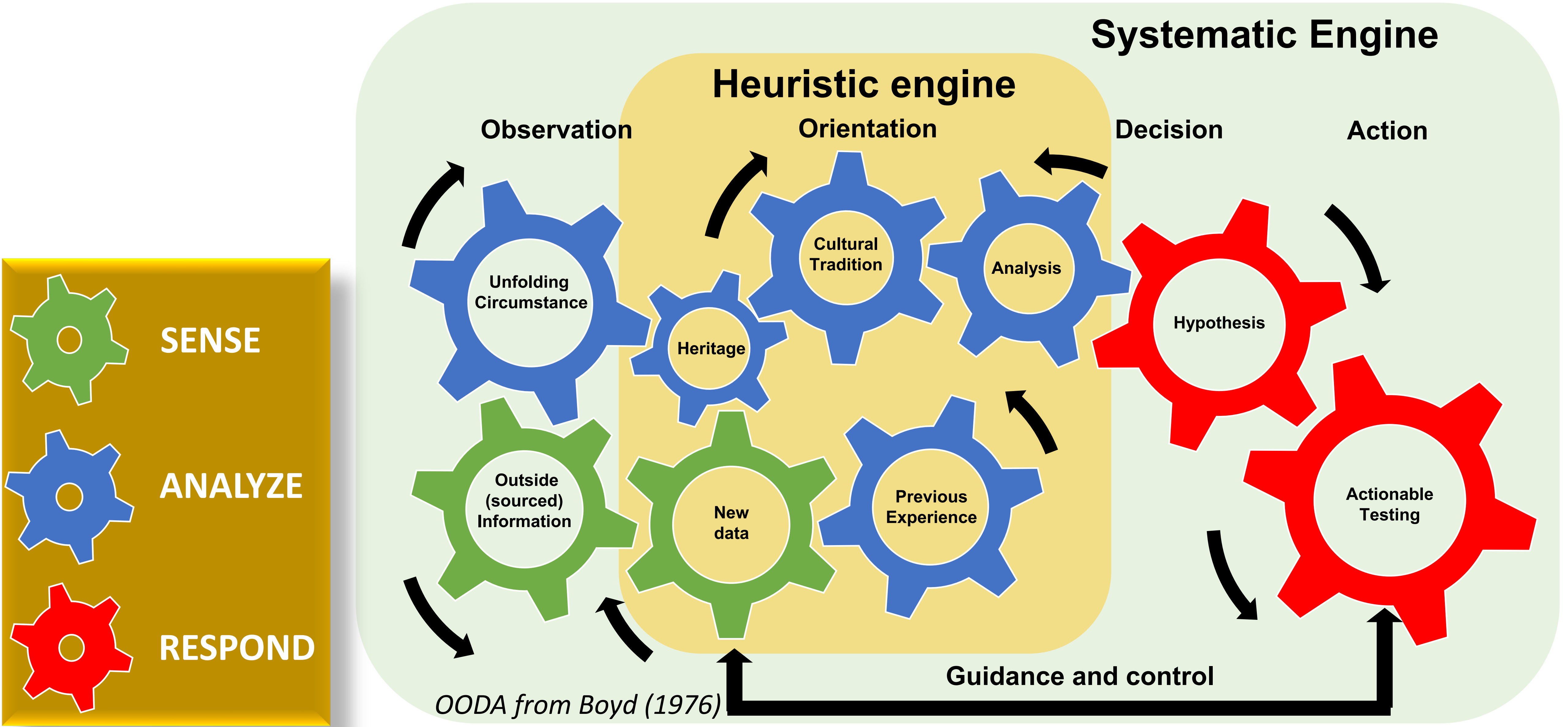
DIDA'S ENGINES : Data-Informed Decision Engines



DIDA'S ENGINES : Data-Informed Decision Engines



DIDA'S ENGINES : Data-Informed Decision Engines



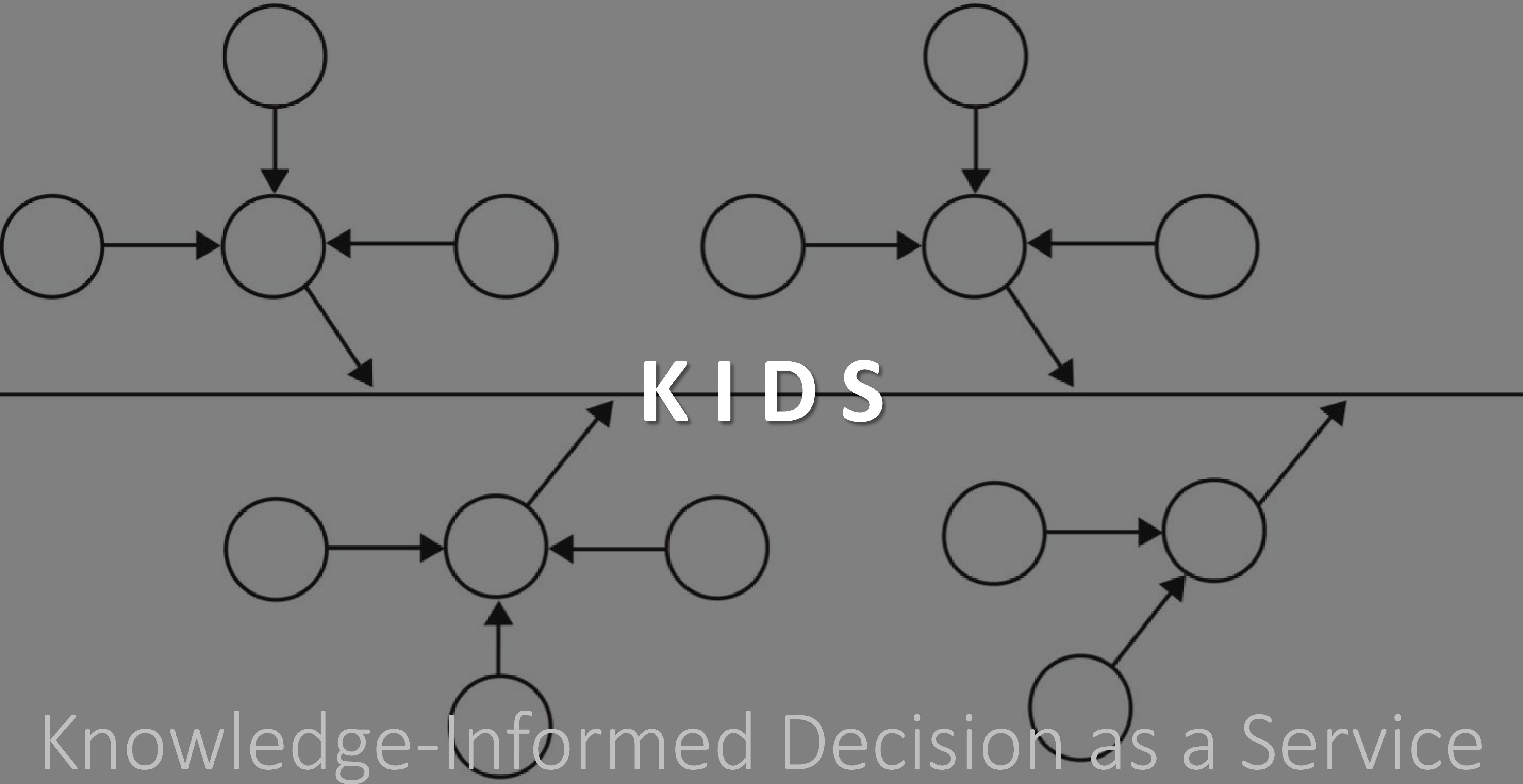
The Value of DIDA'S (Data-Informed Decision as a Service)

Is there consumer demand for this vegetable?

The Value of DIDA’S (Data-Informed Decision as a Service)

To realize the actual value of DIDA’S, the tool must be useful to end-users, if they can use the tool to ask questions and receive actionable information or if it can support the decision making process

The journey to DIDA’S must include and/or create and/or connect a multitude of domains to source data and synthesize relevant information. DIDA’S may lead to knowledge-informed decision as a service.



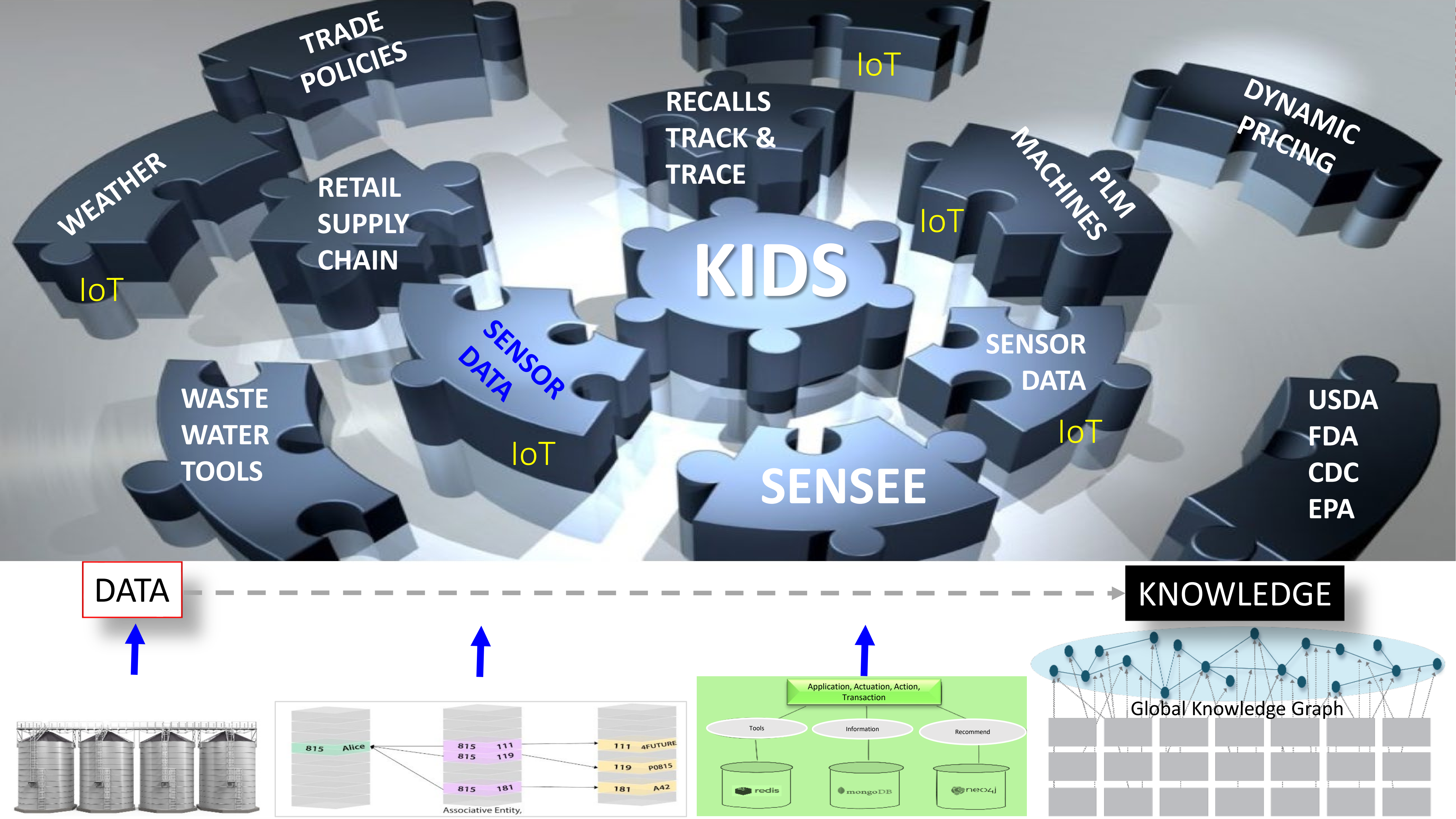
KIDS

Knowledge-Informed Decision as a Service

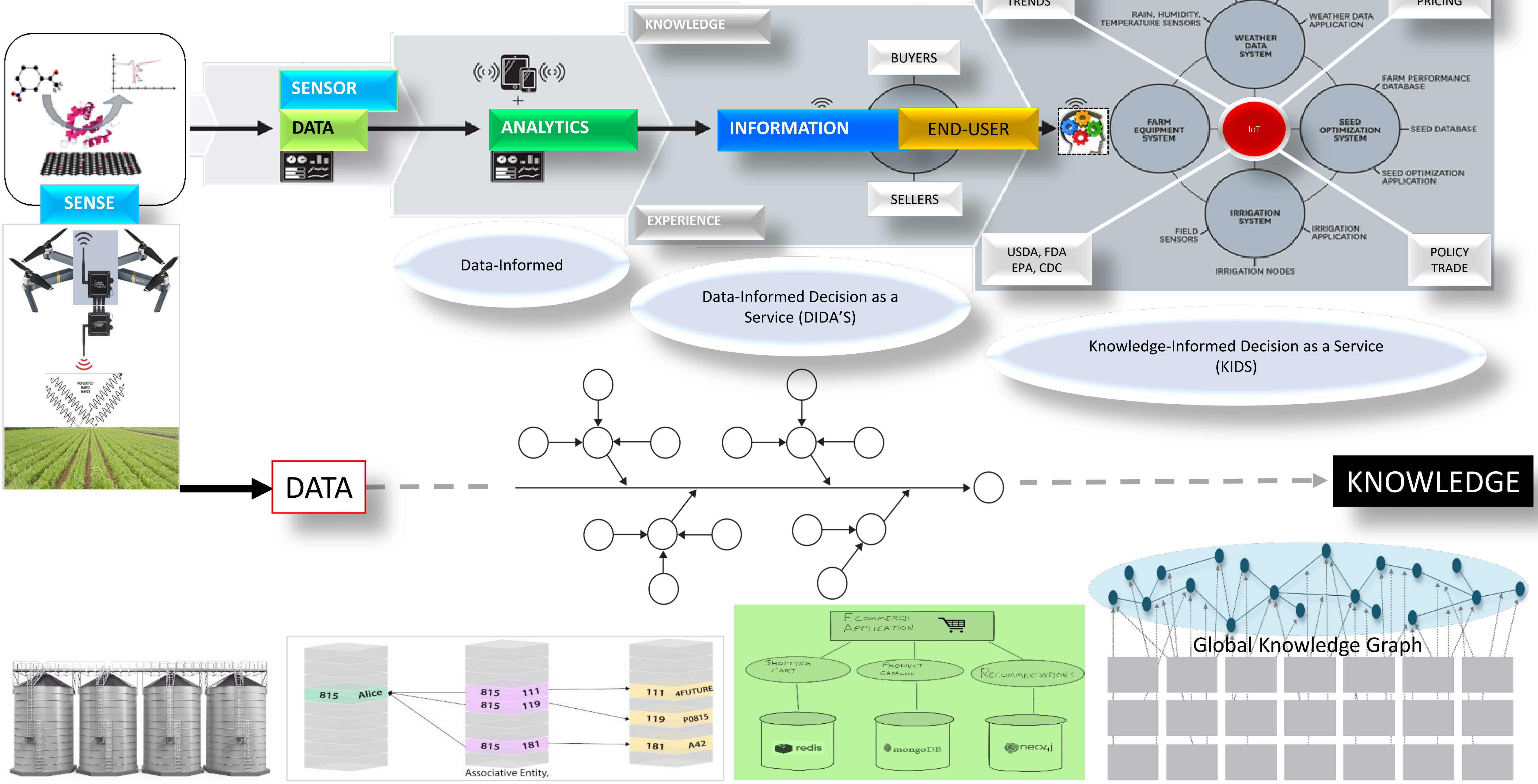
K I D S

Knowledge-Informed Decision as a Service

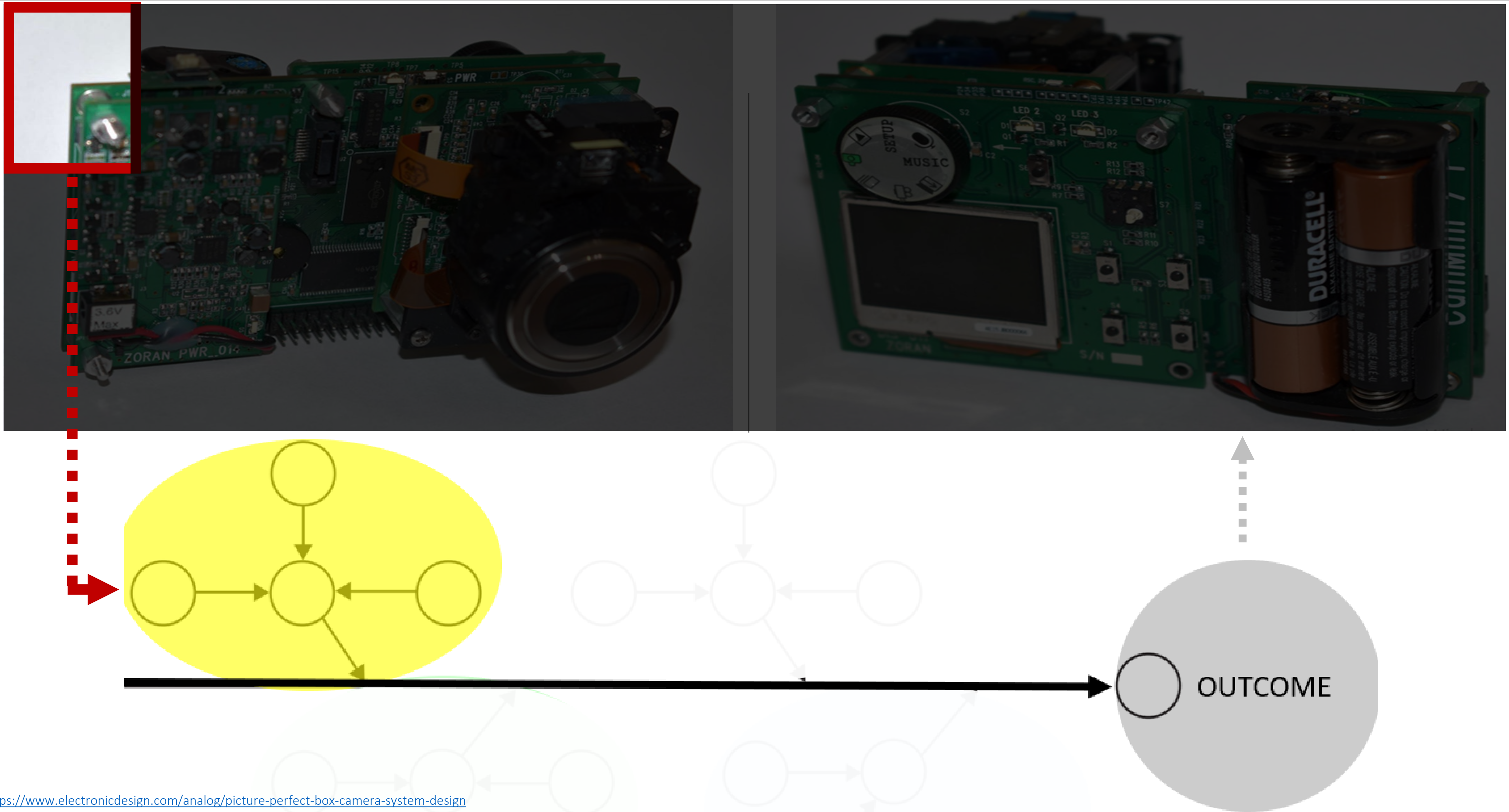
KIDS is an open plan platform concept. Platforms are comprised of multiple applications and integrated solutions with embedded tools and databases that function as complete, seamless environments. Product innovation platforms are intended to support groups of users collaborating across various levels, domains, business units, and the ecosystem. These capabilities are increasingly needed throughout the entire extended enterprise in almost every vertical, agnostic of the type of application or function or users, including farmers, meat packers, produce growers, retail stores, customers, suppliers, and business partners. Developing open platform tools and technologies are not limited to any one domain because these modular tools can be applied, used and re-configured for re-use, almost anywhere, for example: error correction, search engine algorithms, NLU/NLP (natural language processing), automated feature engineering, drag and drop functions, analytics, workflows, and services, such as KIDS, where “open” means ‘plug & play’ user friendly human-computer interactions and interoperability between system of systems.


TRADE POLICIES
IoT
RECALLS TRACK & TRACE
DYNAMIC PRICING
WEATHER
RETAIL SUPPLY CHAIN
PLM MACHINES
IoT
KIDS
IoT
SENSOR DATA
SENSOR DATA
WASTE WATER TOOLS
IoT
IoT
USDA
FDA
CDC
EPA
SENSEE
DATA
KNOWLEDGE
815 Alice
815 111
815 119
815 181
111 4FUTURE
119 P0815
181 A42
Associative Entity,
Application, Actuation, Action, Transaction
Tools
Information
Recommend
redis
mongoDB
neo4j
Global Knowledge Graph

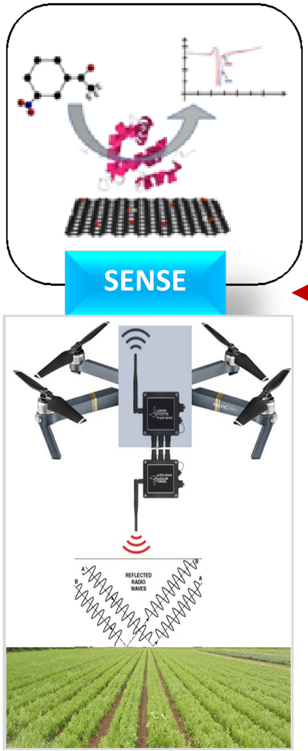
At the end, it is really all about KIDS



We are not even close

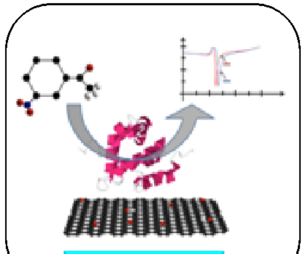


https://www.electronicdesign.com/analog/picture-perfect-box-camera-system-design

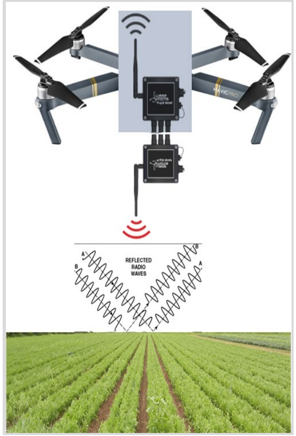


We are just starting here

SENSEE

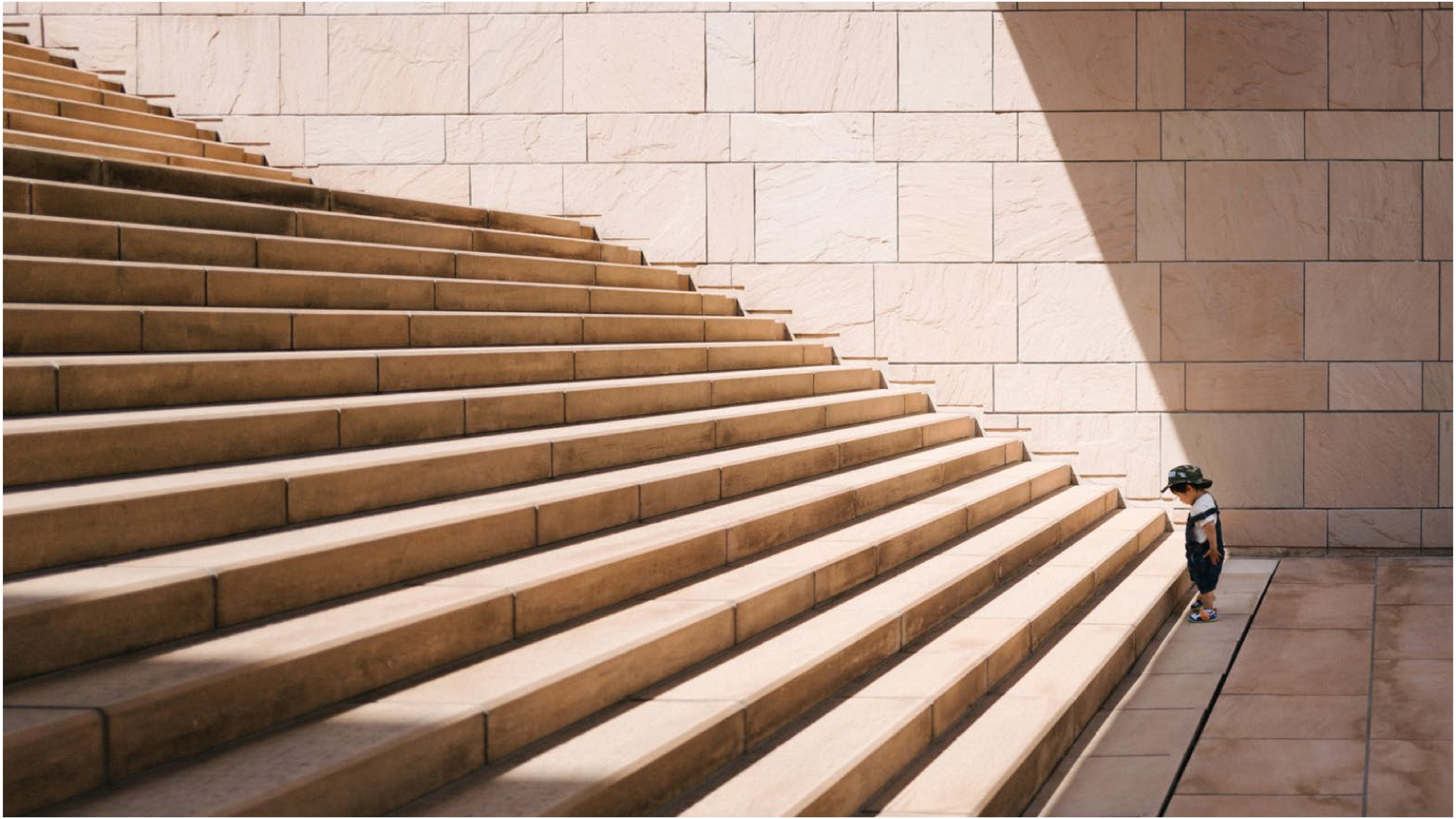


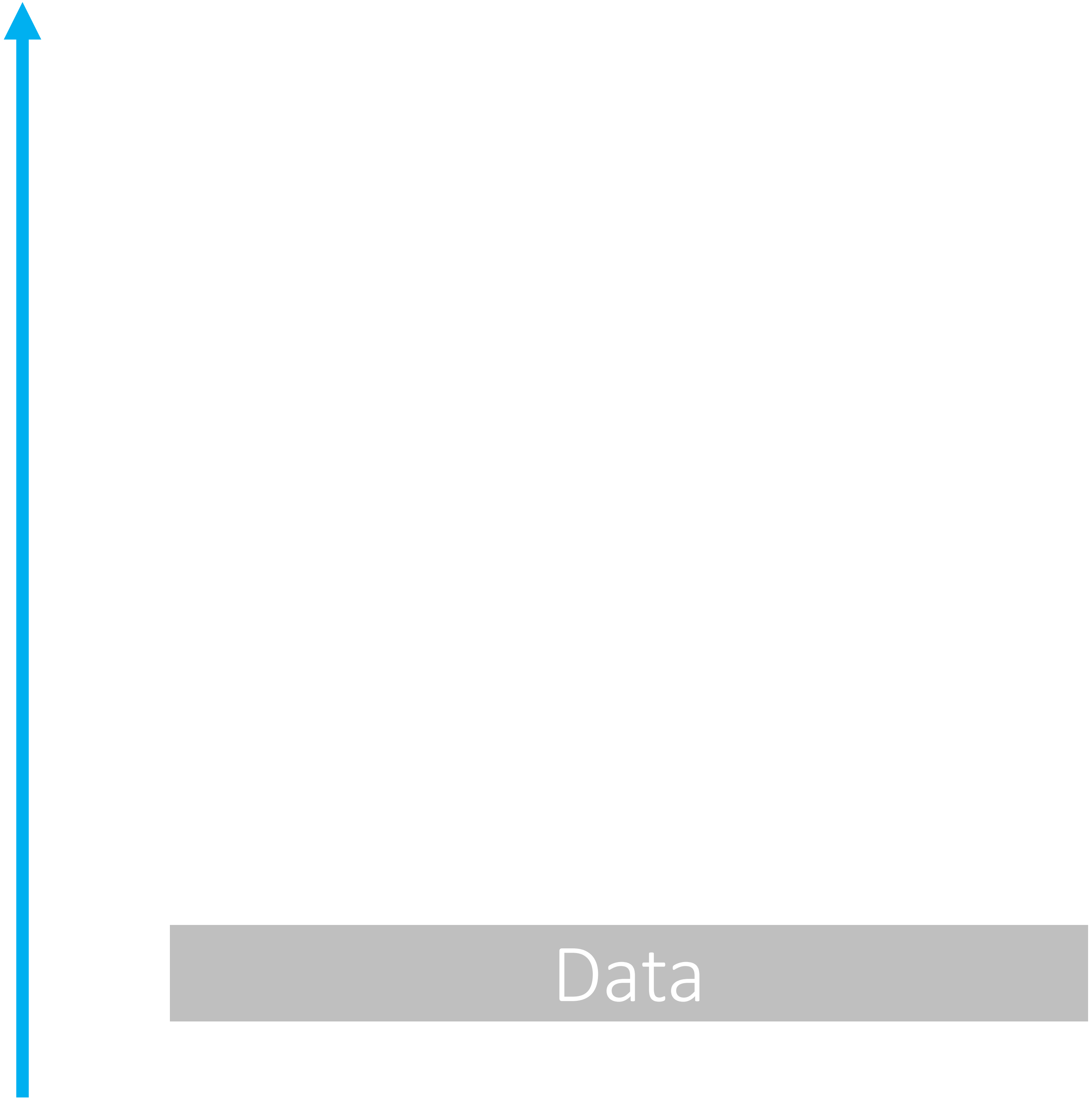
SENSE



We are just starting here

SENsor **SE**arch **E**ngine

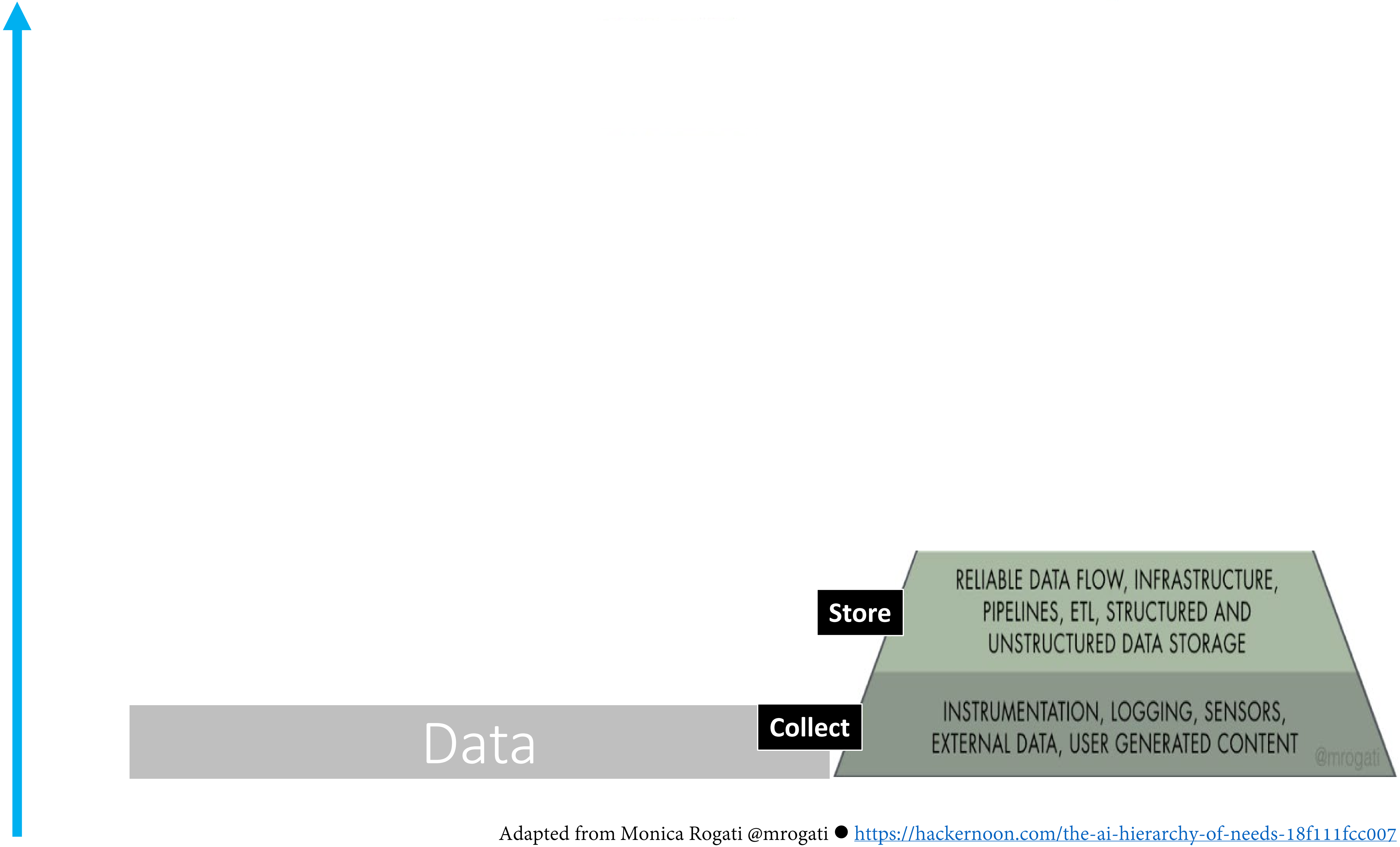



Data

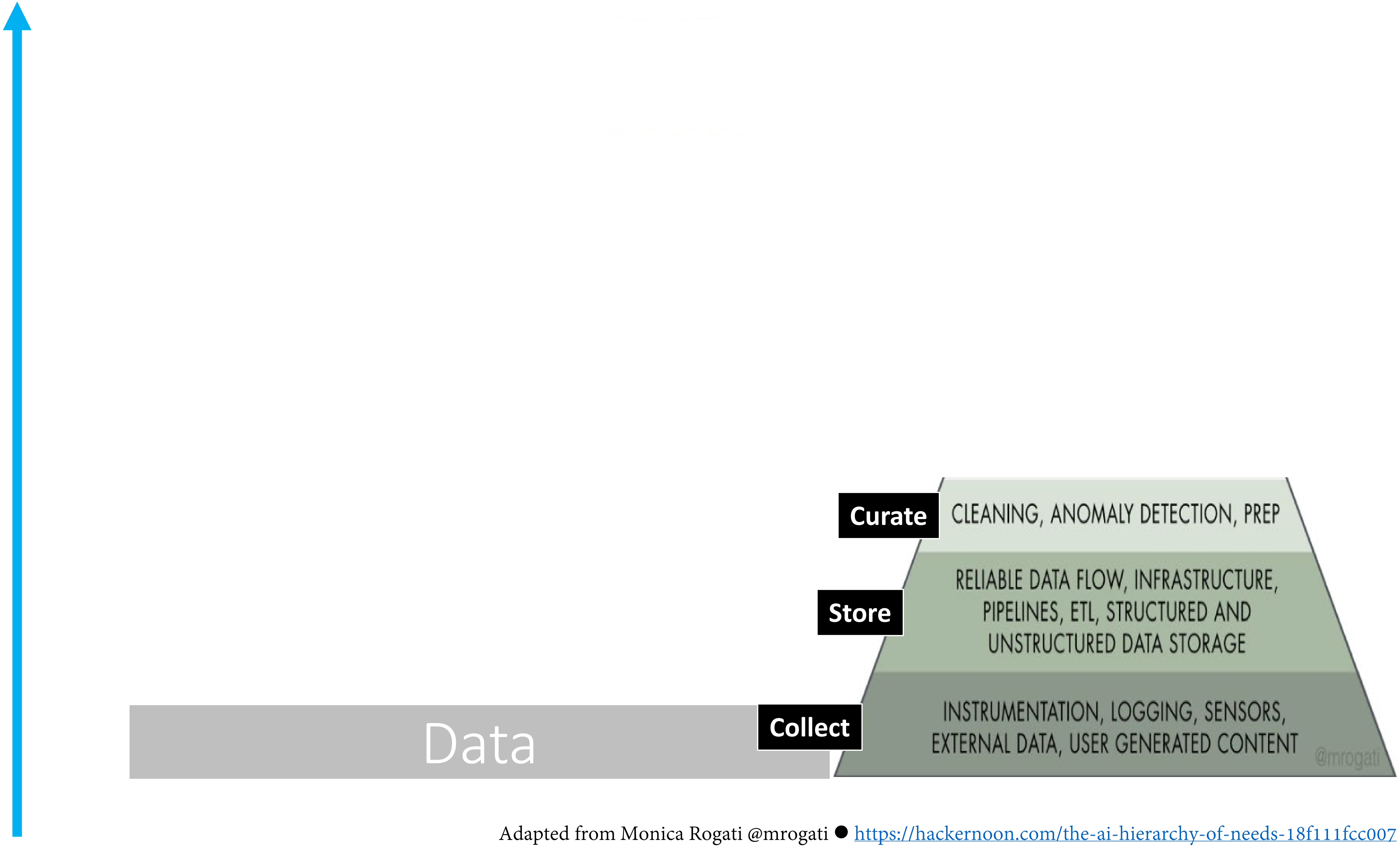
Data

Collect

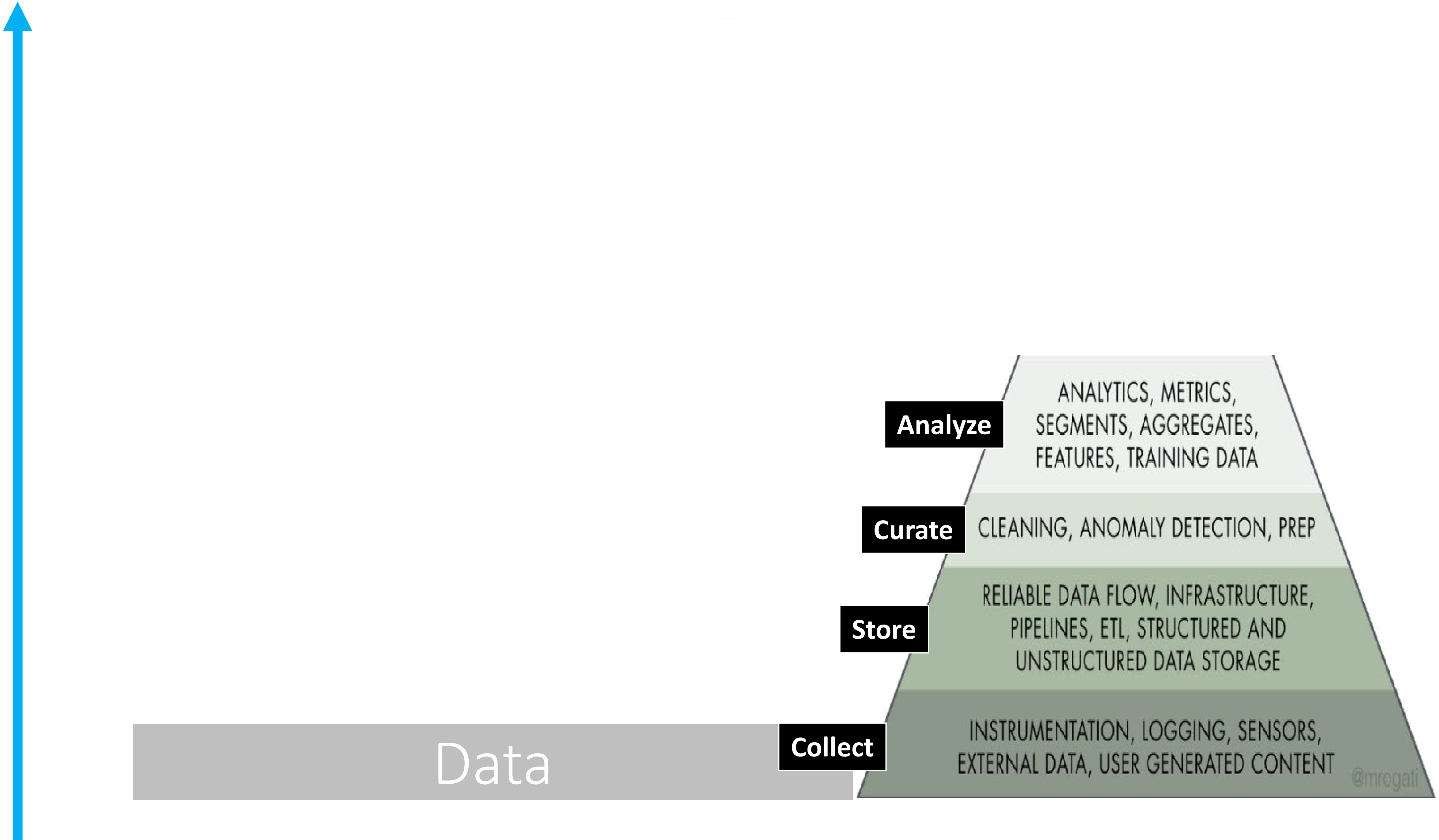
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007



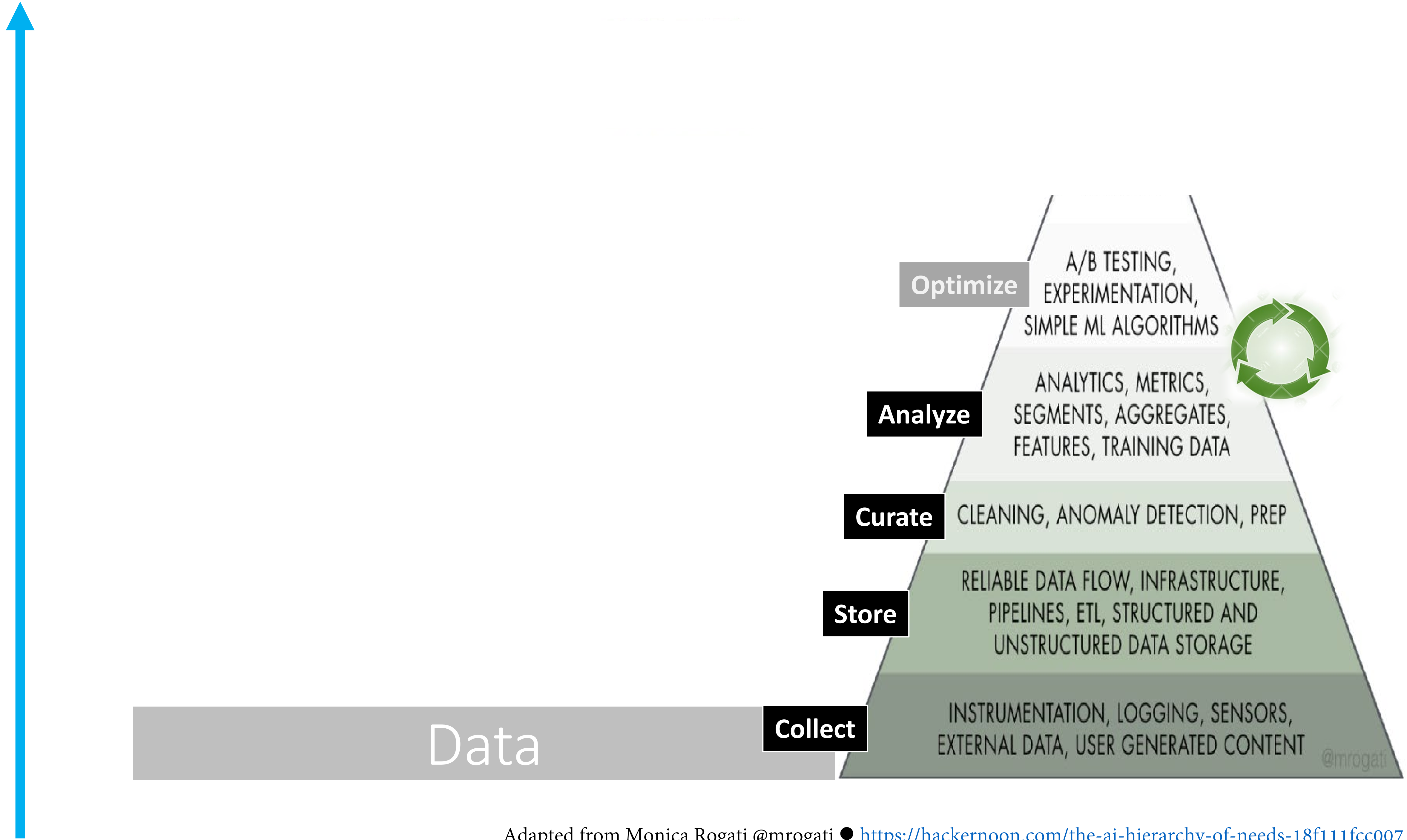
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007



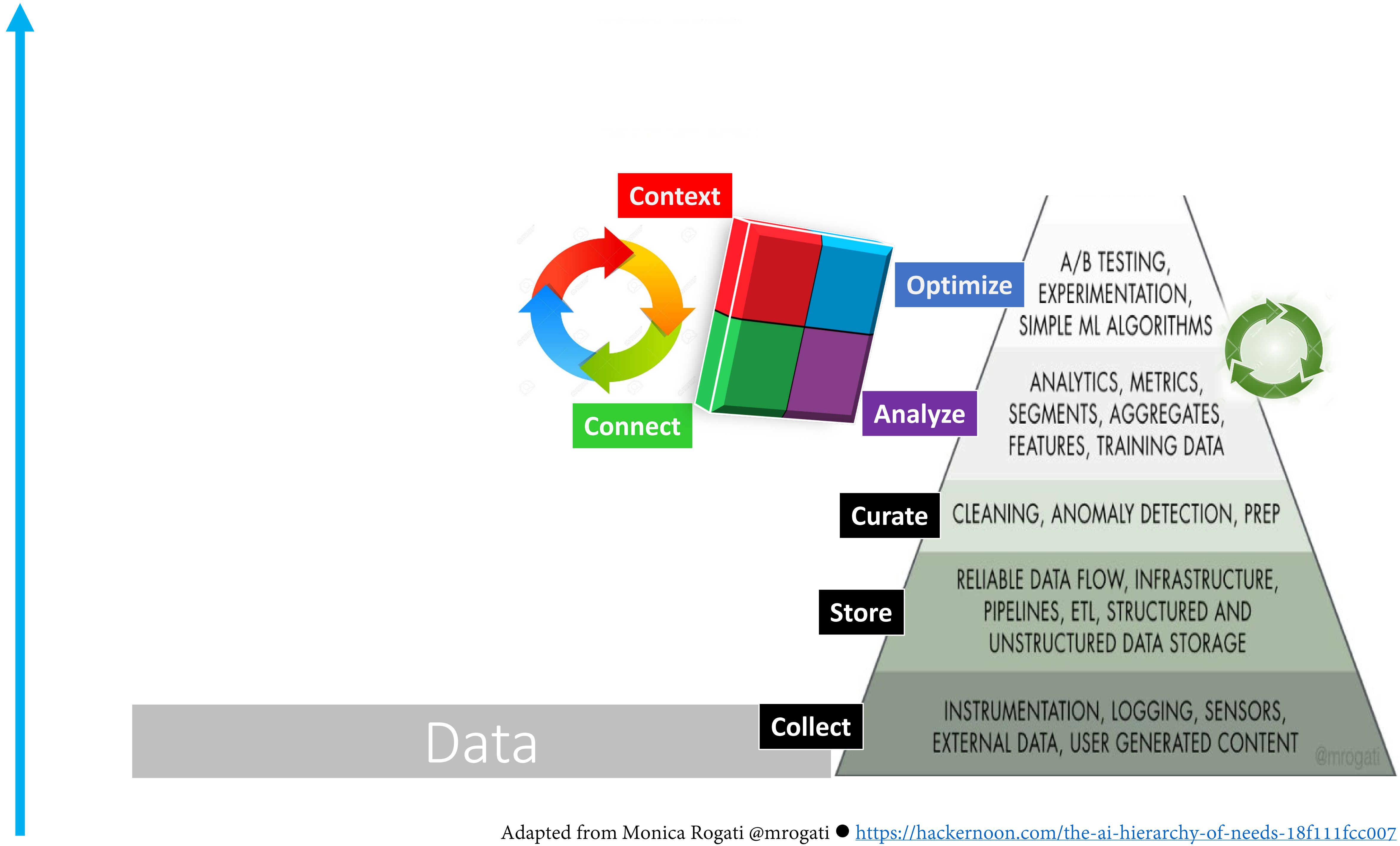
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007



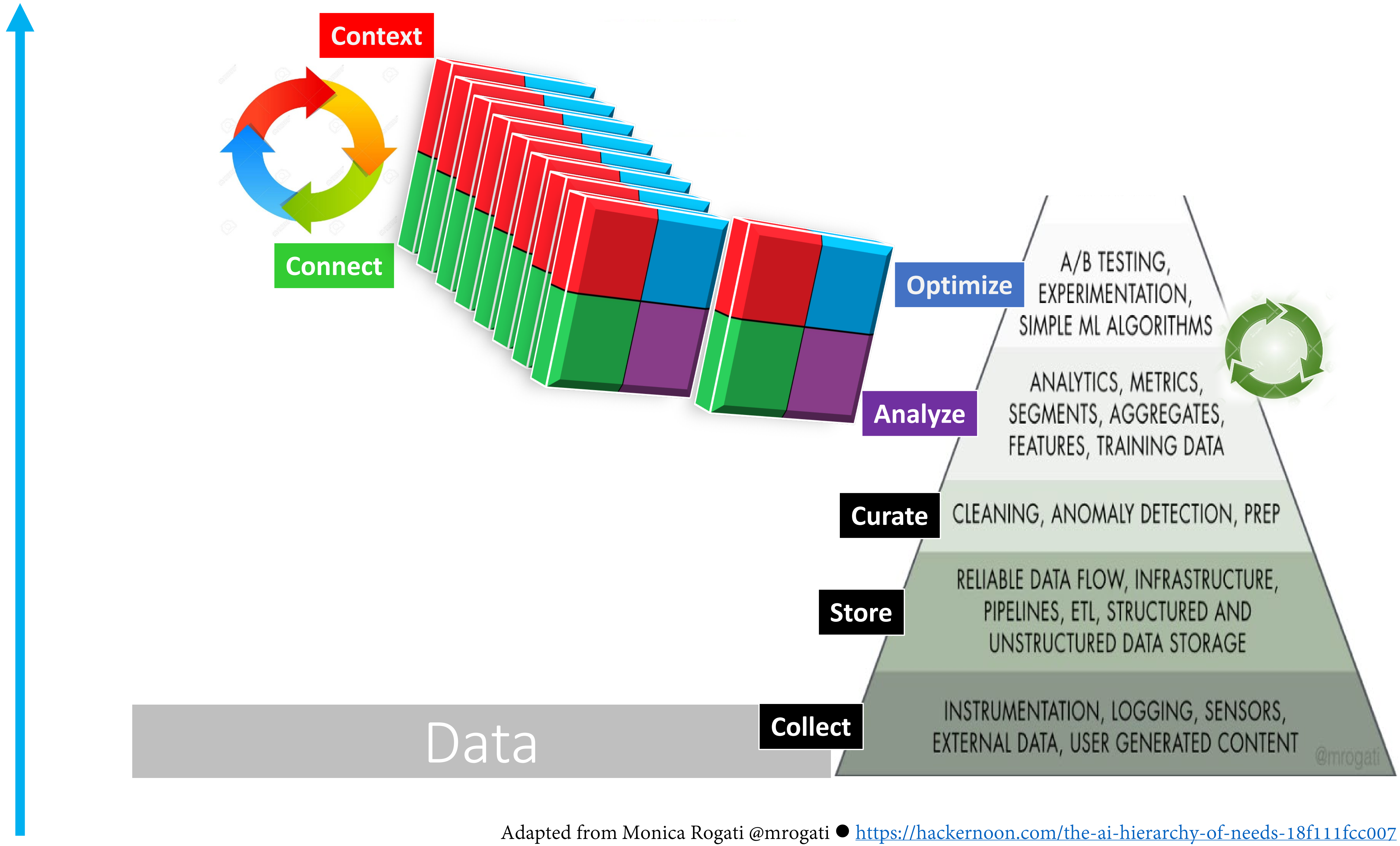
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007



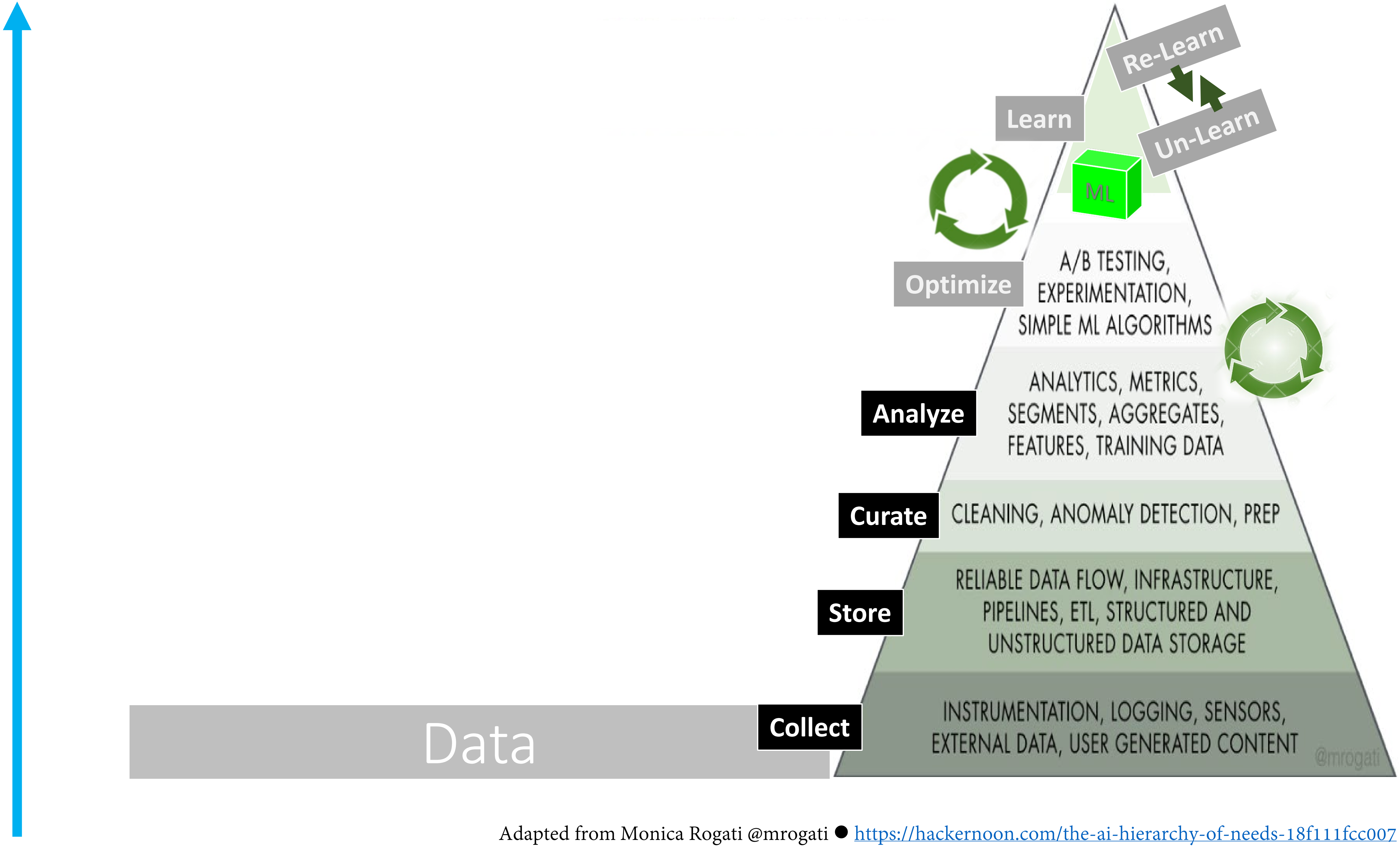
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007



Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

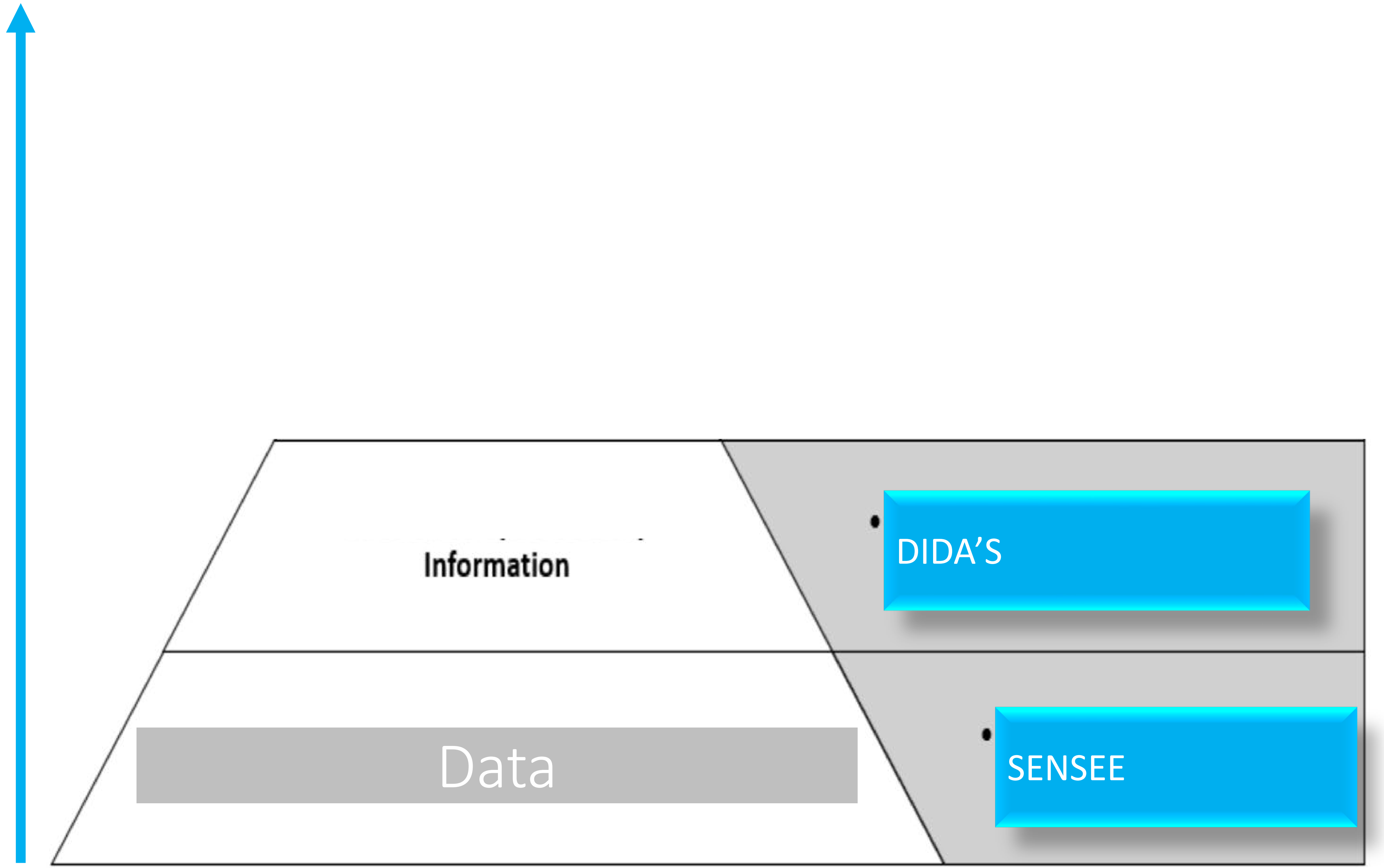


Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

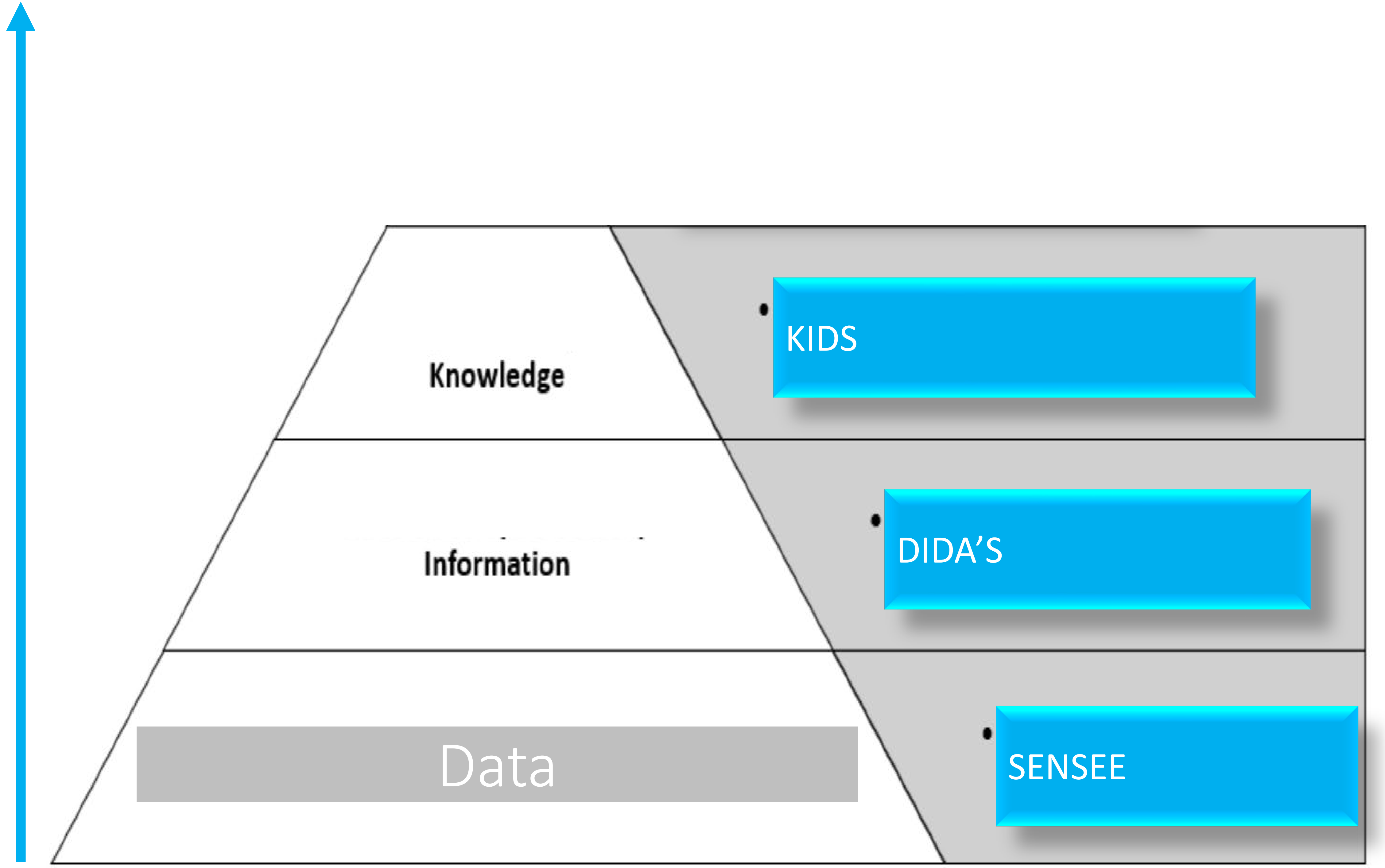


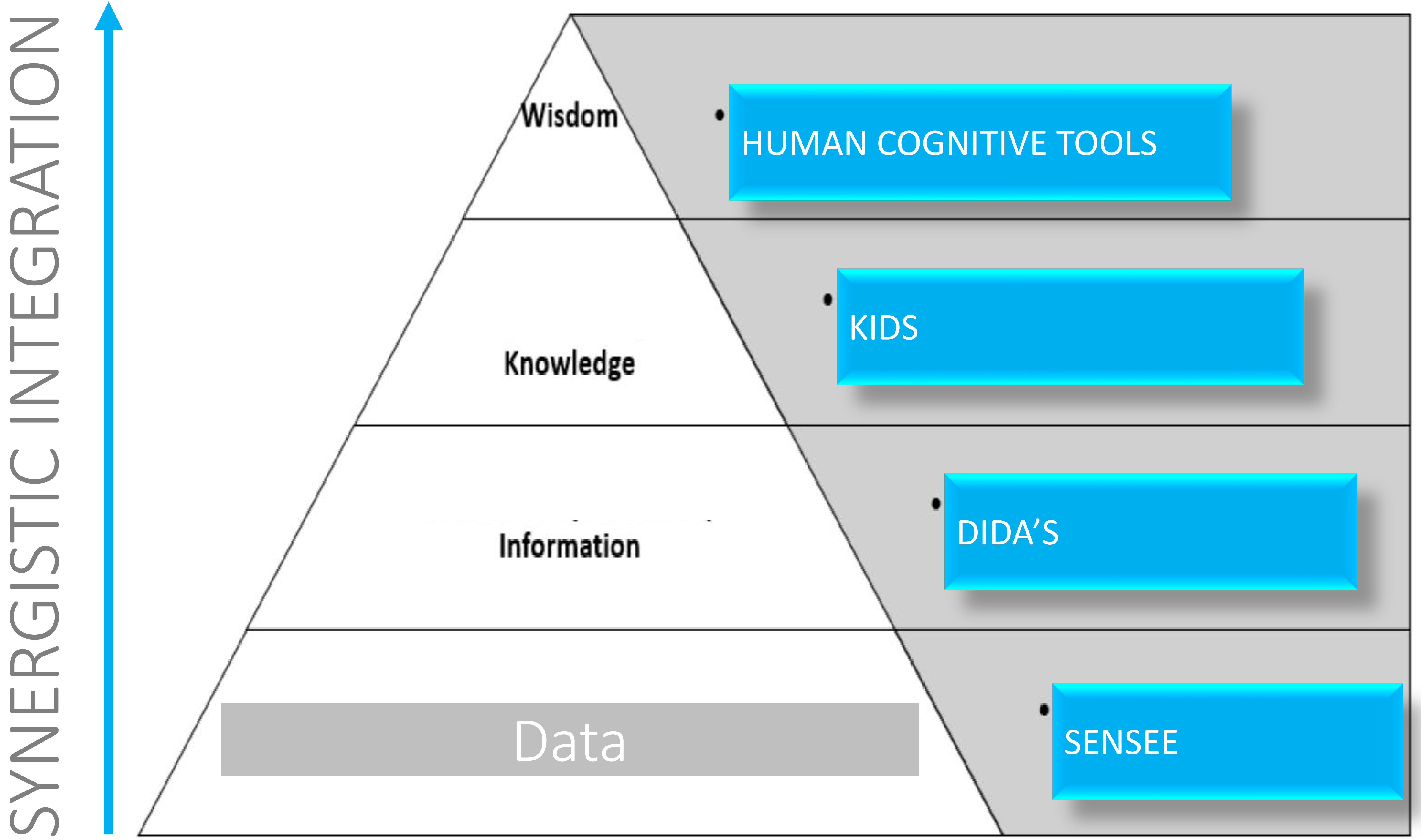
Adapted from Monica Rogati @mrogati ● https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

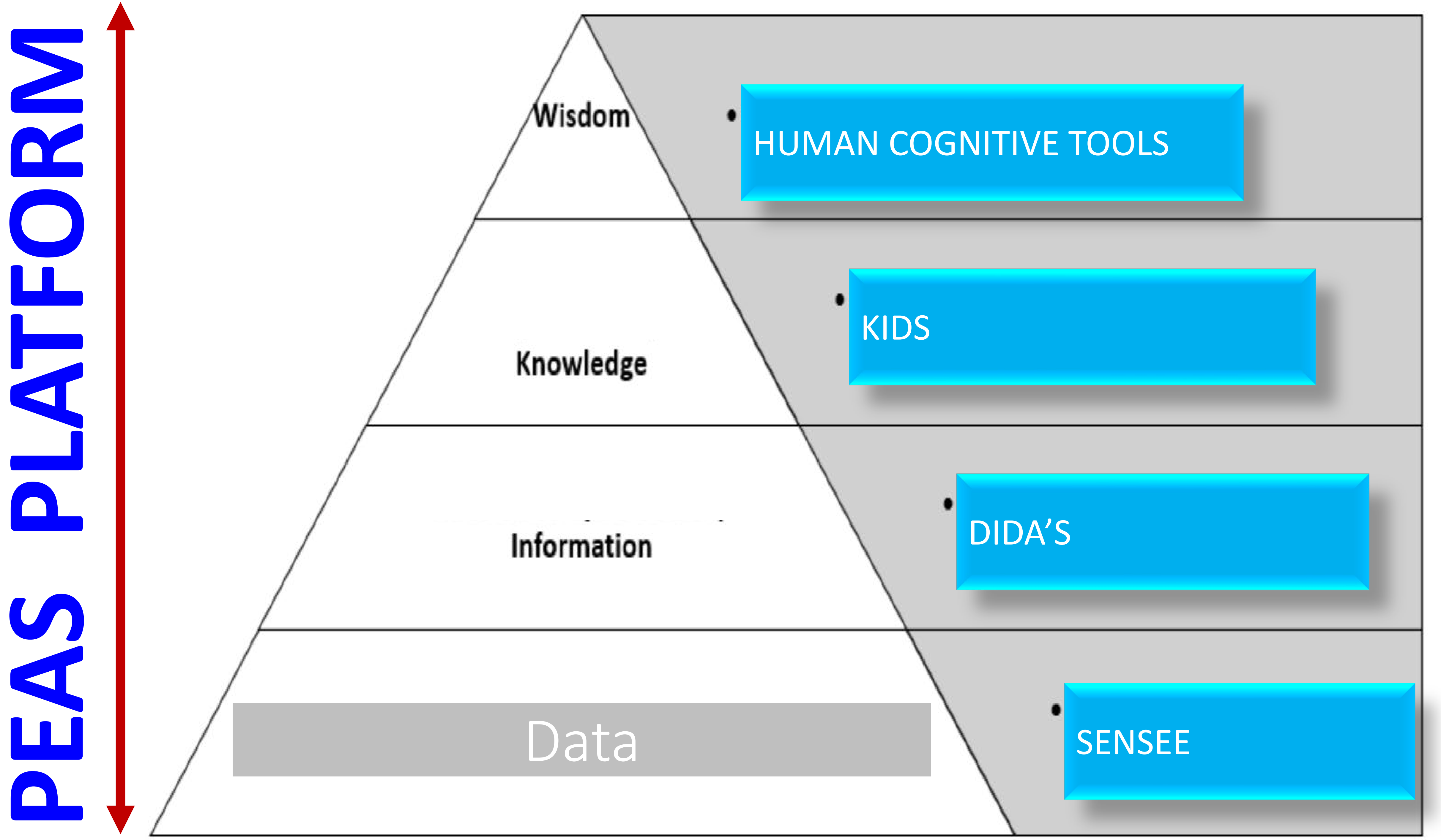

Data
SENSEE



Information
Data
DIDA’S
SENSEE

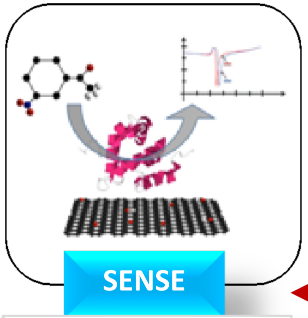

Knowledge
Information
Data
KIDS
DIDA'S
SENSEE

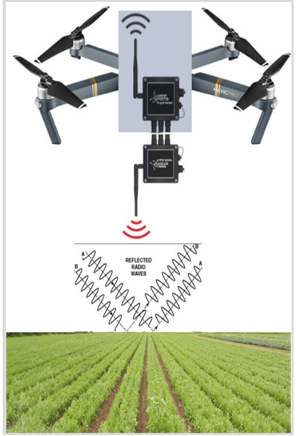

SYNERGISTIC INTEGRATION
Wisdom
Knowledge
Information
Data
HUMAN COGNITIVE TOOLS
KIDS
DIDA'S
SENSEE



PEAS PLATFORM
Wisdom
Knowledge
Information
Data
HUMAN COGNITIVE TOOLS
KIDS
DIDA'S
SENSEE

SENSEE





SENsor **SE**arch **E**ngine

Attempt to create an open source curated repository for different types of sensors created by academic and industrial labs, globally. Expect to connect with similar data from sensor manufacturers.

SENSEE

The curated repository containing descriptions of sensors may also serve as an unit or module for the hypothetical open source library to contain information about tools or technologies related to management of agricultural wastewater systems (AWS).

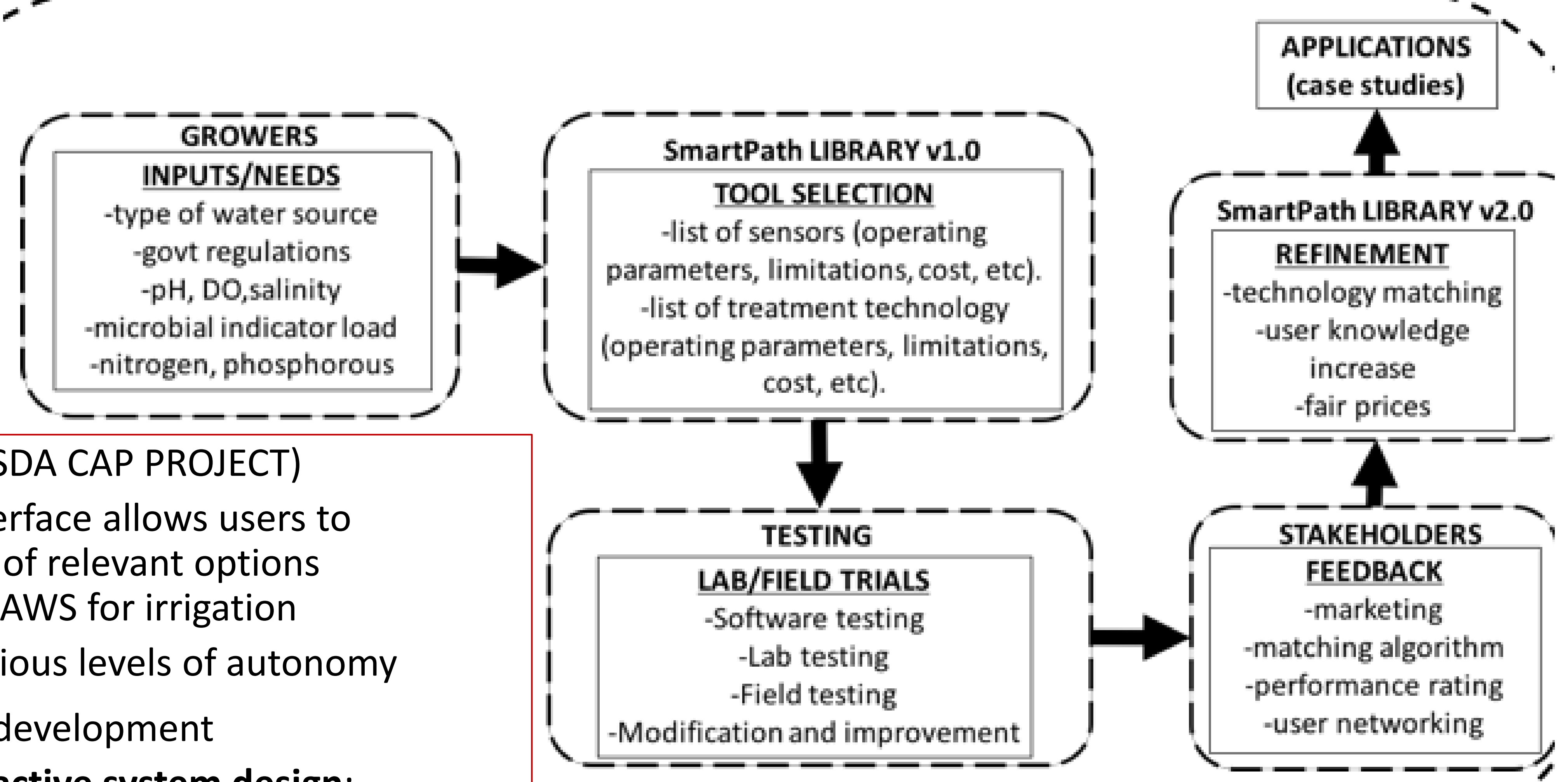
The purpose of the AWS library is to serve end-users (farmers, growers) who may ask questions pertaining to AWS for irrigation. Questions may be related to the detection of heavy metals and microbes (thus, sensors) or waste water treatment technologies (separate module to be developed by USDA SmartPath Project, not a part of SENSEE).

Development of an open source AWS technology repository (with mobile access)

Development of an open source AWS technology repository (with mobile access)

- SmartPath Library (USDA CAP PROJECT)
    - User-friendly interface allows users to explore a variety of relevant options related to use of AWS for irrigation
    - Open source, various levels of autonomy
- Tools/widgets under development
    - Software for **proactive system design**; selection of appropriate technologies (sensors from SENSEE and treatment module)
    - Software for assisting growers with **technology adoption** and assist with trade policies, pricing
    - Software for providing **decision support** for monitoring/treatment (SENSEE, DIDA’S, KIDS)

Development of an open source AWS technology repository (with mobile access)



- SmartPath Library (USDA CAP PROJECT)
  - User-friendly interface allows users to explore a variety of relevant options related to use of AWS for irrigation
  - Open source, various levels of autonomy
- Tools/widgets under development
  - Software for **proactive system design**; selection of appropriate technologies (sensors from SENSEE and treatment module)
  - Software for assisting growers with **technology adoption** and assist with trade policies, pricing
  - Software for providing **decision support** for monitoring/treatment (SENSEE, DIDA'S, KIDS)

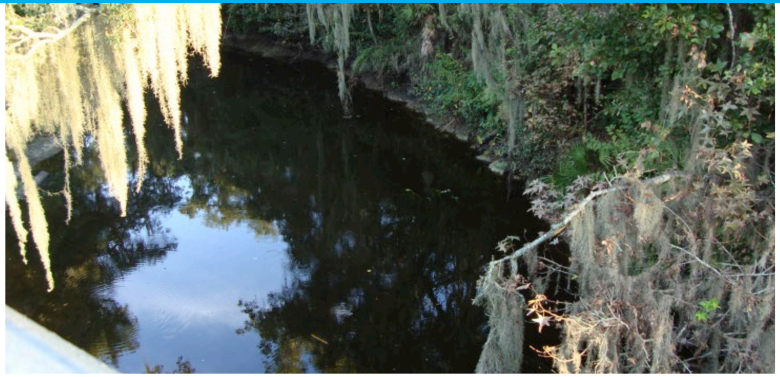
SENSEE

Expert users (academics, students, industrial labs) may be the short term beneficiary from the curated descriptions of sensors, if SENSEE contains a critical mass of sensor descriptions (expectation: sensor descriptions from 1,000 to 10,000 labs, globally). Google search may reveal millions of documents. SENSEE presents a curated list.

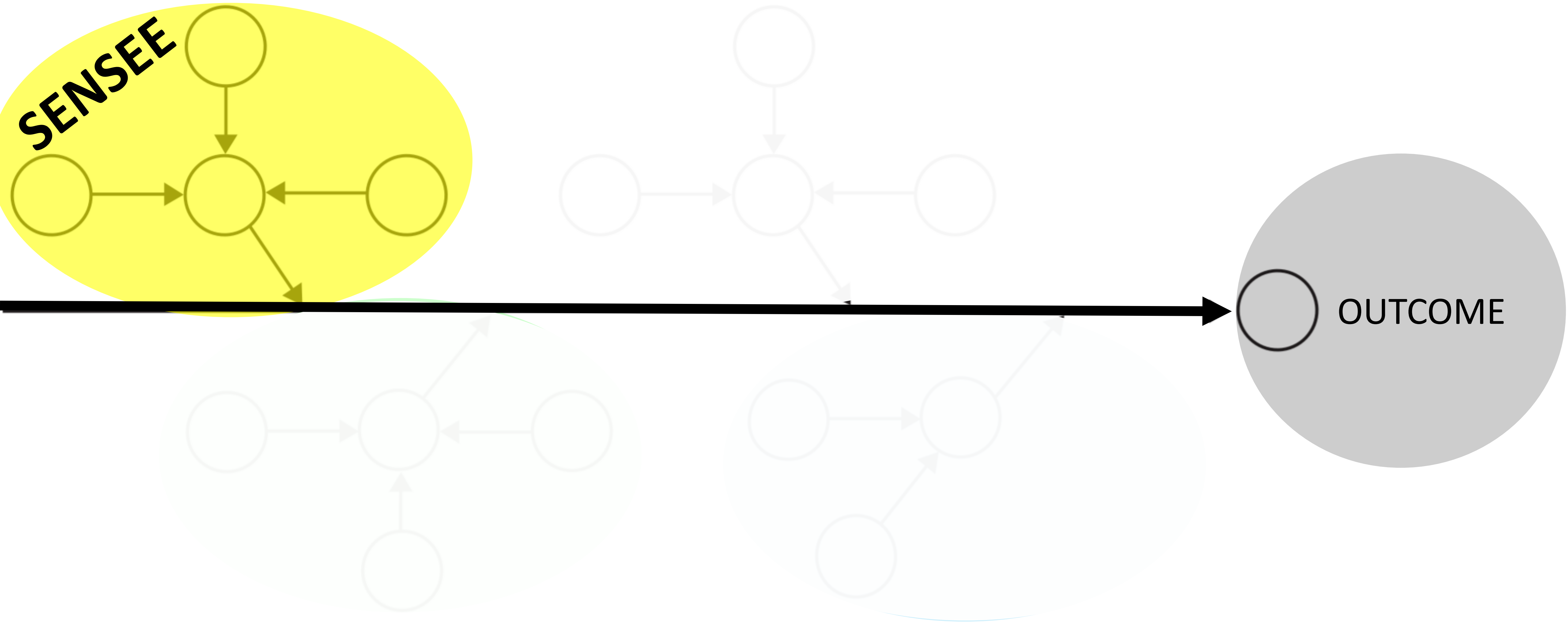
In the future, SENSEE will contain sensor descriptions (categories, attributes) **and** select data from specific sensors. Sensor data ingested in SENSEE will be determined by use cases for specific end-users (farms, grocery stores, food warehouse, food processing plant, packaging operations, retail distribution, food logistics, global transportation).

SENSEE is a start … but woefully inadequate as a solution

- End-user perspective and questions from the field (agro-ecosystem) are complex:
  - Is my ***water quality in compliance*** with FSMA produce safety rule (PSR)?
  - What are the ***costs*** associated with agricultural wastewater (AWS) reuse?
  - Does reuse of AWS add excessive ***management*** issues?
  - ***Who monitors*** return flows, aquifer recharge, and water quality?
  - What are the perceived and real ***health risks*** associated with AWS for irrigation?
  - Can ***technologies*** (sensors + treatment systems) add ***quantifiable*** value?
  - Are there ***legal implications*** of real time water quality data acquisition?
  - Are there ***economic penalties*** for buyers if data log shows poor water quality?

SENSEE is a start … but woefully inadequate as a solution



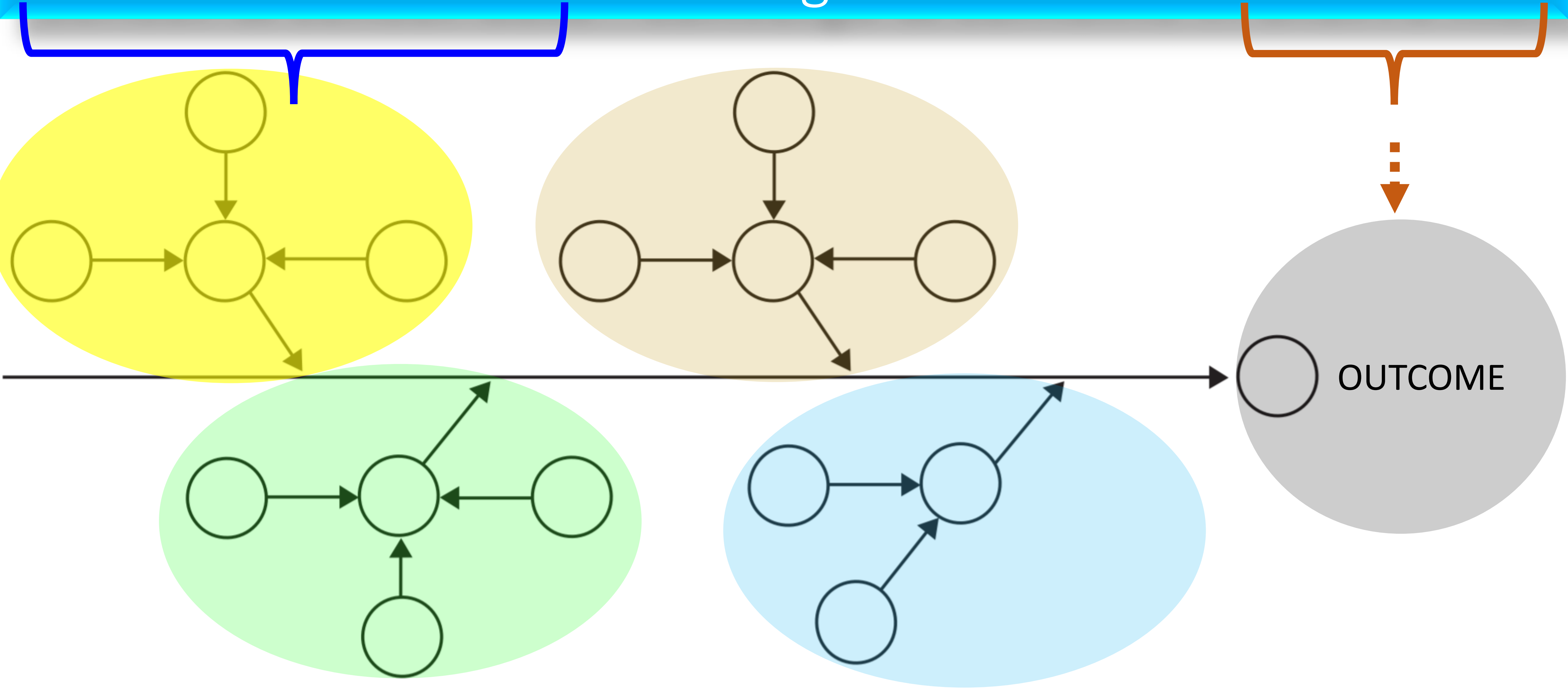
How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

End-user perspective and questions from the field (agro-ecosystem) are complex

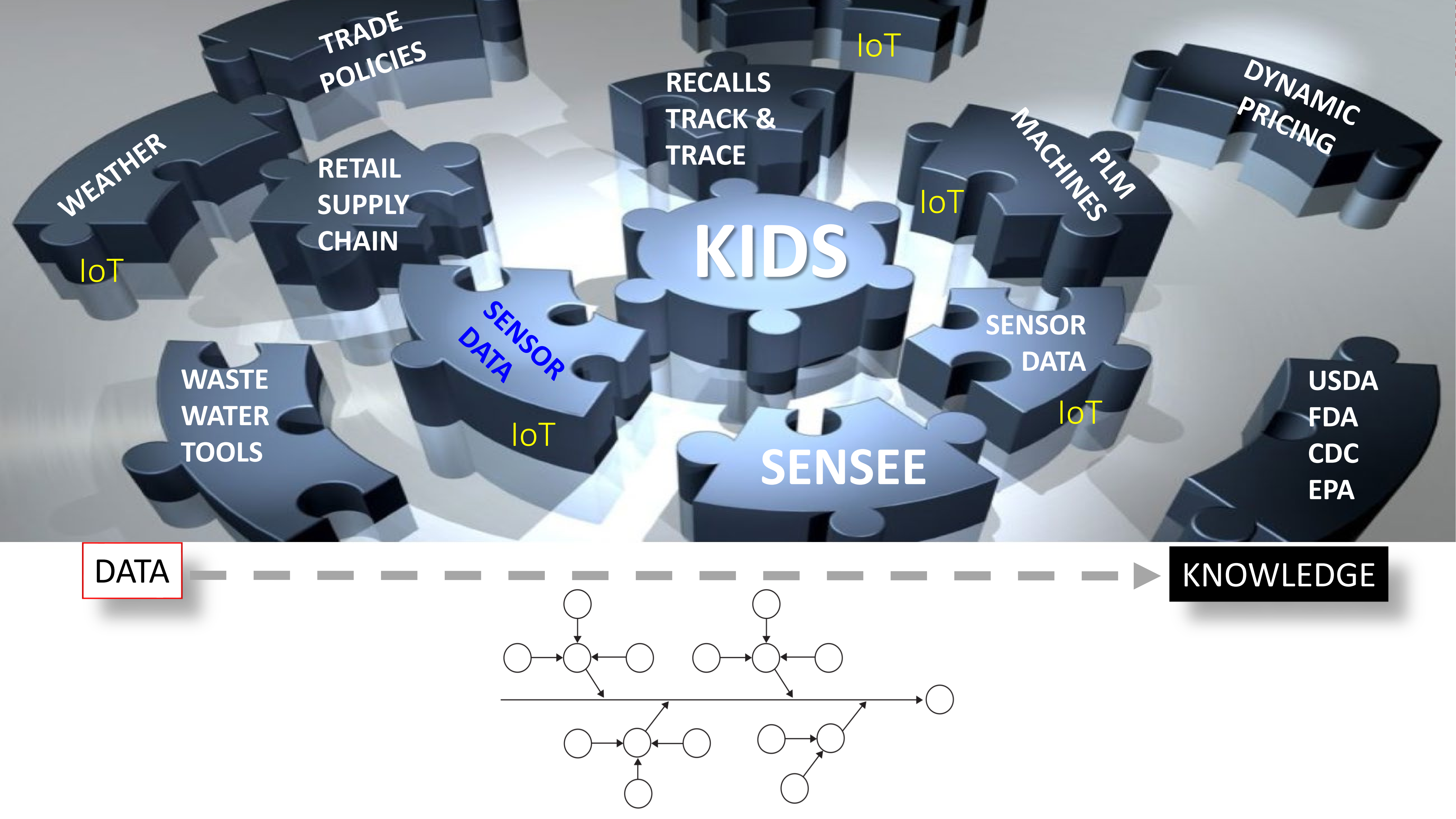
KIDS

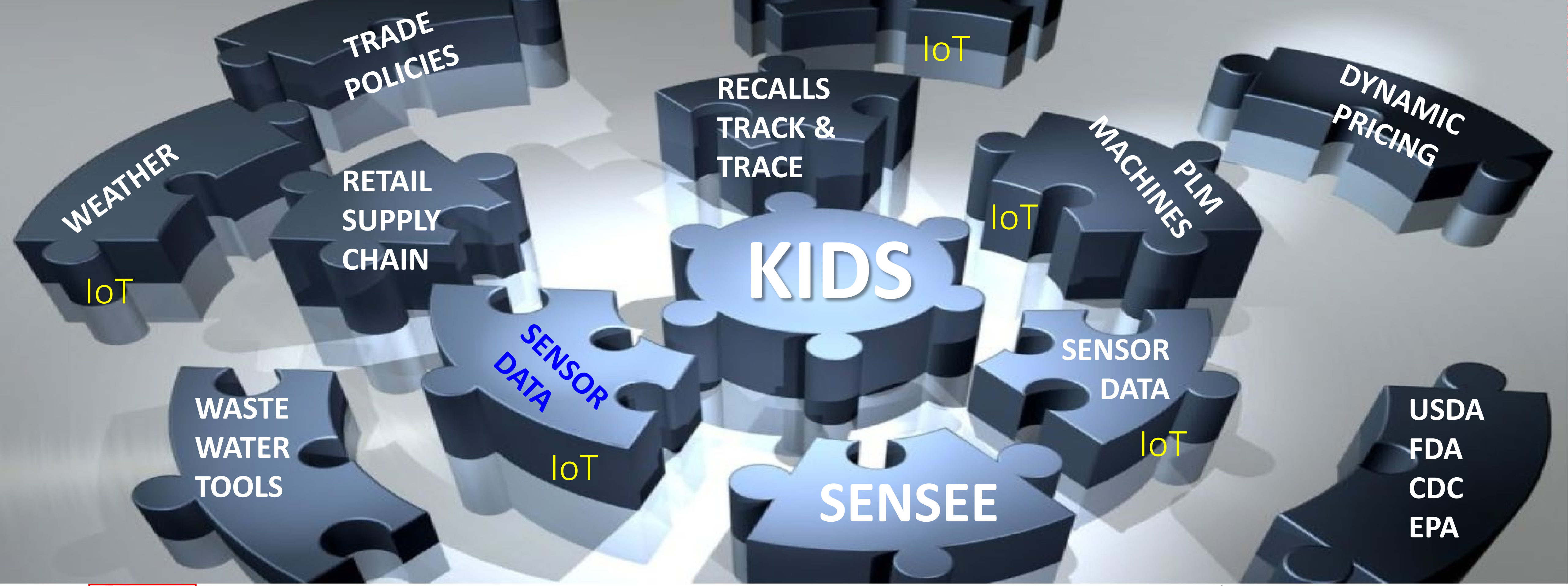
Do you sense the convergence?

Sensor & sensor data - converge & combine in DIDA'S KIDS



How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?


TRADE POLICIES
IoT
RECALLS TRACK & TRACE
DYNAMIC PRICING
WEATHER
RETAIL SUPPLY CHAIN
PLM MACHINES
IoT
KIDS
IoT
SENSOR DATA
SENSOR DATA
WASTE WATER TOOLS
IoT
IoT
USDA FDA CDC EPA
SENSEE
DATA
KNOWLEDGE



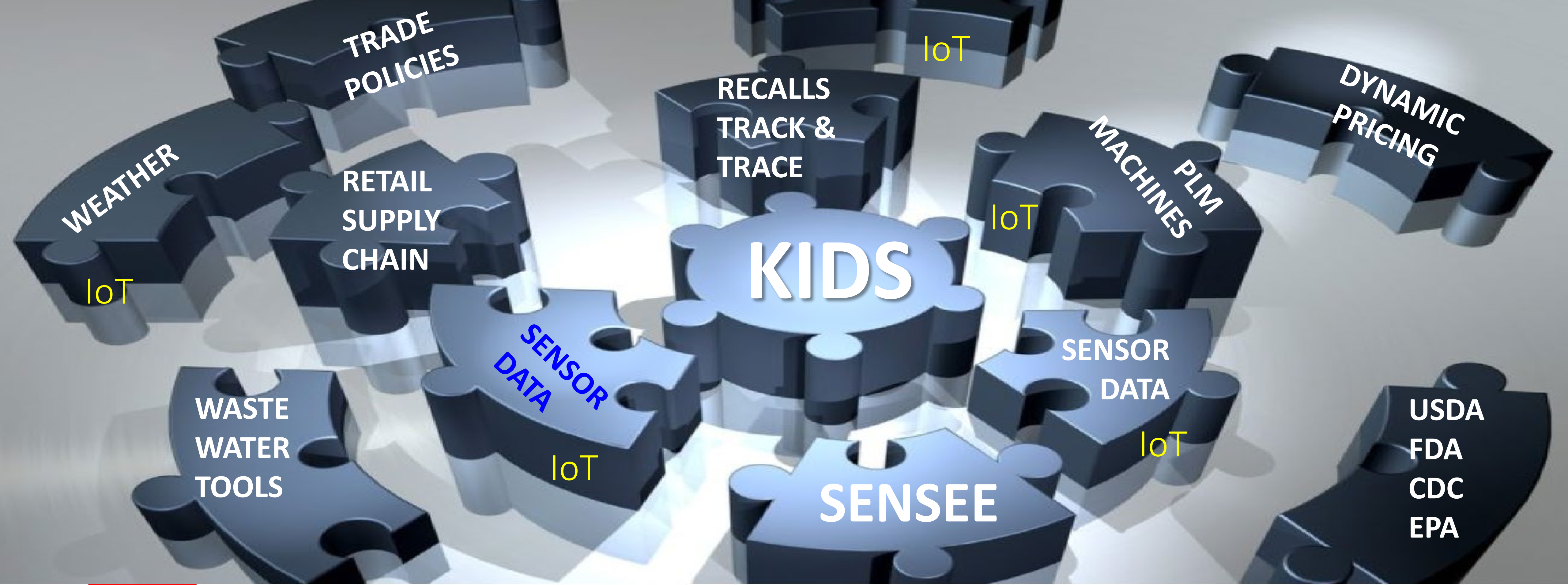
DATA → KNOWLEDGE

KIDS IS NOT A "THING"

ONE SIZE WILL NOT FIT ALL

KIDS IS A CONCEPTUAL DESIGN METAPHOR

KIDS IS A SYNERGISTIC KNOWLEDGE INTEGRATION PLATFORM



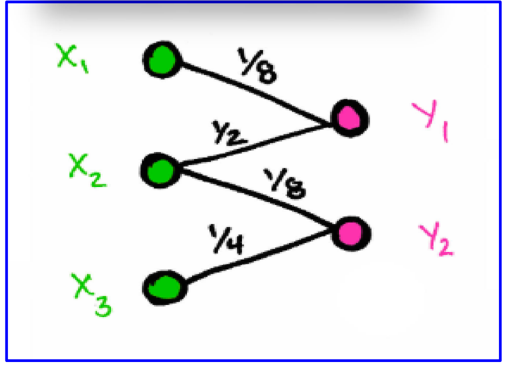
DATA → KNOWLEDGE

End-user perspective and questions from the field (agro-ecosystem) are complex:

- Is my ***water quality in compliance*** with FSMA produce safety rule (PSR)?
- What are the ***costs*** associated with agricultural wastewater (AWS) reuse?
- Does reuse of AWS add excessive ***management*** issues?
- ***Who monitors*** return flows, aquifer recharge, and water quality?
- What are the perceived and real ***health risks*** associated with AWS for irrigation?
- Can ***technologies*** (sensors + treatment systems) add ***quantifiable*** value?
- Are there ***legal implications*** of real time water quality data acquisition?
- Are there ***economic penalties*** for buyers if data log shows poor water quality?

WHY SENSEE IS A TINY PART OF A SYSTEM

It may not be difficult to grasp that the questions from field users demand immense cross-pollination of data and information to converge with knowledge, logic and reasoning, to generate even a basic response.

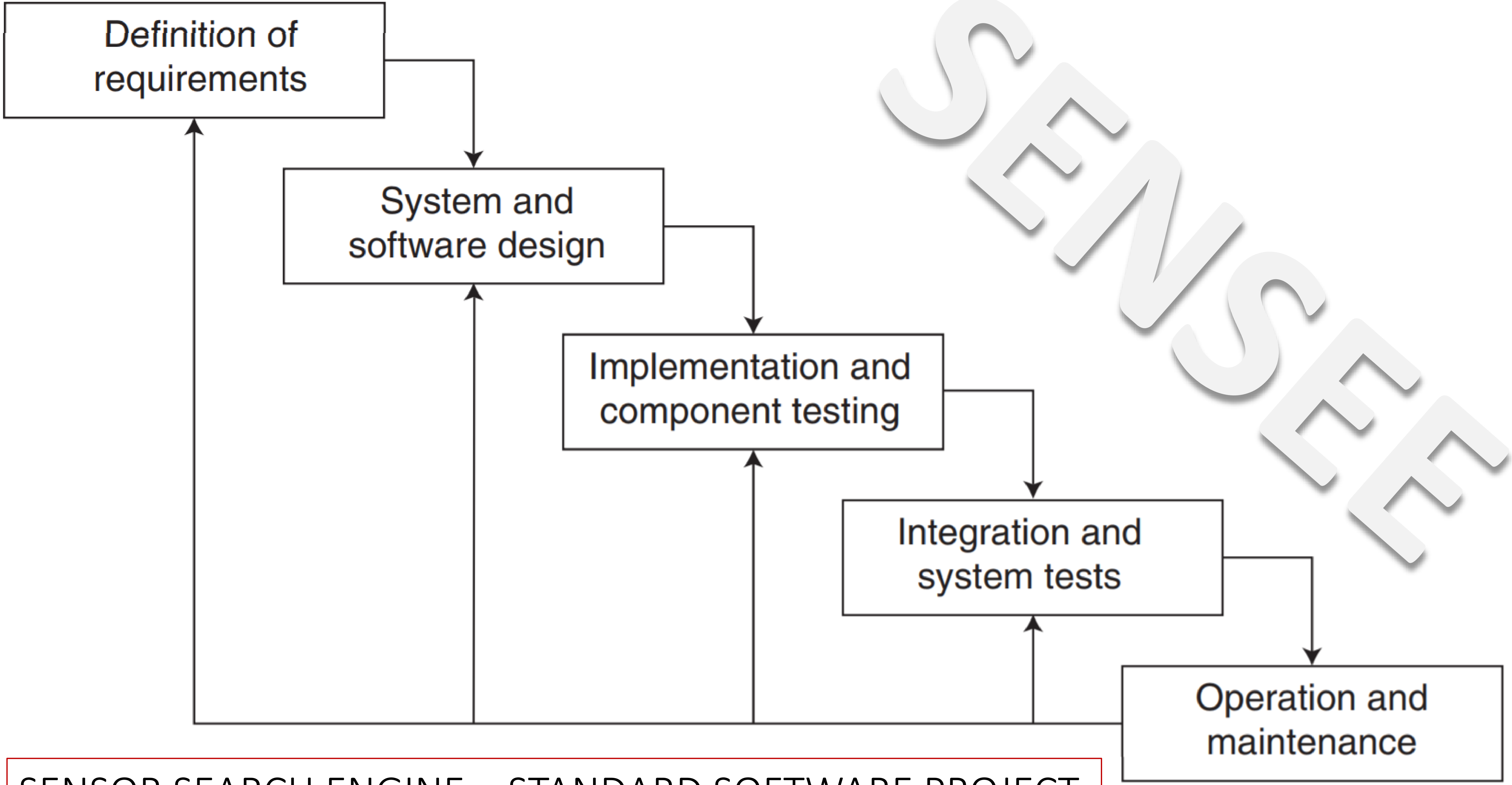


SENSEE

AT THIS TIME, SENSEE 1.0 CONTAINS ONLY SENSOR DESCRIPTIONS (CATEGORIES, ATTRIBUTES). SENSEE CAN ANSWER SELECT QUESTIONS.

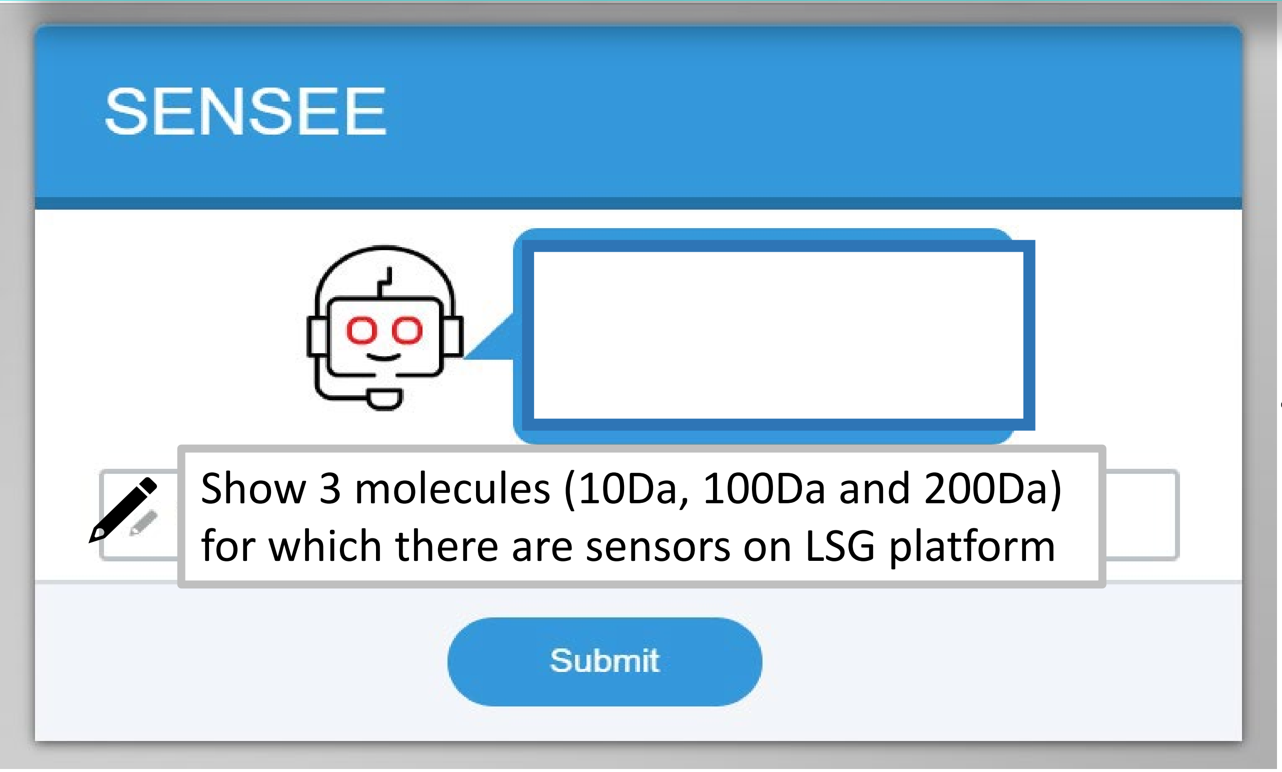
Expert users (academics, students, industrial labs) may benefit from the curated descriptions of sensors. In future, SENSEE may contain a critical mass of sensor descriptions. Google search reveal millions of docs but SENSEE is a curated list.

In the future, SENSEE will contain sensor descriptions (categories, attributes) **<u>and</u>** select **<u>data</u>** from specific sensors. Sensor data ingested in SENSEE will be determined by use cases for specific end-users (farms, grocery stores, food warehouse, food processing, packaging, retail distribution, logistics, agri-business, machine tools, supply chain).



SENSOR SEARCH ENGINE – STANDARD SOFTWARE PROJECT

Experts can query SENSEE and perhaps receive a decent answer, in the near future



SENSEE WEB SERVICE (MOBILE APP EQUIVALENT)

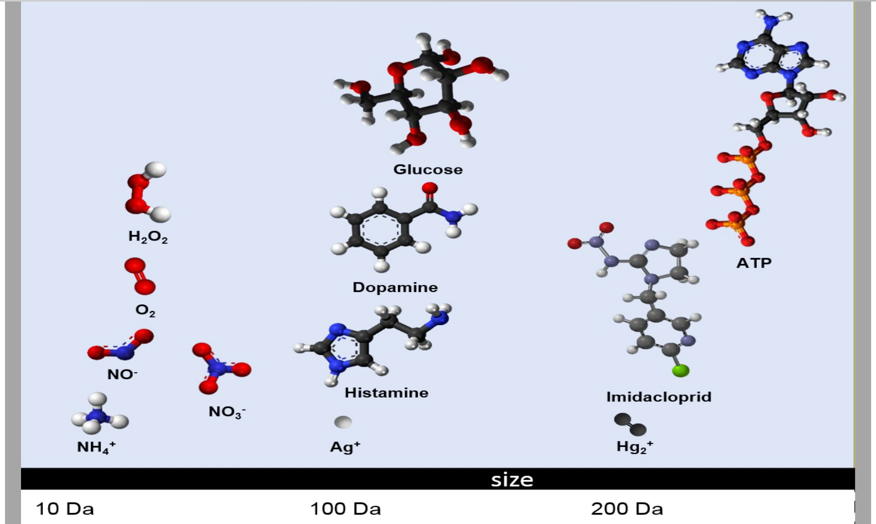
Experts can query SENSEE and perhaps receive a decent answer, in the near future

SENSEE

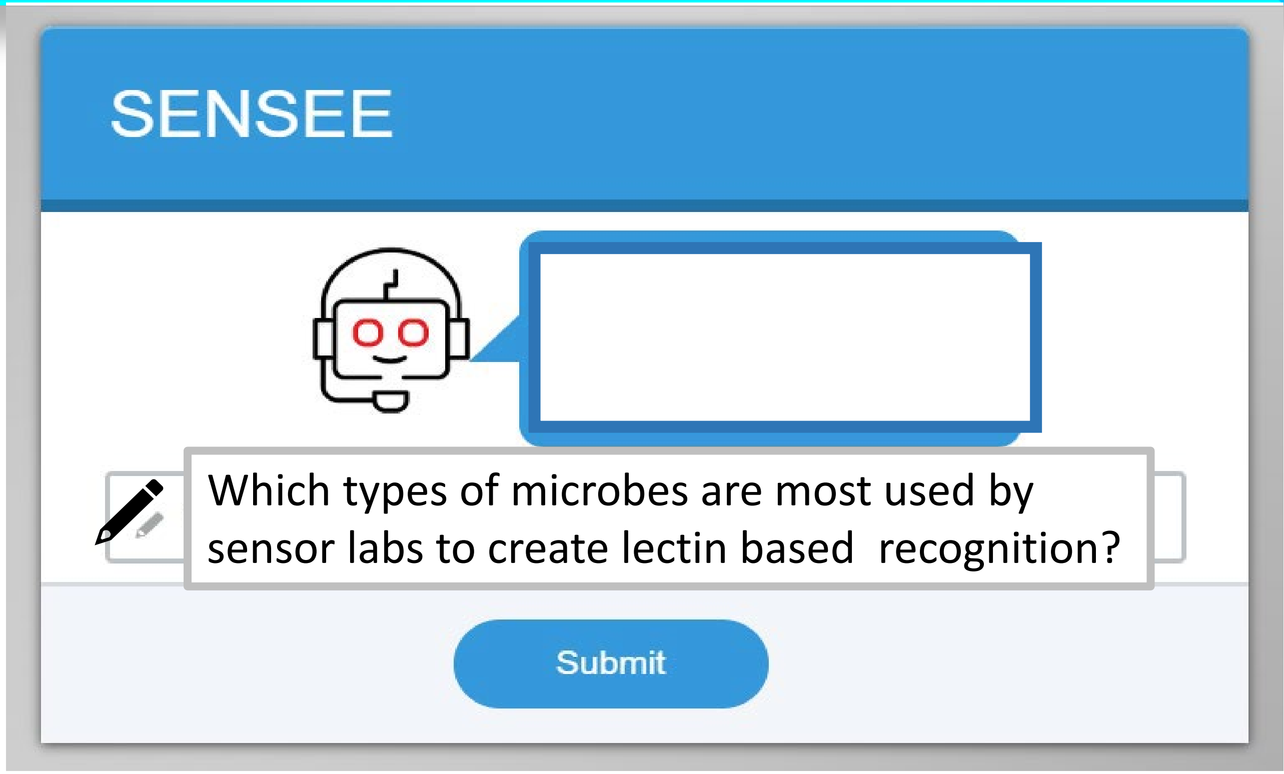
I found 3 classes of molecules. See results.

Show 3 molecules (10Da, 100Da and 200Da) for which there are sensors on LSG platform

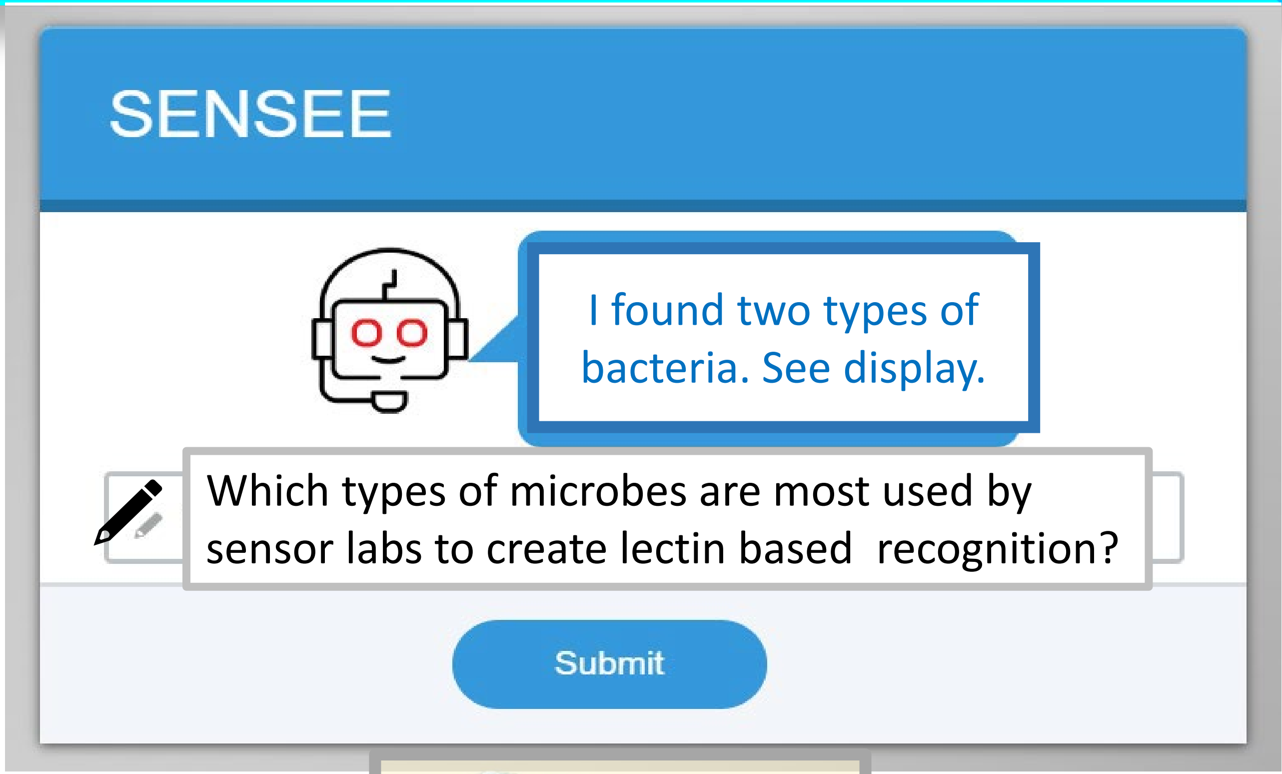
Submit

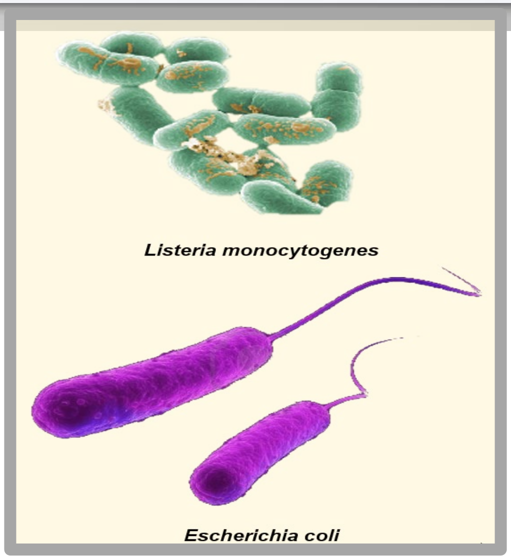


Experts can query SENSEE and perhaps receive a decent answer, in the near future

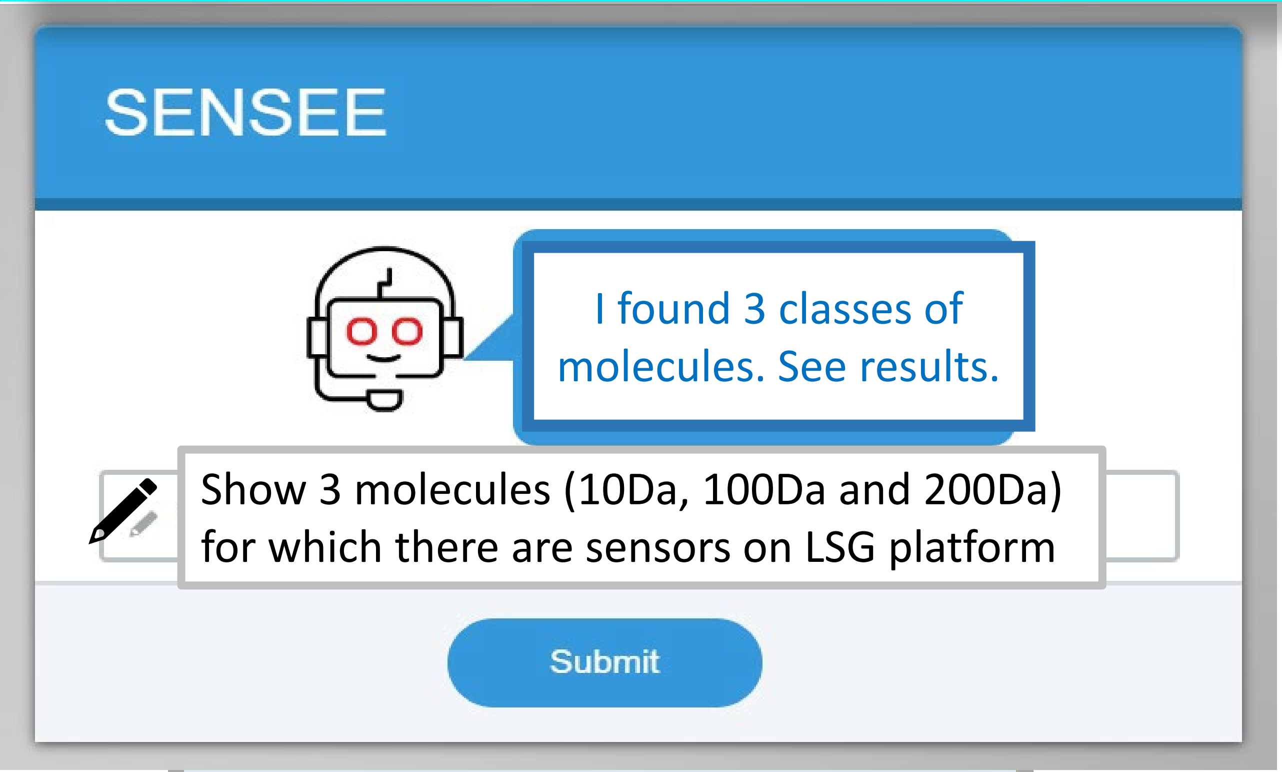


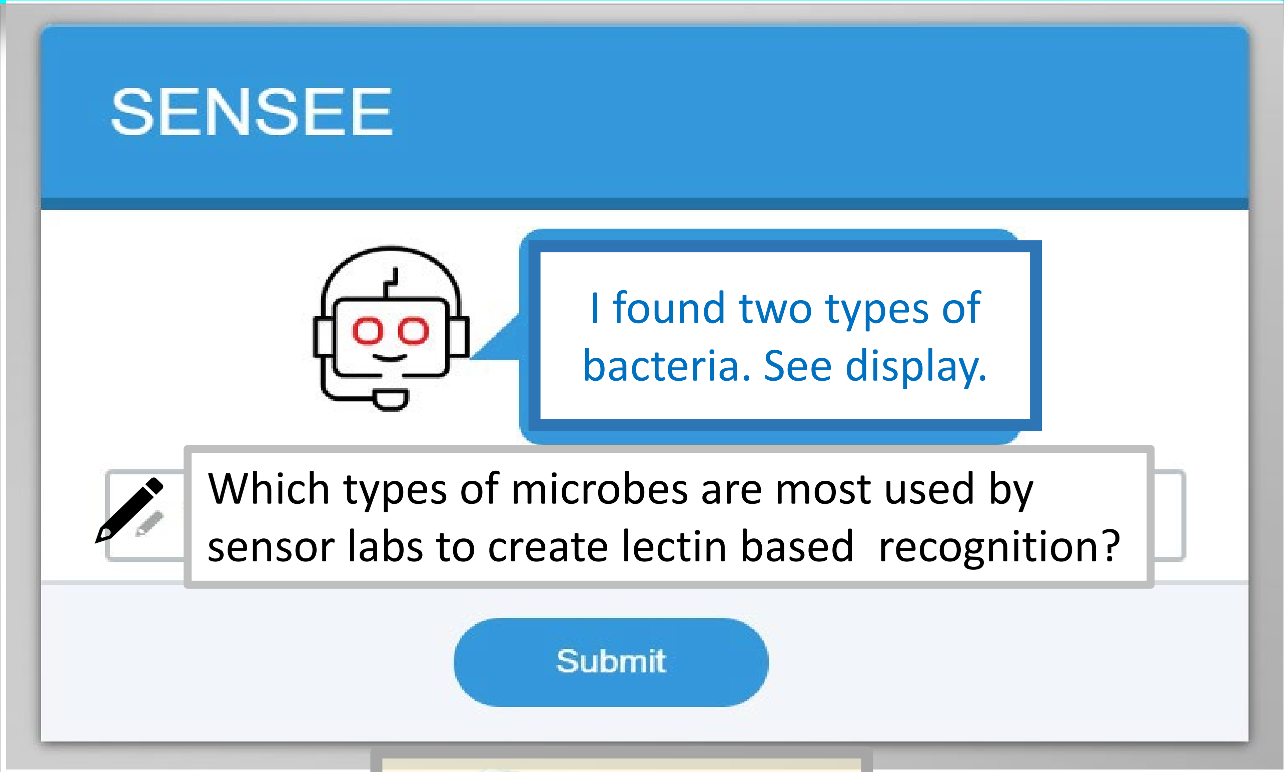
Experts can query SENSEE and perhaps receive a decent answer, in the near future

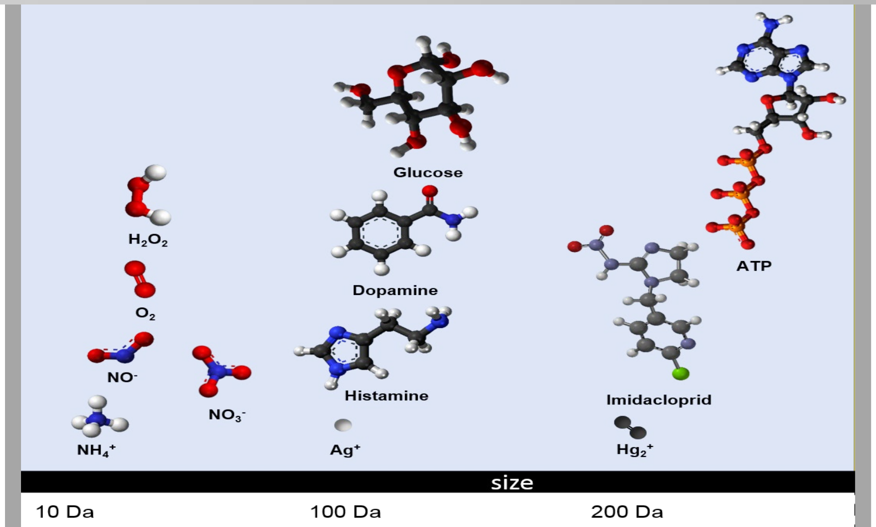


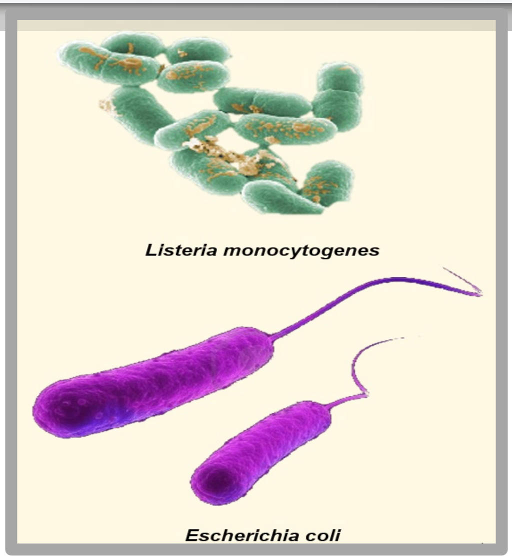


Experts may query SENSEE but is SENSEE capable of facing complex questions?



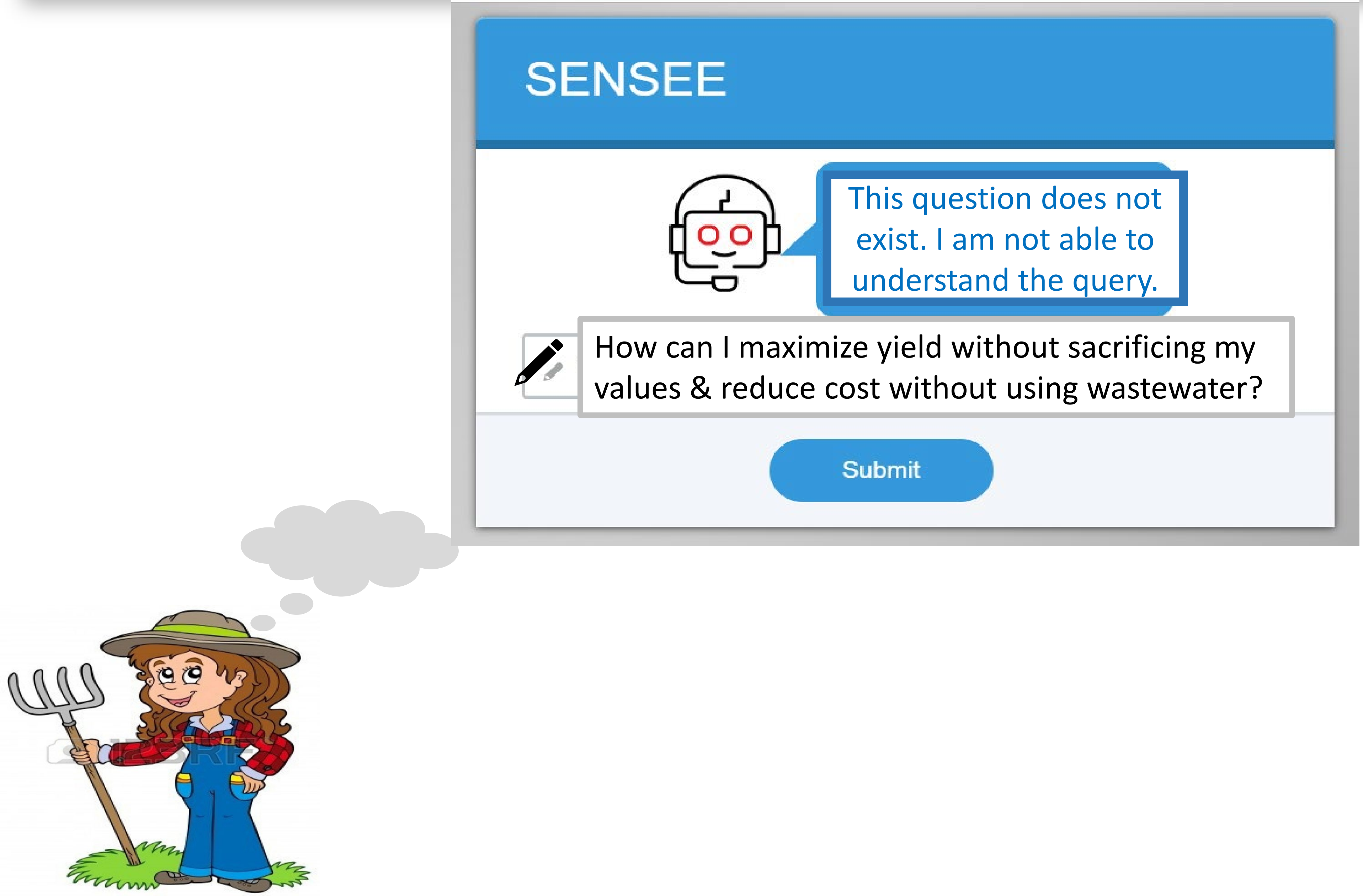






End-user perspective and questions from the field (agro-ecosystem) are complex

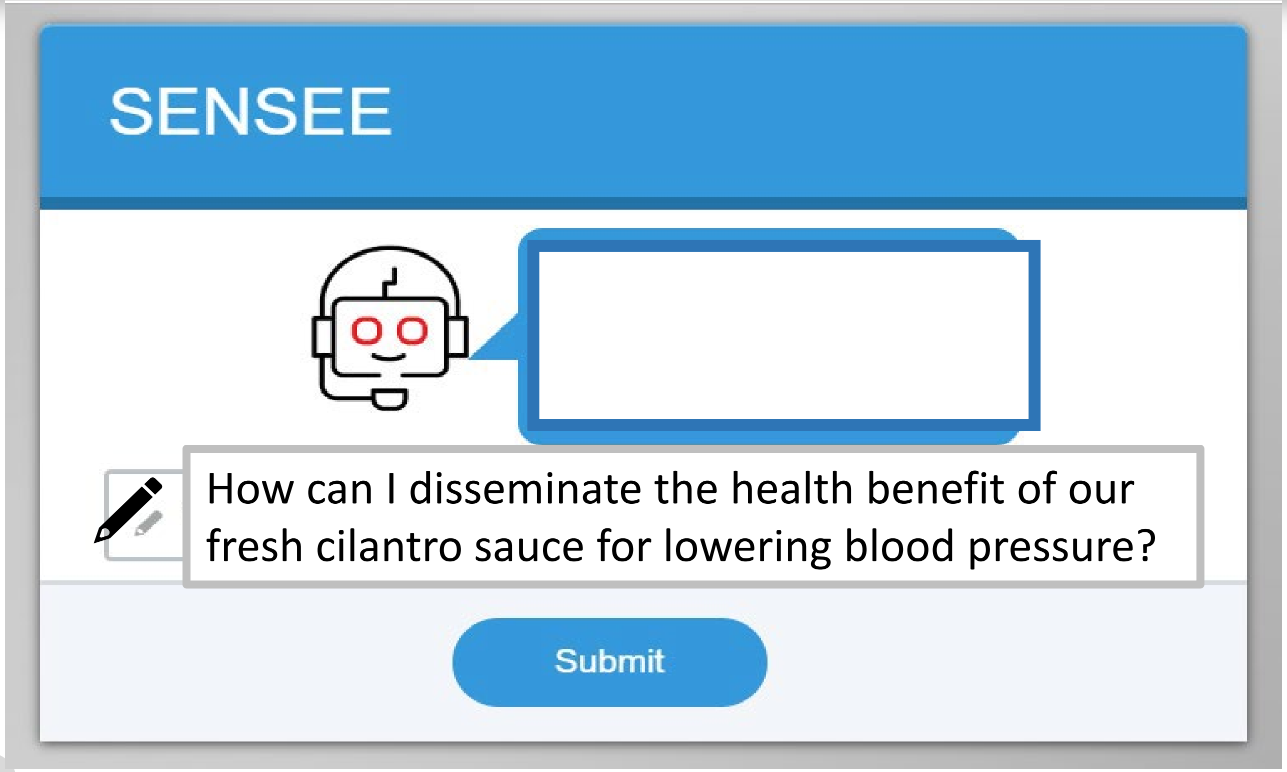
End-user perspective and questions from the field (agro-ecosystem) are complex



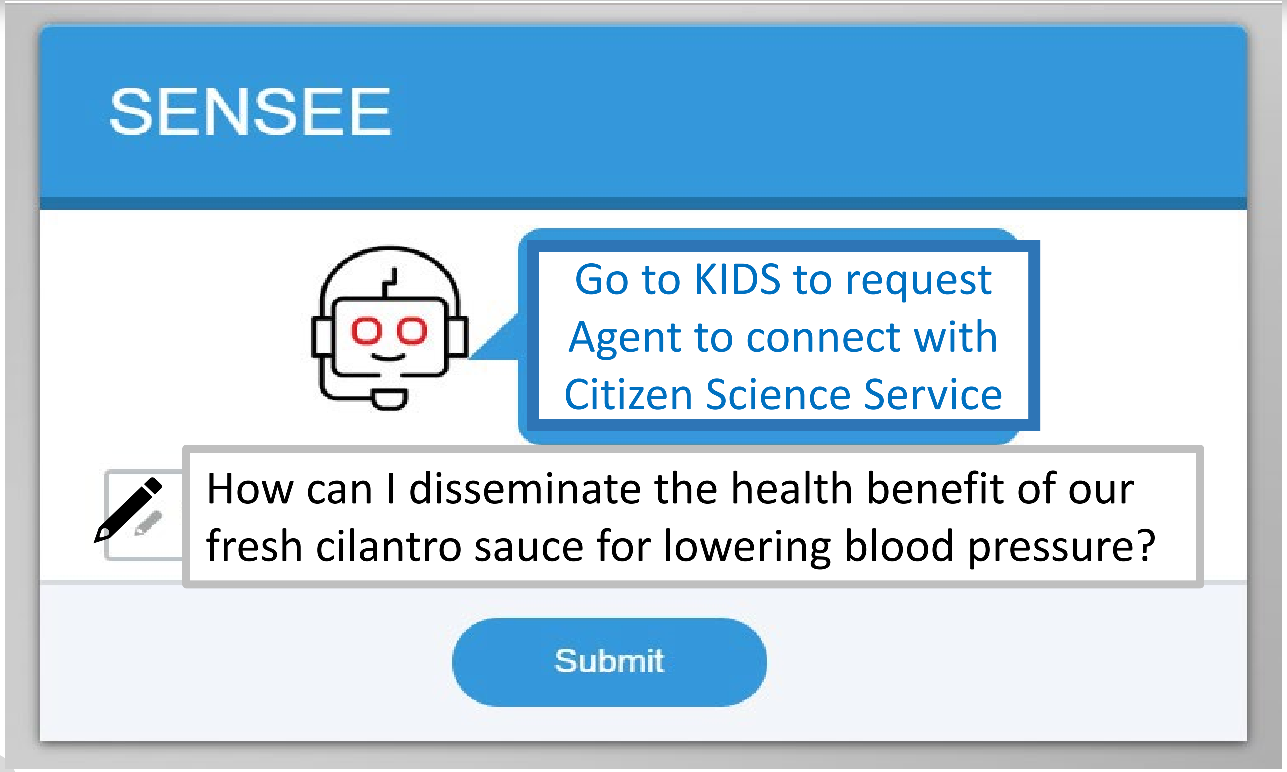
SENSEE 1.0 CANNOT HELP

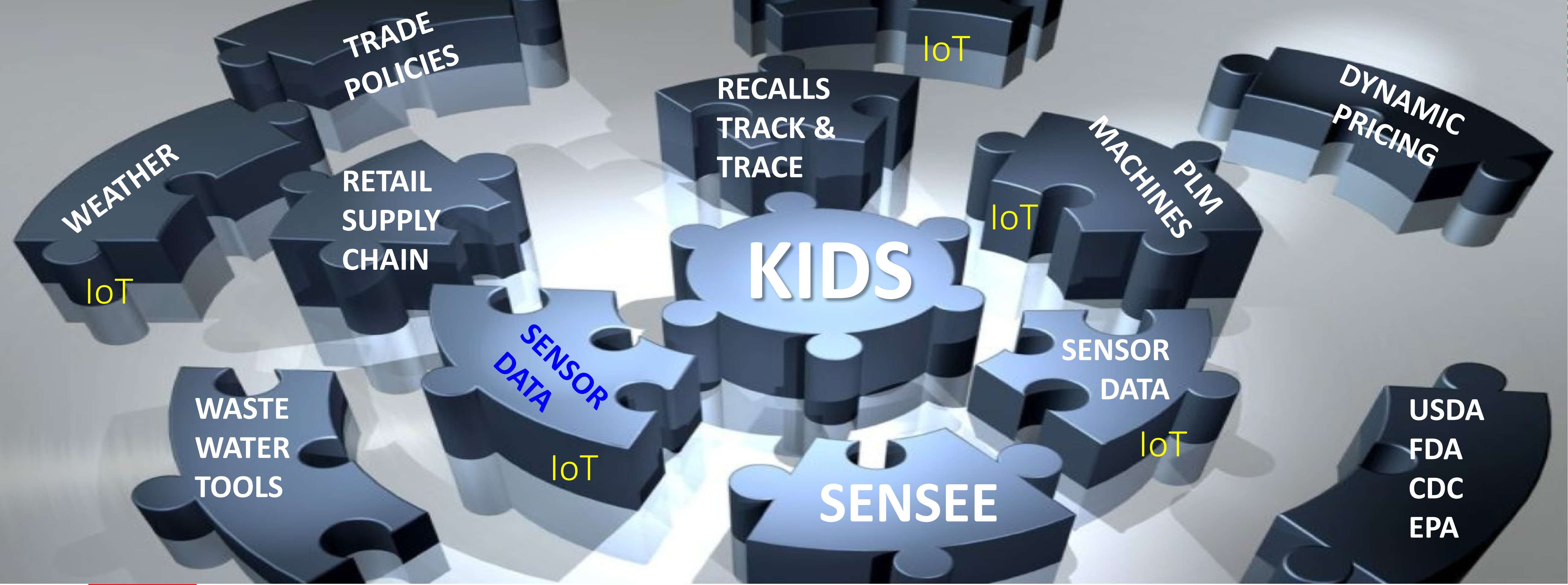
CAN ART HELP? CAN KIDS HELP?

End-user perspective and questions from the field (agro-ecosystem) are complex



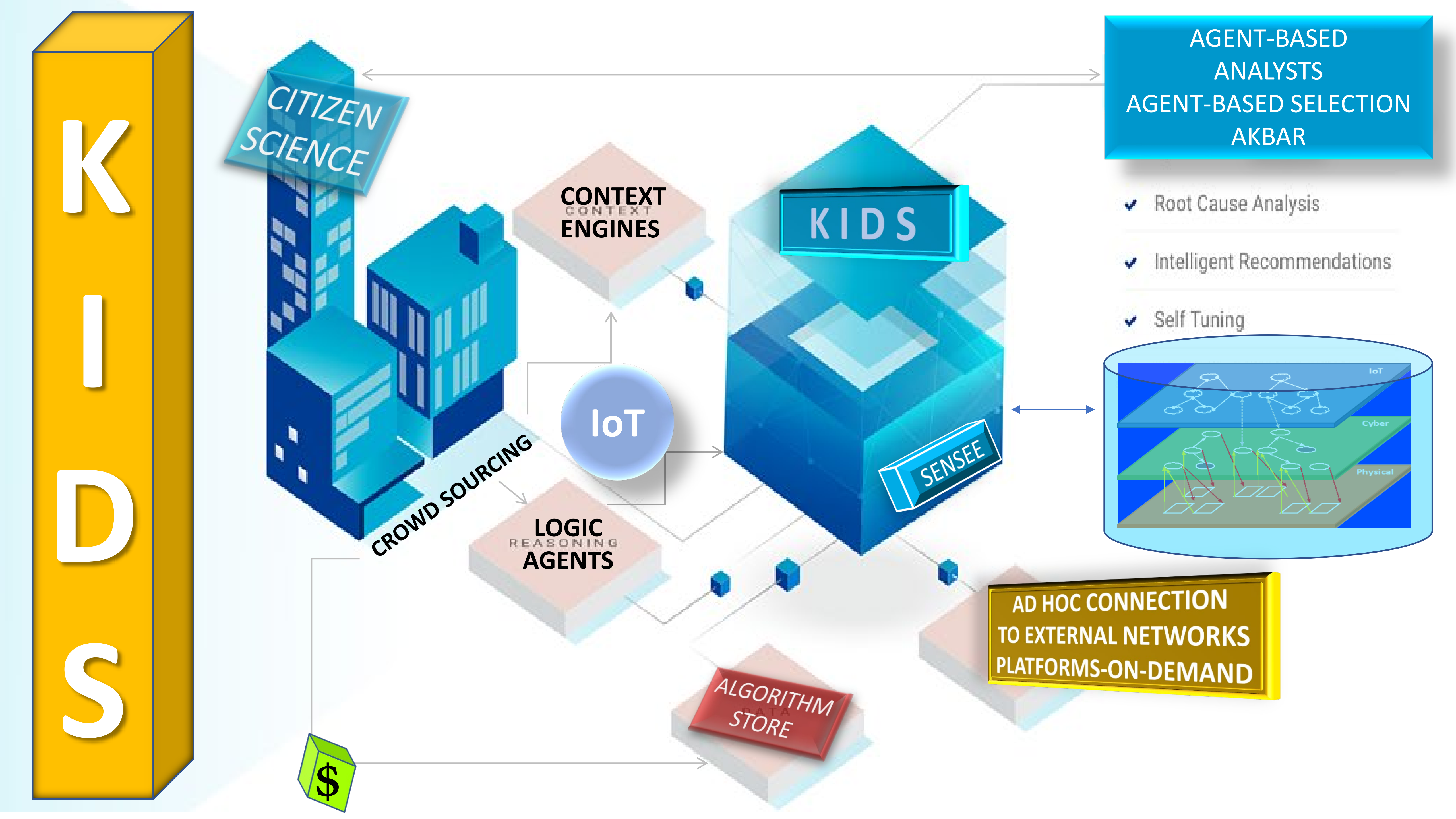
End-user perspective and questions from the field (agro-ecosystem) are complex



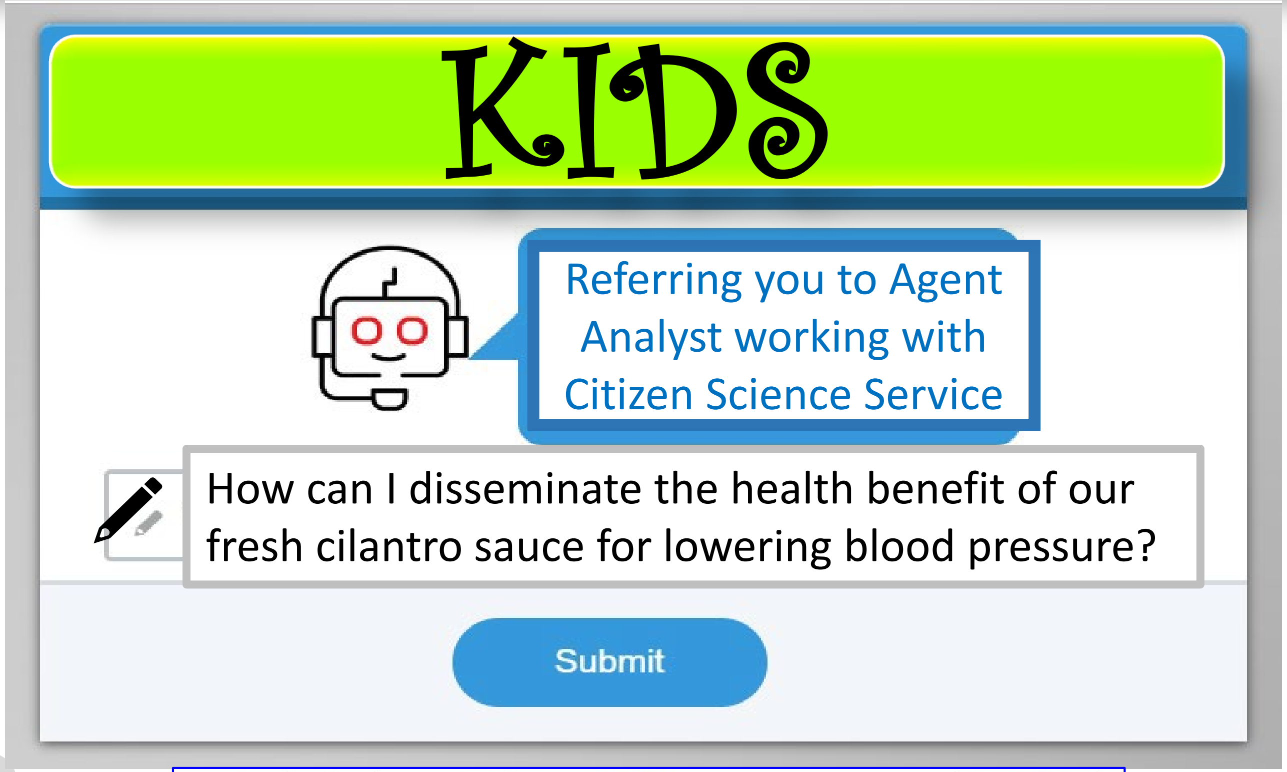


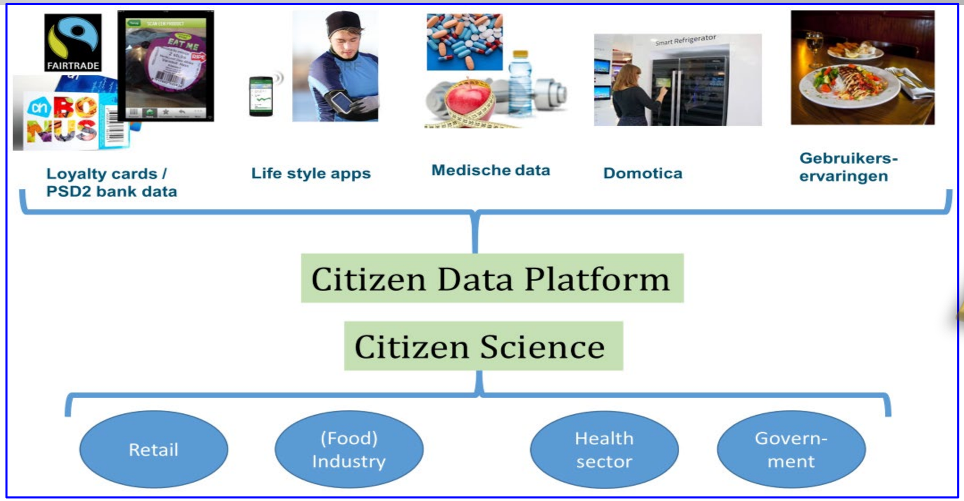
DATA → KNOWLEDGE

HOW KIDS IS DESIGNED TO ADDRESS COMPLEXITY. LET US TAKE ANOTHER LOOK.


K I D S
CITIZEN SCIENCE
CONTEXT
CONTEXT
ENGINES
KIDS
AGENT-BASED
ANALYSTS
AGENT-BASED SELECTION
AKBAR
Root Cause Analysis
Intelligent Recommendations
Self Tuning
IoT
Cyber
Physical
IoT
SENSEE
CROWD SOURCING
LOGIC
REASONING
AGENTS
AD HOC CONNECTION
TO EXTERNAL NETWORKS
PLATFORMS-ON-DEMAND
ALGORITHM
DATA
STORE
$

End-user perspective and questions from the field (agro-ecosystem) are complex

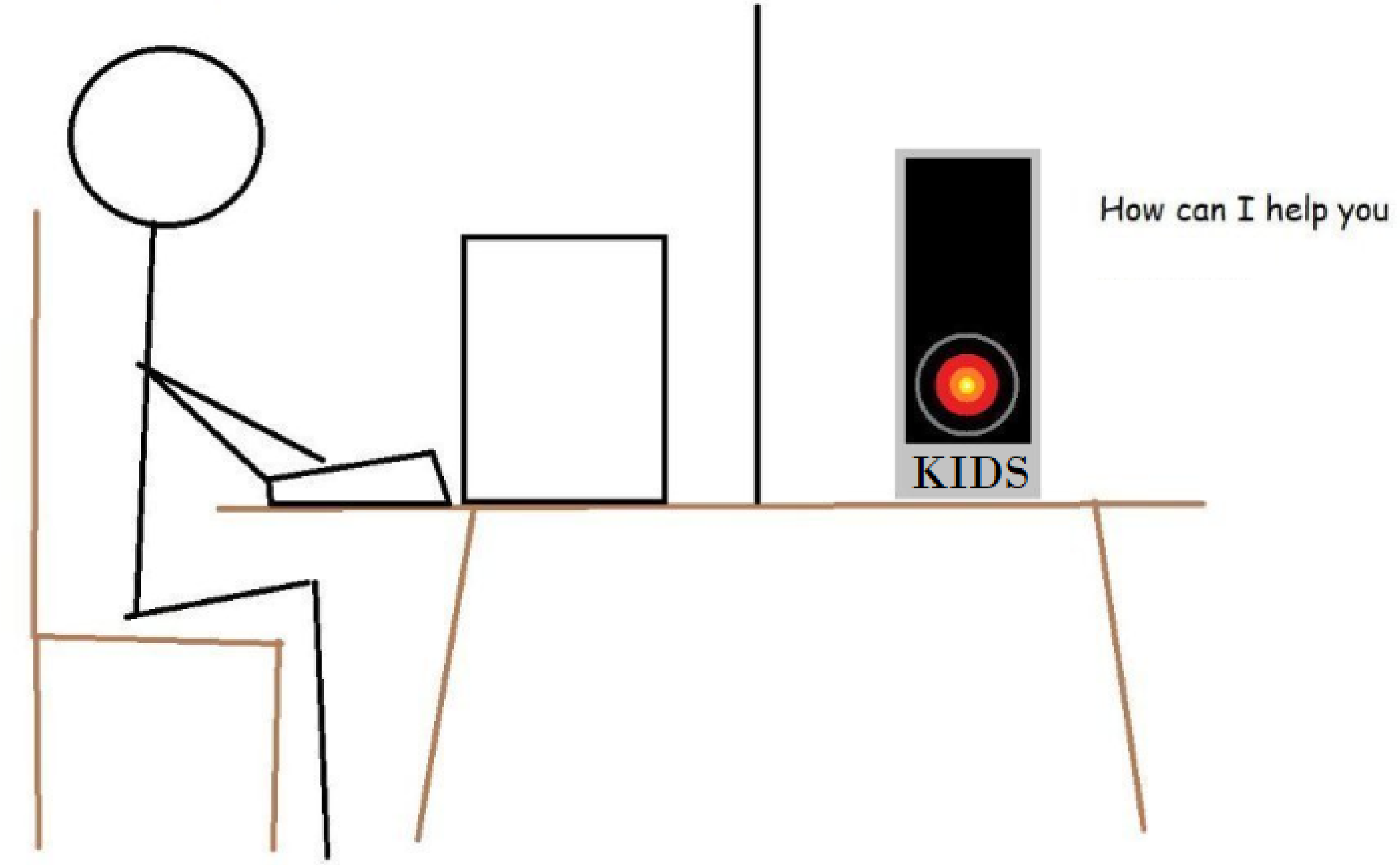




https://www.slideshare.net/knoesis/shreyansh-thesis-defense

KIDS

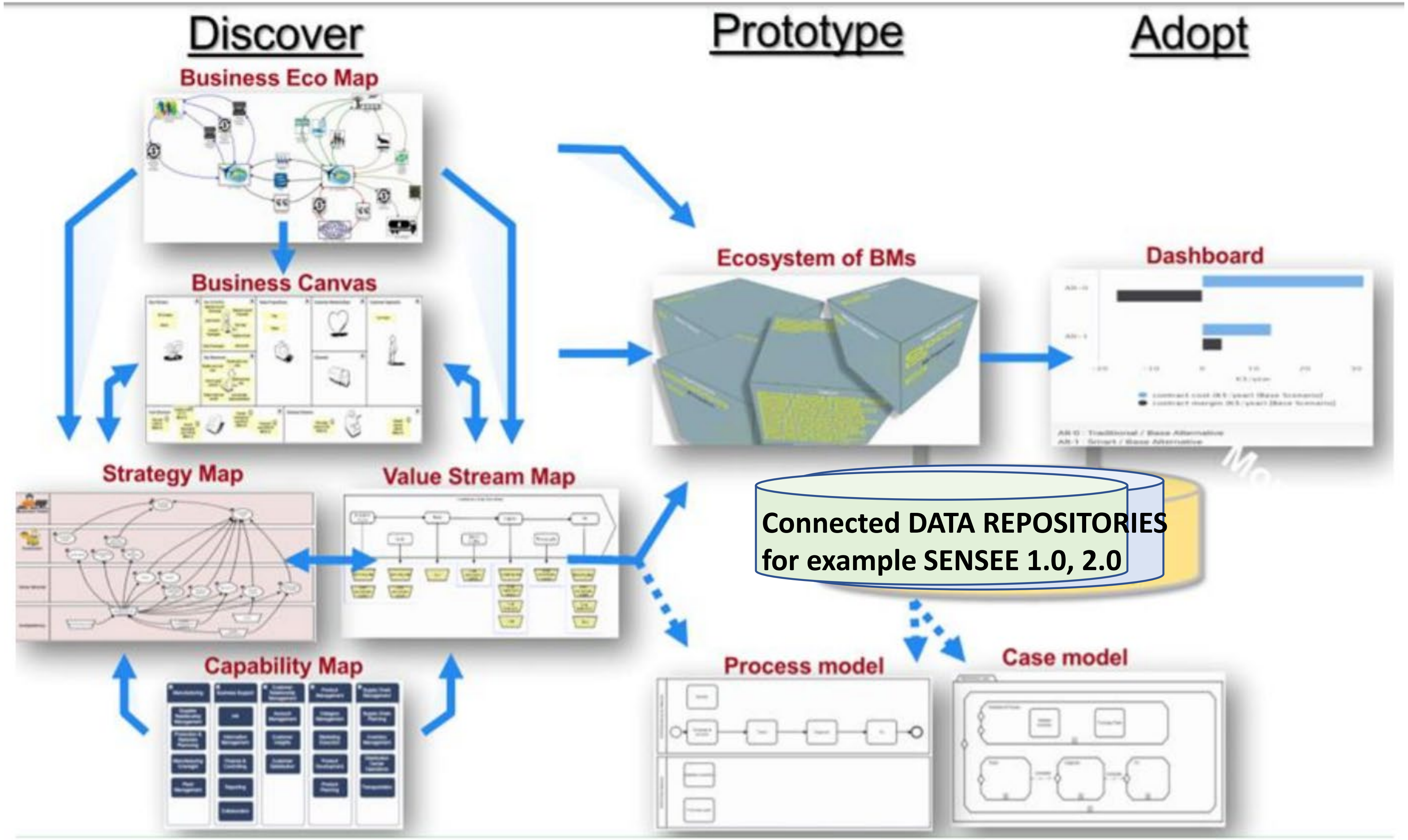
The use of Agents as virtual chat bots are common and may be modified for the agro-ecosystem.



Microsoft Virtual Agent and Dynamics 365 for Finance and Operations

Dynamic discovery & composability of graph networks for continuous value modeling relative to demand.

KIDS



www.vdmbee.com/wp-content/vmp-library/e-books/VDML_Based_CAV_Business_Modeling_V2.2.pdf

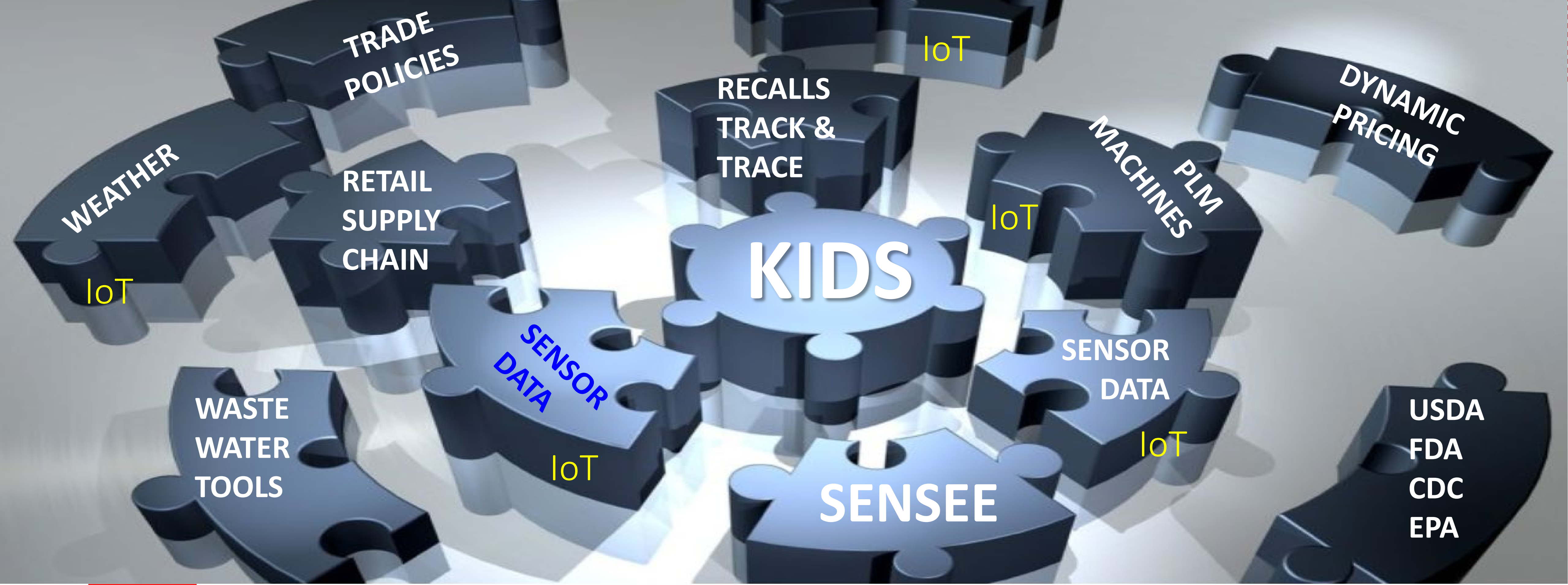
Continuous Business Model Planning - https://www.vdmbee.com/wp-content/vmp-library/e-books/VMBO_2019_paper_v1.pdf

End-user perspective and questions from the field (agro-ecosystem) are complex

PEAS, KIDS

makes sense?

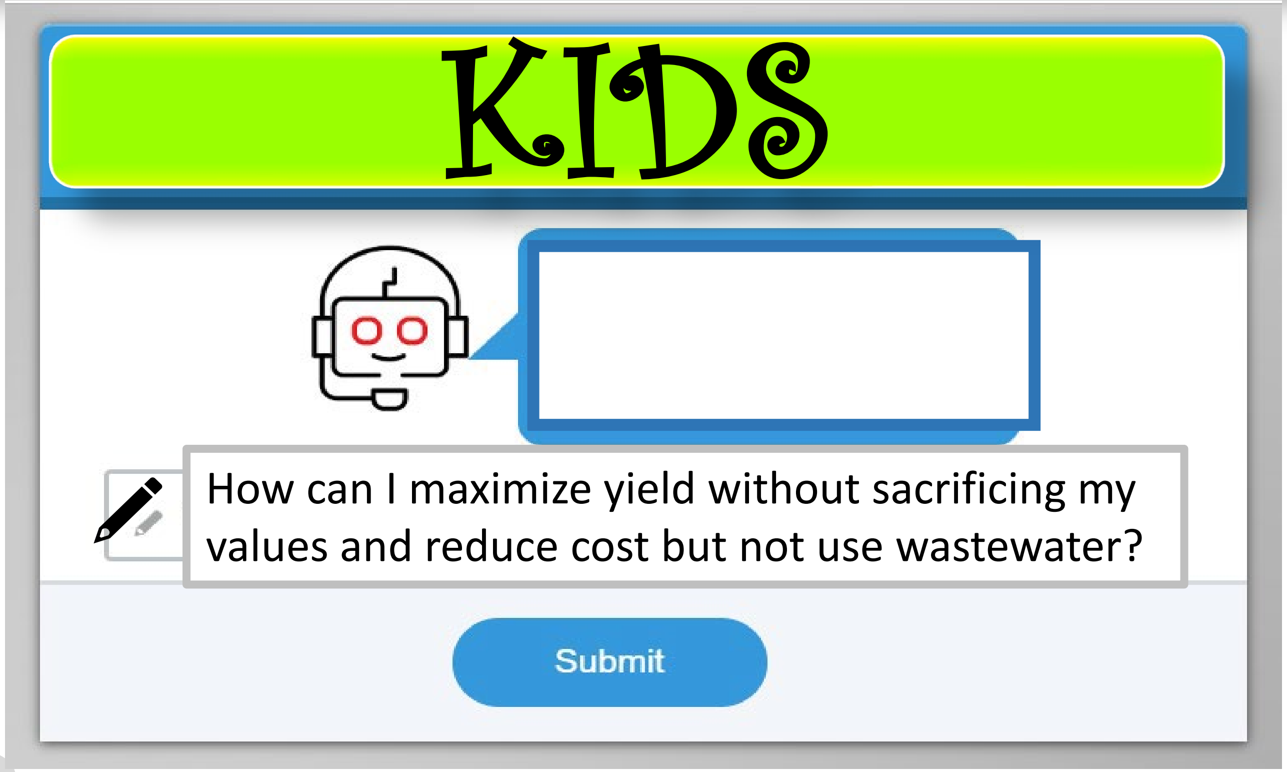
WHY SENSEE IS JUST A TINY STEP IN OUR JOURNEY TO KNOWLEDGE-INFORMED DECISIONS (KIDS)



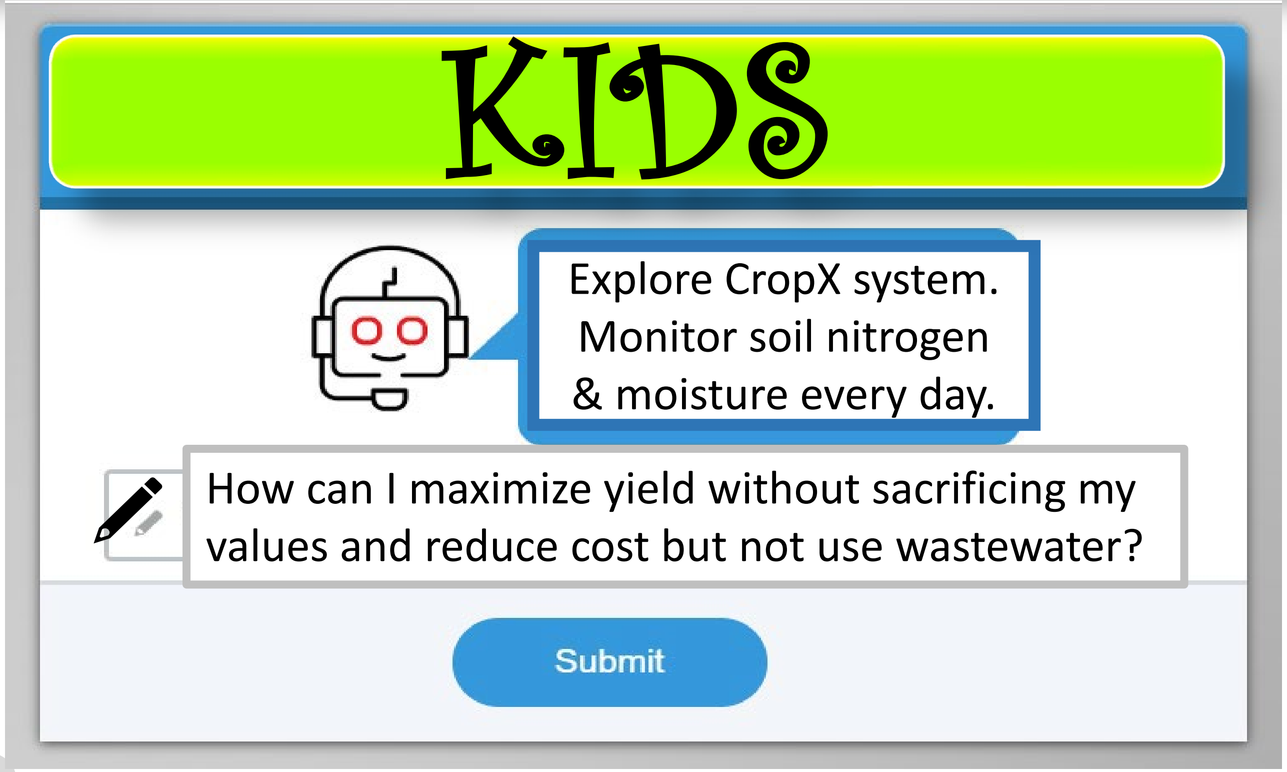
DATA → KNOWLEDGE

How can I maximize yield without sacrificing my values and reduce cost but not use wastewater?

Can KIDS answer end-user questions? We don't know but that is the expectation.



Can KIDS answer end-user questions? We don’t know but that is the expectation.



SENSEE

A journey of a thousand miles begins with a single step.

Developing SENSEE 1.0

Quote from Lao Tzu

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Spreadsheet

Sensor Properties

Library
PoC
DB

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Spreadsheet

Sensor Properties

	A	B	C	D	E	F	G	H	I	J	K	L
1	**MW [Da]**	**Category**	**Target**	**Recognition-transduction scheme**	**Platform**	**LOD [M]**	**Max range [M]**	**Selectivity (interferent species tested)**	**Response time [sec]**	**Durability**	**USECAT1**	**Link to paper(s)**
2	1	small molecule	H+	H+ ionophore (liquid)	glass capillary	1.00E-13	1.00E-04	excellent (K+, Na+, Ca2+, Mg2+)	2	High	hydroponics	https://www.ncbi.nlm.nih.gov/pubmed/26088926
3	1	small molecule	H+	anthocyanin/nanocellulose	paper filter	1.00E-15	1.00E-02	excellent	2	high	irrigation water	https://www.ncbi.nlm.nih.gov/pubmed/28884510
4	18	small molecule	Ammonium	NH4+ ionophore (liquid)	glass capillary	5.00E-09	1.00E-01	excellent	5	medium	wastewater	http://www.allelopathyjournal.org/archives/?Year=2016&Vol=37&Issue=2&Month=3
5	18	small molecule	Ammonium	NH4+ ionophore (solid)	LSG	2.80E-05	5.00E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
6	30	small molecule	N/O radicals	nanoplatinum/nanoceria	Pt electrode	1.00E-08	3.00E-06	medium	1	medium	ocean water	https://pubs.rsc.org/en/Content/ArticleLanding/2017/AY/C7AY01964E#!divAbstract
7	32	small molecule	DO	Pt porphyrin-nTiO2	fiber optic	1.00E-06	5.00E-06	excellent, temp sens	1	High	hydroponic media	https://www.sciencedirect.com/science/article/pii/S0925400514001117
8	32	small molecule	DO	Pt porphyrin	96 well	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	https://www.sciencedirect.com/science/article/pii/S016770121300331X
9	32	small molecule	DO	Pt porphyrin	glass vial	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	
10	34	small molecule	H2O2	fractal nPt	Pt electrode	5.00E-09	5.00E-05	excellent	1	high	ocean water	https://www.ncbi.nlm.nih.gov/pubmed/27121177
11	39	small molecule	K+	K+ ionophore (liquid)	glass capillary	1.00E-06	2.50E-01	excellent	2	low	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/24961073
12	41	small molecule	Ca2+	Ca2+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-01	excellent	1	low	Hoaglands media	https://onlinelibrary.wiley.com/doi/full/10.1002/jpln.201700319
13	58	small molecule	acetone	chemosensory proteins-nPt	Pt electrode	5.00E-06	1.00E-05	high	10	low	buffer	https://pubs.rsc.org/en/content/articlelanding/2018/an/c8an00065d#!divAbstract
14	62	small molecule	Nitrate	NO3- ionophore (liquid)	glass capillary	1.00E-06	2.00E-01	excellent	2	medium	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/18985610
15	62	small molecule	Nitrate	NO3- ionophore (solid)	LSG	2.00E-05	1.50E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
16	108	small molecule	Ag+	Ag+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-02	excellent	2	high	wound dressing	https://link.springer.com/article/10.1007/s11356-014-3058-6
17	111	small molecule	histamine	diamine oxidase-nCu	LSG	6.30E-05	1.00E-03	excellent	2	medium	fermented fish	https://www.mdpi.com/2079-6374/8/2/42
18	147	small molecule	Glutamate	CNT/nPt/GlOx	Pt electrode	1.00E-06	1.00E-03	excellent	2	low	INS1 tissue culture	https://www.sciencedirect.com/science/article/pii/S0165027010001196
19	147	small molecule	Glutamate	CNT/nPt/GlOx	Si biochip	1.00E-06	5.00E-01	excellent	2	low	INS1 tissue culture	https://pubs.rsc.org/en/content/articlelanding/2011/jm/c1jm11561h#!divAbstract
20	154	small molecule	catecholamines	nPt	LSG	5.00E-07	3.00E-03	excellent	2	high	ocean water	
21	154	small molecule	catecholamines	graphene anchored nCuO	LSG	3.00E-07	3.00E-03	high	2	medium	buffer	https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.8b02510
22	176	small molecule	indole acetic acid	fractal nPt	Pt/Ir microwire	1.00E-06	1.00E-03	high	1	high	root growth media	https://link.springer.com/article/10.1007/s00344-017-9688-4
23	181	small molecule	Glucose	nPt/GOx	graphene paper	8.00E-08	1.00E-03	excellent	2	medium	buffer	https://www.ncbi.nlm.nih.gov/pubmed/27209574
24	181	small molecule	Glucose	nPt/GOx	Pt/Ir microwire	1.00E-07	5.00E-06	excellent	1	medium	blood	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166557
25	181	small molecule	Glucose	nPt/nCe/GOx	Pt electrode	1.00E-07	3.00E-06	excellent	1	medium	buffer	https://www.sciencedirect.com/science/article/pii/S0956566314000992

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Spreadsheet

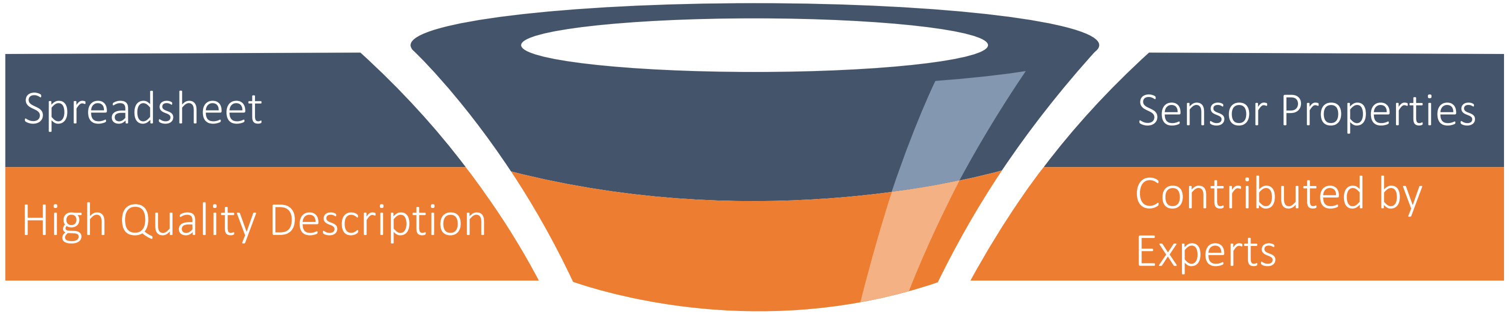
Sensor Properties

	A	B	C	D	E	F	G	H	I	J	K	L
1	**MW [Da]**	**Category**	**Target**	**Recognition-transduction scheme**	**Platform**	**LOD [M]**	**Max range [M]**	**Selectivity (interferent species tested)**	**Response time [sec]**	**Durability**	**USECAT1**	**Link to paper(s)**
2	1	small molecule	H+	H+ ionophore (liquid)	glass capillary	1.00E-13	1.00E-04	excellent (K+, Na+, Ca2+, Mg2+)	2	High	hydroponics	https://www.ncbi.nlm.nih.gov/pubmed/26088926
3	1	small molecule	H+	anthocyanin/nanocellulose	paper filter	1.00E-15	1.00E-02	excellent	2	high	irrigation water	https://www.ncbi.nlm.nih.gov/pubmed/28884510
4	18	small molecule	Ammonium	NH4+ ionophore (liquid)	glass capillary	5.00E-09	1.00E-01	excellent	5	medium	wastewater	http://www.allelopathyjournal.org/archives/?Year=2016&Vol=37&Issue=2&Month=3
5	18	small molecule	Ammonium	NH4+ ionophore (solid)	LSG	2.80E-05	5.00E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
6	30	small molecule	N/O radicals	nanoplatinum/nanoceria	Pt electrode	1.00E-08	3.00E-06	medium	1	medium	ocean water	https://pubs.rsc.org/en/Content/ArticleLanding/2017/AY/C7AY01964E#!divAbstract
7	32	small molecule	DO	Pt porphyrin-nTiO2	fiber optic	1.00E-06	5.00E-06	excellent, temp sens	1	High	hydroponic media	https://www.sciencedirect.com/science/article/pii/S0925400514001117
8	32	small molecule	DO	Pt porphyrin	96 well	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	https://www.sciencedirect.com/science/article/pii/S016770121300331X
9	32	small molecule	DO	Pt porphyrin	glass vial	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	
10	34	small molecule	H2O2	fractal nPt	Pt electrode	5.00E-09	5.00E-05	excellent	1	high	ocean water	https://www.ncbi.nlm.nih.gov/pubmed/27121177
11	39	small molecule	K+	K+ ionophore (liquid)	glass capillary	1.00E-06	2.50E-01	excellent	2	low	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/24961073
12	41	small molecule	Ca2+	Ca2+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-01	excellent	1	low	Hoaglands media	https://onlinelibrary.wiley.com/doi/full/10.1002/jpln.201700319
13	58	small molecule	acetone	chemosensory proteins-nPt	Pt electrode	5.00E-06	1.00E-05	high	10	low	buffer	https://pubs.rsc.org/en/content/articlelanding/2018/an/c8an00065d#!divAbstract
14	62	small molecule	Nitrate	NO3- ionophore (liquid)	glass capillary	1.00E-06	2.00E-01	excellent	2	medium	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/18985610
15	62	small molecule	Nitrate	NO3- ionophore (solid)	LSG	2.00E-05	1.50E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
16	108	small molecule	Ag+	Ag+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-02	excellent	2	high	wound dressing	https://link.springer.com/article/10.1007/s11356-014-3058-6
17	111	small molecule	histamine	diamine oxidase-nCu	LSG	6.30E-05	1.00E-03	excellent	2	medium	fermented fish	https://www.mdpi.com/2079-6374/8/2/42
18	147	small molecule	Glutamate	CNT/nPt/GlOx	Pt electrode	1.00E-06	1.00E-03	excellent	2	low	INS1 tissue culture	https://www.sciencedirect.com/science/article/pii/S0165027010001196
19	147	small molecule	Glutamate	CNT/nPt/GlOx	Si biochip	1.00E-06	5.00E-01	excellent	2	low	INS1 tissue culture	https://pubs.rsc.org/en/content/articlelanding/2011/jm/c1jm11561h#!divAbstract
20	154	small molecule	catecholamines	nPt	LSG	5.00E-07	3.00E-03	excellent	2	high	ocean water	
21	154	small molecule	catecholamines	graphene anchored nCuO	LSG	3.00E-07	3.00E-03	high	2	medium	buffer	https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.8b02510
22	176	small molecule	indole acetic acid	fractal nPt	Pt/Ir microwire	1.00E-06	1.00E-03	high	1	high	root growth media	https://link.springer.com/article/10.1007/s00344-017-9688-4
23	181	small molecule	Glucose	nPt/GOx	graphene paper	8.00E-08	1.00E-03	excellent	2	medium	buffer	https://www.ncbi.nlm.nih.gov/pubmed/27209574
24	181	small molecule	Glucose	nPt/GOx	Pt/Ir microwire	1.00E-07	5.00E-06	excellent	1	medium	blood	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166557
25	181	small molecule	Glucose	nPt/nCe/GOx	Pt electrode	1.00E-07	3.00E-06	excellent	1	medium	buffer	https://www.sciencedirect.com/science/article/pii/S0956566314000992

Library
PoC
DB

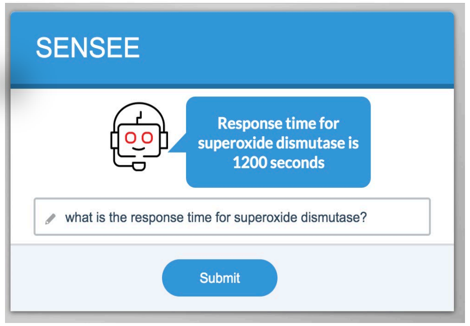
Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Sensor R&D Community



Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

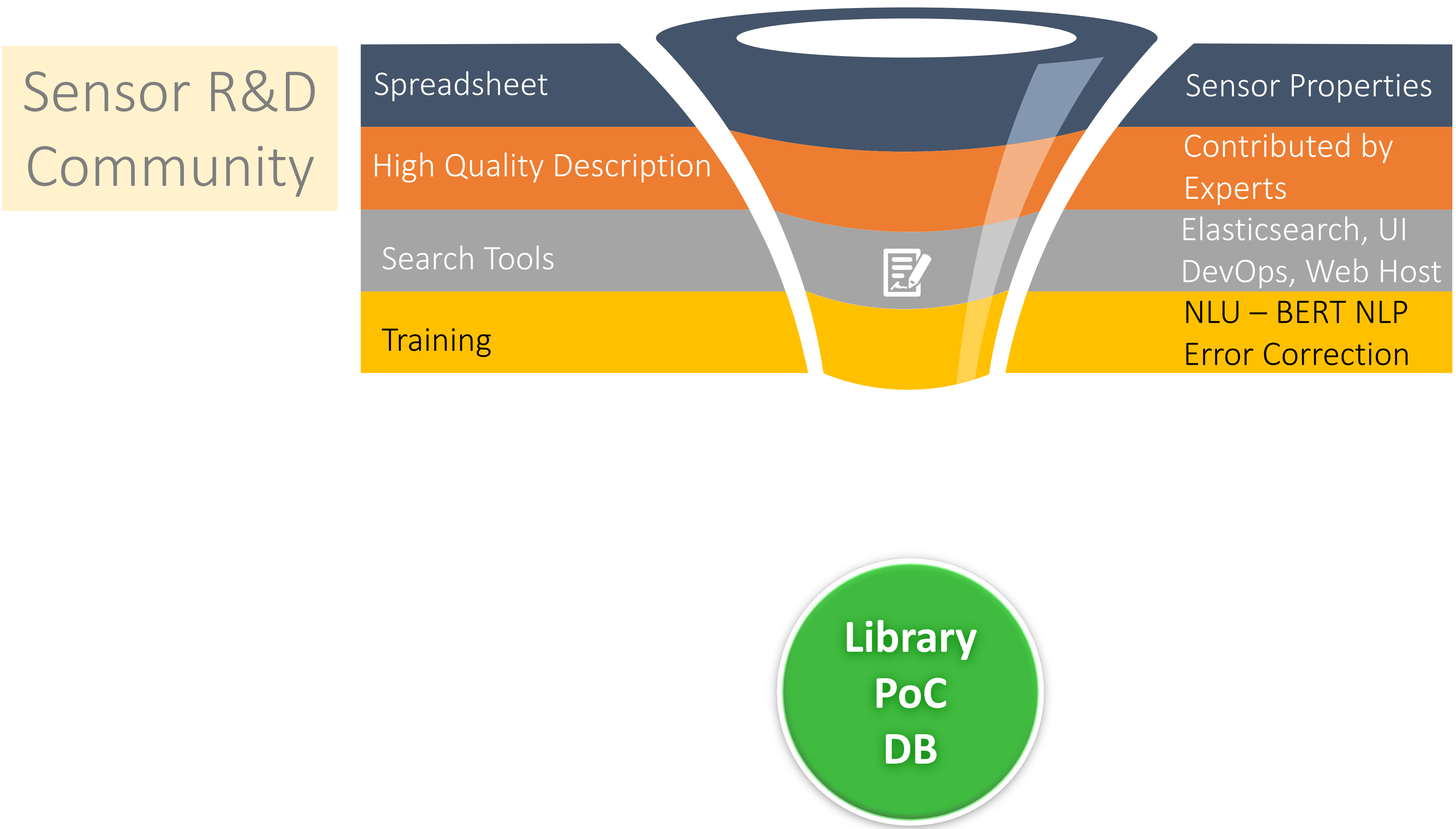
Sensor R&D Community



http://139.162.7.63/SENSEE/

Library
PoC
DB

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

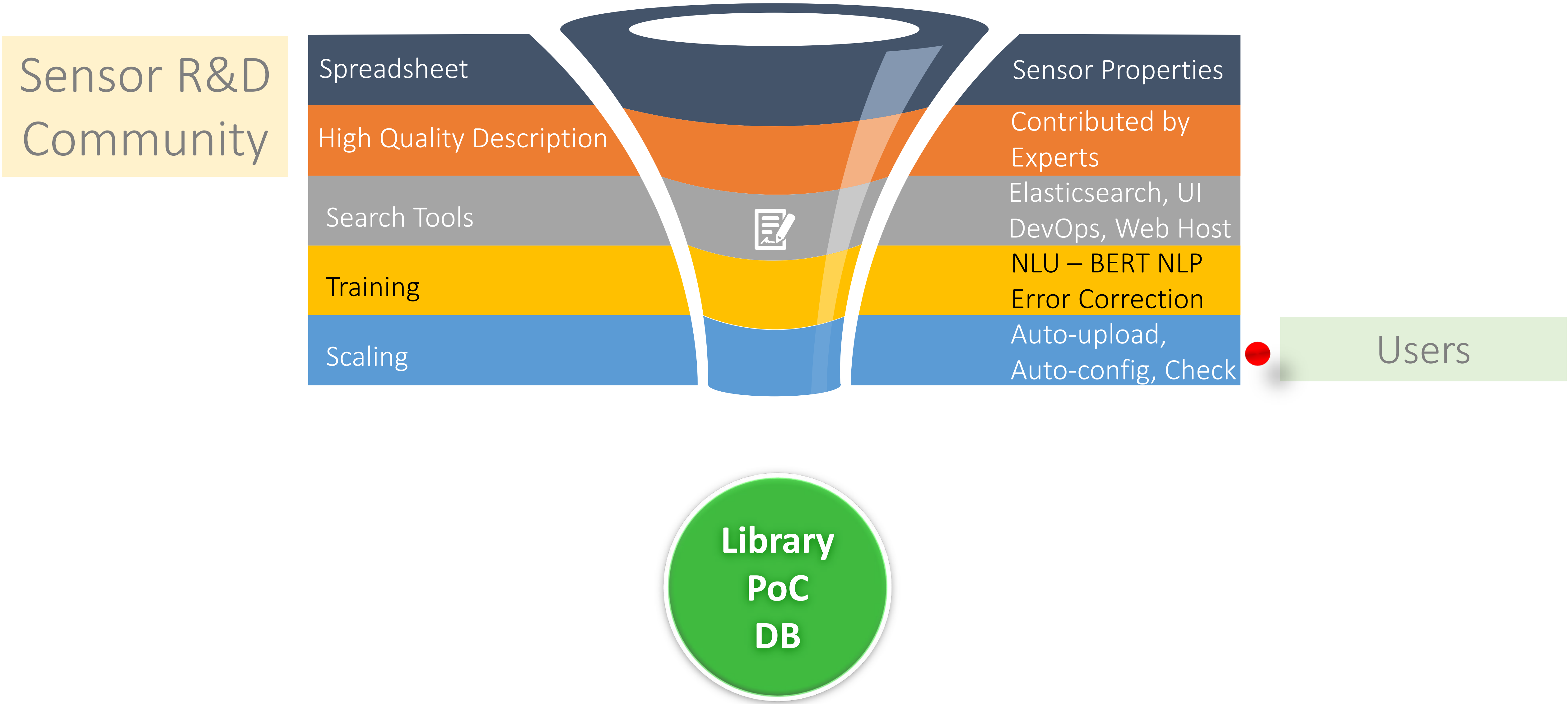


Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Sensor R&D Community

Library
PoC
DB

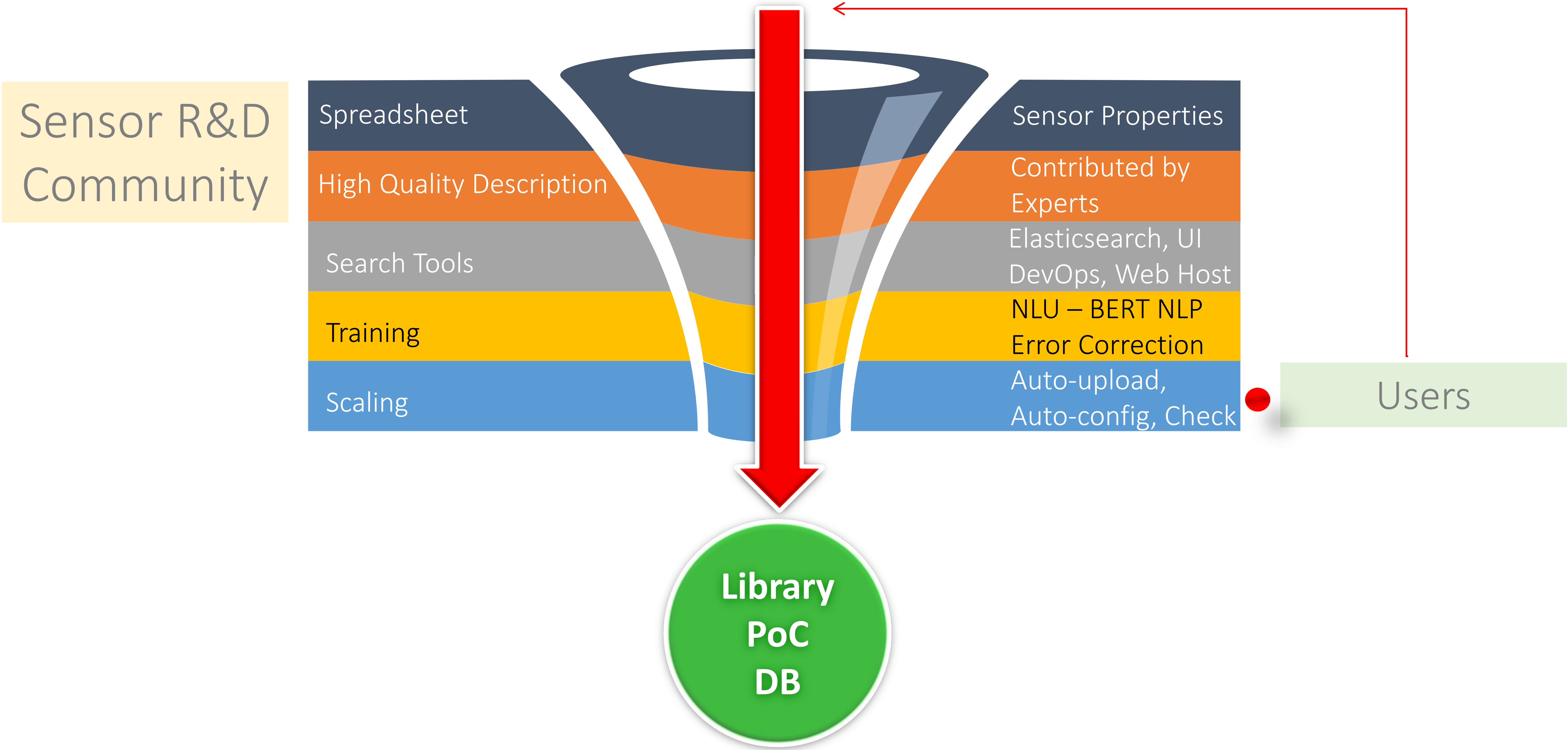
Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)



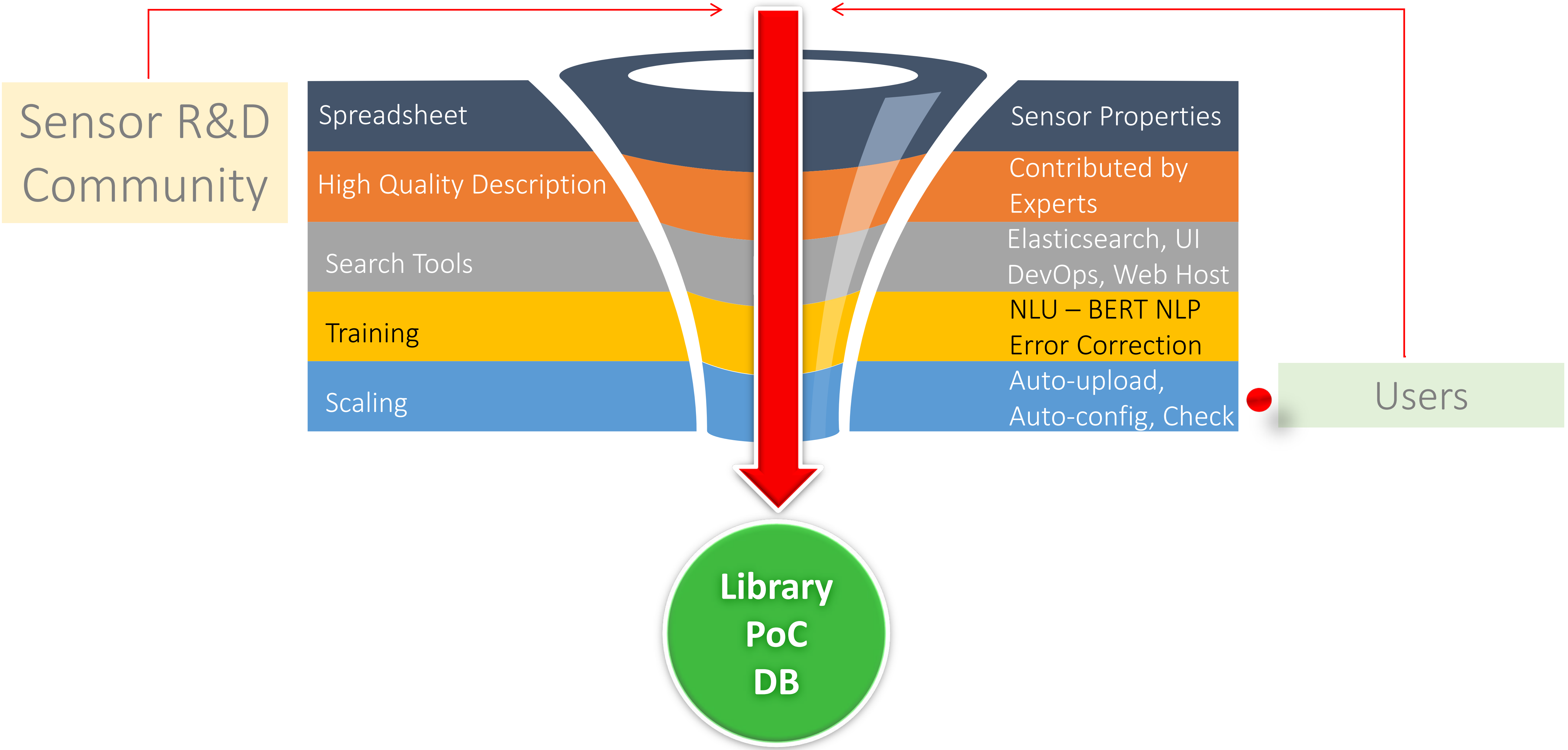
SENSEE 1.0 PROOF OF CONCEPT – TEST QUESTIONS

155) What molecules can be detected in breast milk using biosensors?
156) What is the difference in sensitivity between glucose biosensors based on graphene or platinum foil?
157) What is the most sensitive biosensor based on carbon nanotubes?
158) How many biosensors have been proposed for glucose determination?
159) Anthocyanin is used as a target for which biosensor?
160) Which biosensors can be used for hydroponic medium?
161) In which samples, glutamate and/or glutamine was determined using biosensors?
162) Which biosensors were proposed for catecholamine determination?
163) What is the lowest limit of detection for graphene-based biosensors?
164) What is the maximal range for nitrate biosensors?
165) What platforms can be used for ammonium detection and mercury detection?
166) Most durable recognition-transduction scheme for interferon gamma biosensors?
167) Best limit of detection achieved with phosphotriesterase-based biosensors?
168) How many biosensors were described for ATP determination?
169) What platforms were proposed for ATP-sensitive biosensors?
170) What is the average LOD of K+ sensors?
171) Which platform could be used for selective glutamate analysis?
172) What is largest analyte/molecule for which there is a sensor in the database?
173) Is there any cost associated with any type of sensor?
174) How many labs are making sensors to detect lead in water?
175) Are there sensors to detect air-borne viruses in the air?

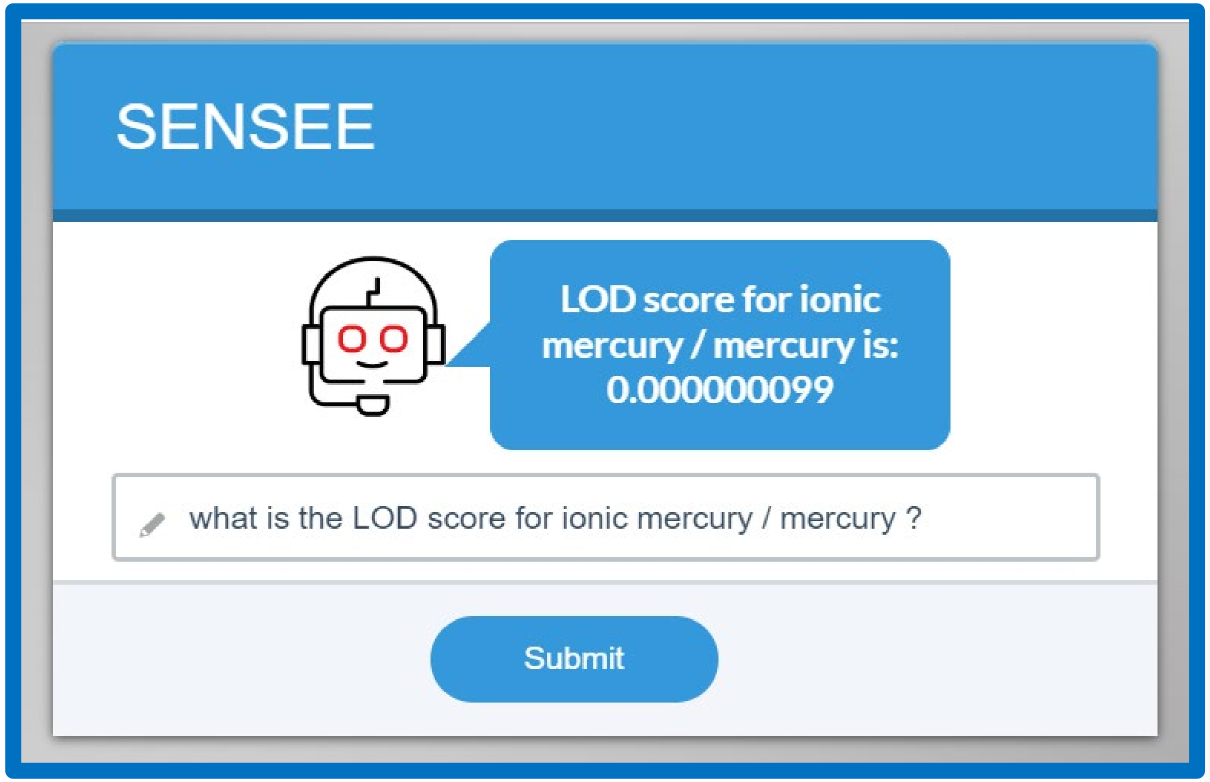
Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

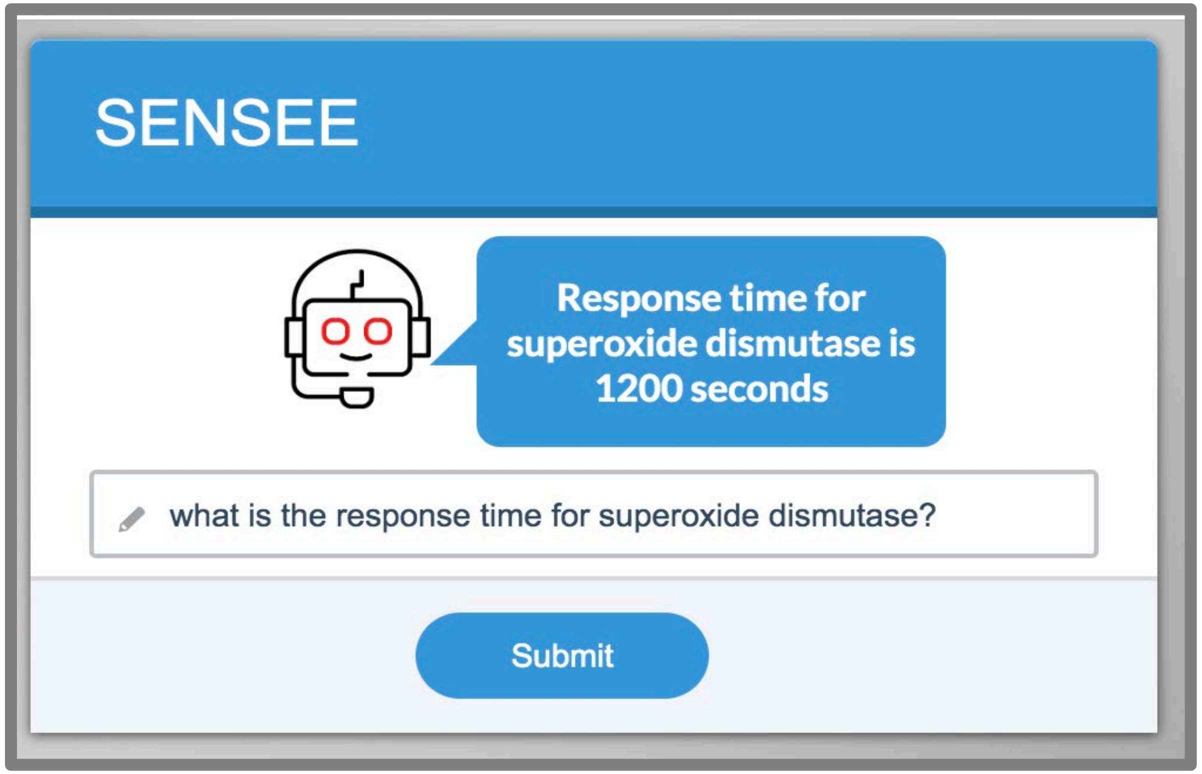


Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)



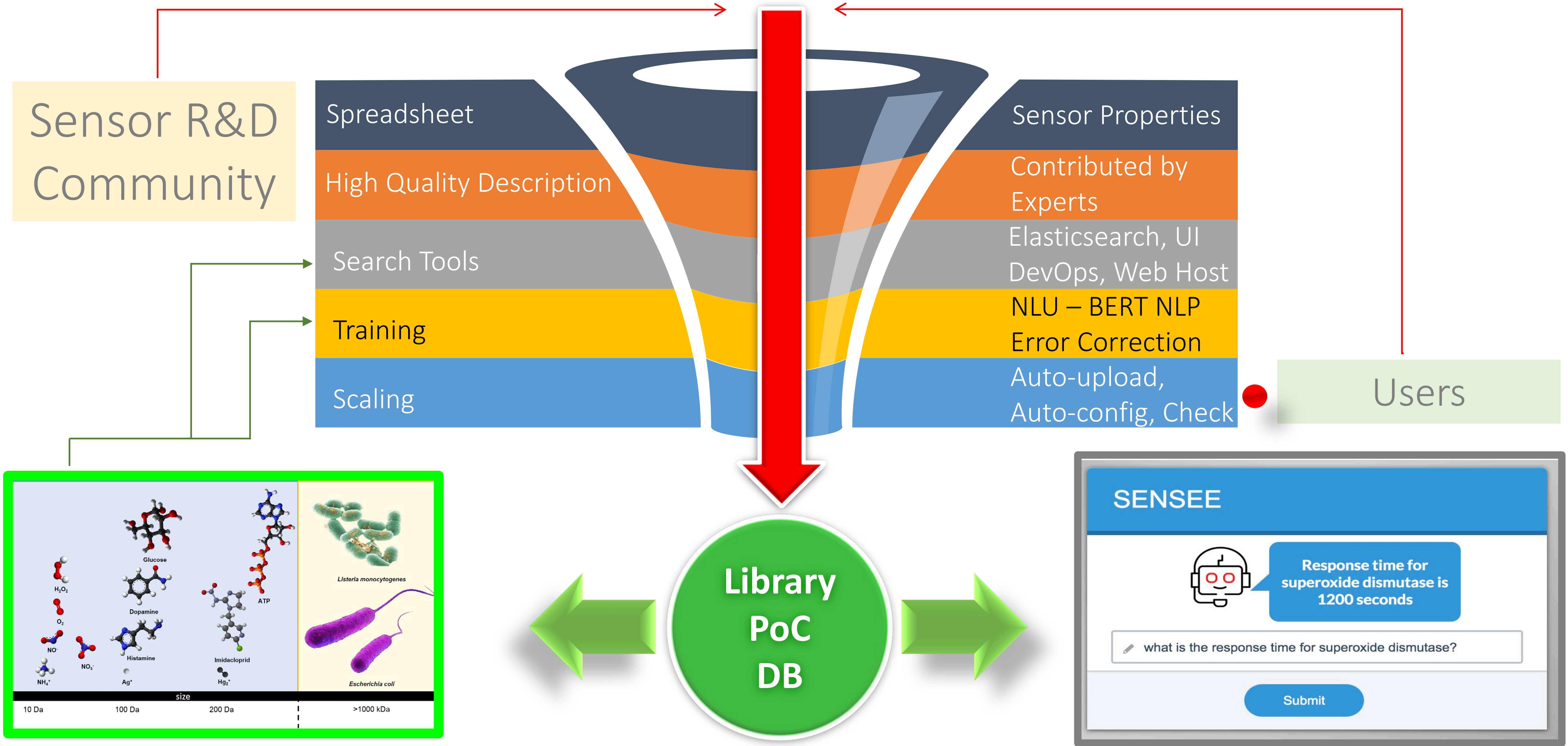
SENSEE 1.0 PROOF OF CONCEPT – DIALOG BOX APP





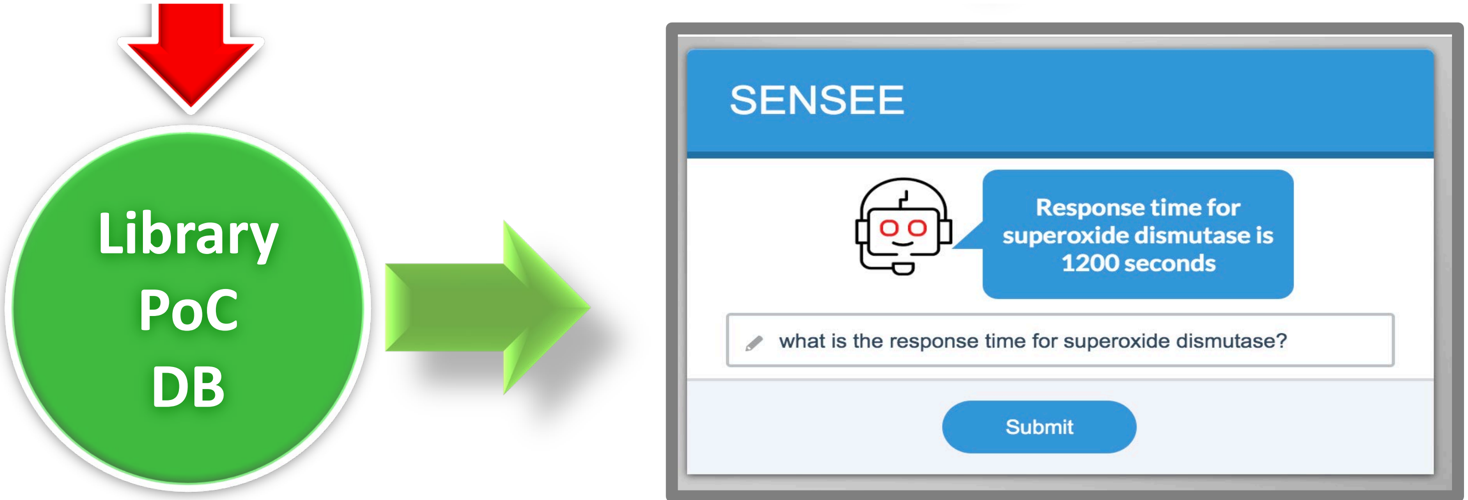
SENSEE PoC UI ● http://146.185.133.187/SENSEE1/ ● http://139.162.7.63/SENSEE/

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)



Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

This is where we are …

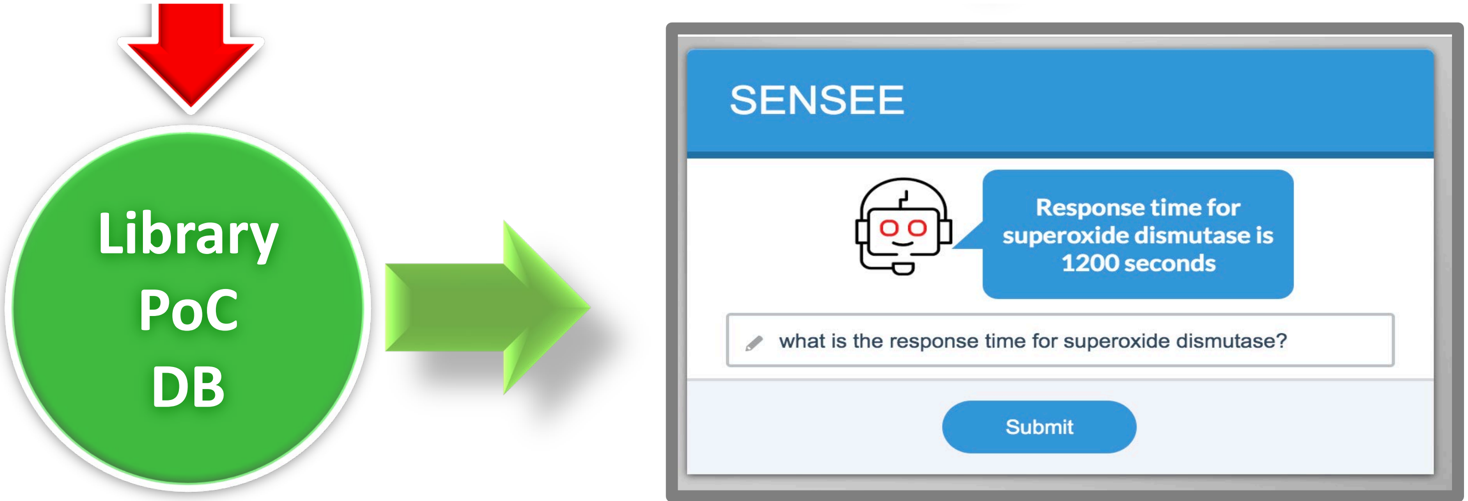


Development of SENSEE 1.0 – This is where we are

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

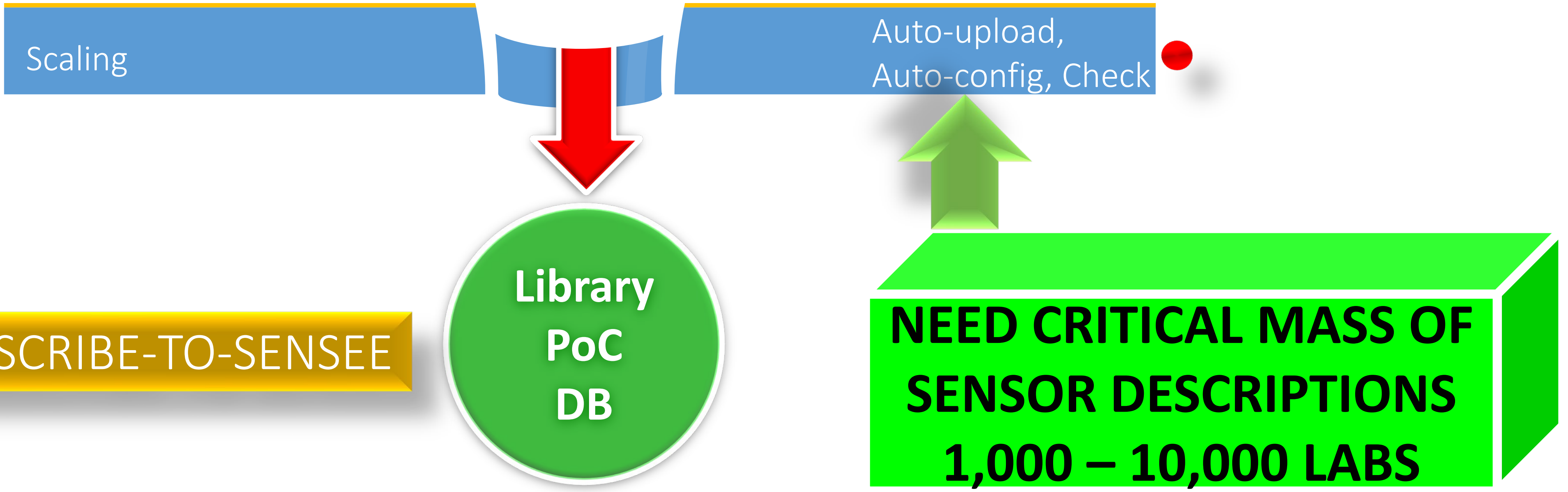
To move forward …

http://bit.ly/SUBSCRIBE-TO-SENSEE



Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

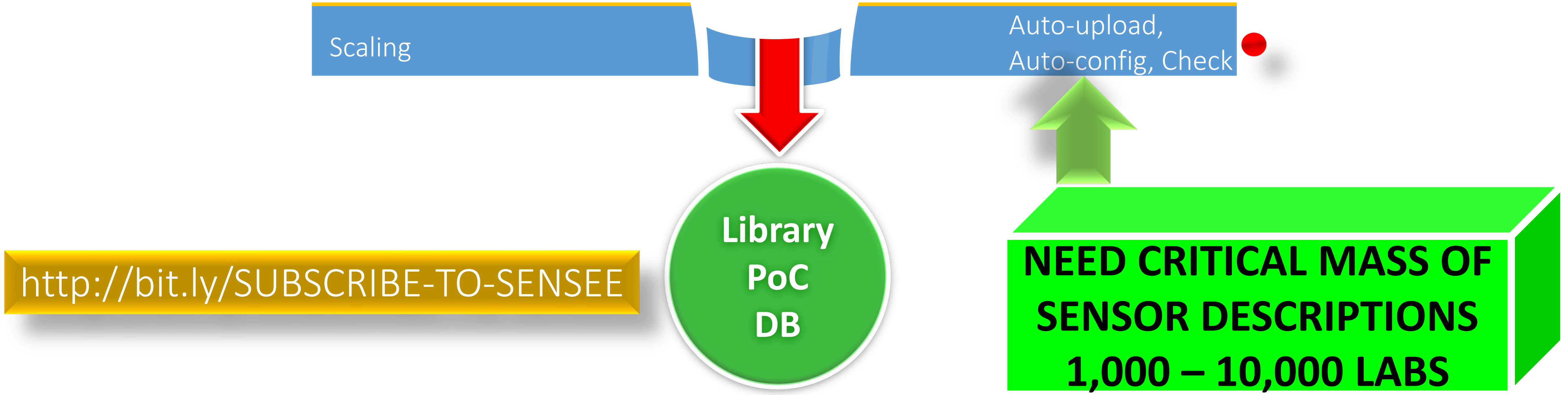
Research Community we need your help



http://bit.ly/SUBSCRIBE-TO-SENSEE

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

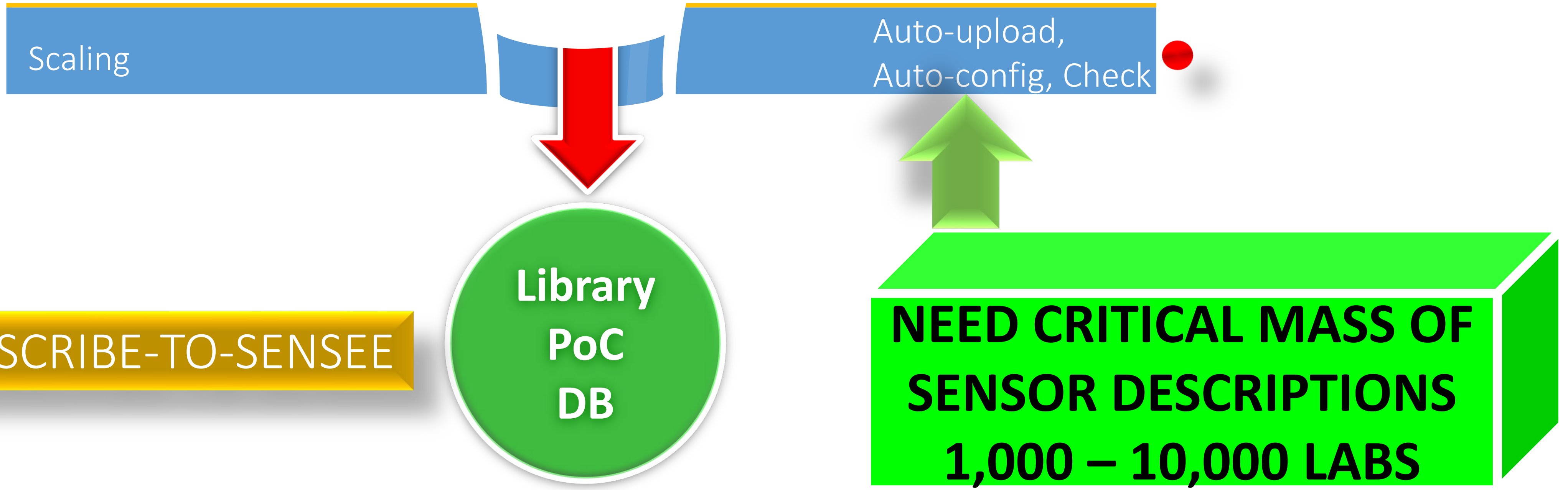
It is useless without sensor descriptions



Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

NO sensor data until 2.0

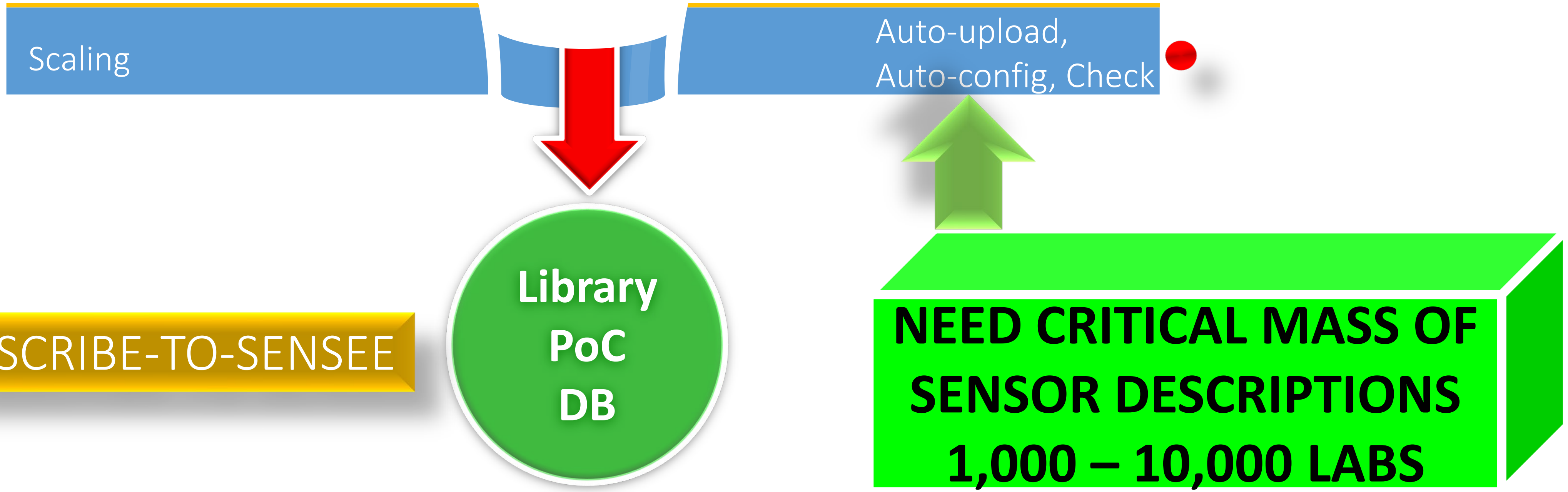
Your sensor descriptions



http://bit.ly/SUBSCRIBE-TO-SENSEE

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

XL auto-UPLOAD tool for your sensor descriptions



http://bit.ly/SUBSCRIBE-TO-SENSEE

Your Spreadsheet

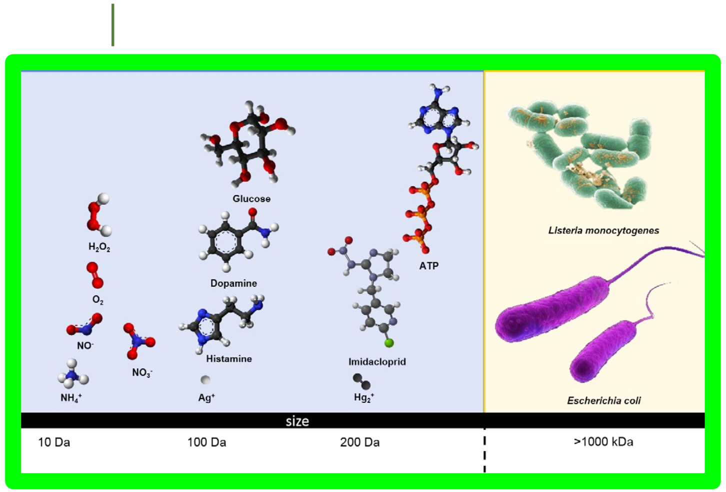
Sensor Properties

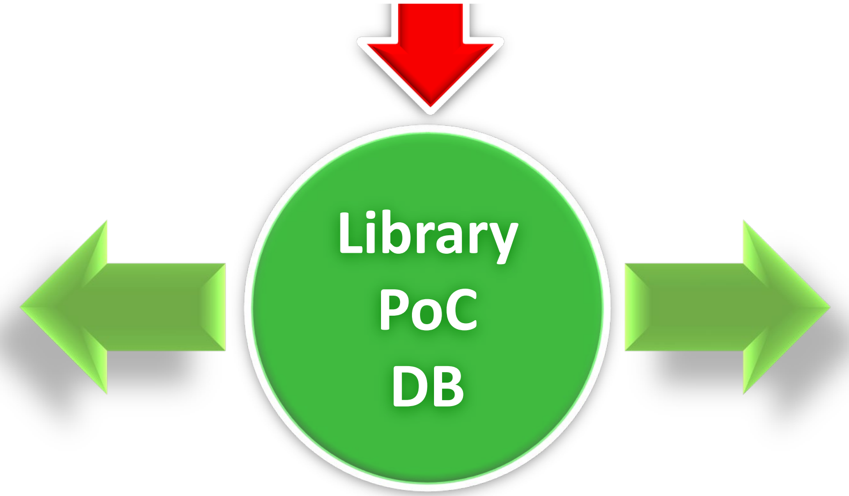
1	MW [Da]	Category	Target	Recognition-transduction scheme	Platform	LOD [M]	Max range [M]	Selectivity (interferent species tested)	Response time [sec]	Durability	USECAT1	Link to paper(s)
2	1	small molecule	H+	H+ ionophore (liquid)	glass capillary	1.00E-13	1.00E-04	excellent (K+, Na+, Ca2+, Mg2+)	2	High	hydroponics	https://www.ncbi.nlm.nih.gov/pubmed/26088926
3	1	small molecule	H+	anthocyanin/nanocellulose	paper filter	1.00E-15	1.00E-02	excellent	2	high	irrigation water	https://www.ncbi.nlm.nih.gov/pubmed/28884510
4	18	small molecule	Ammonium	NH4+ ionophore (liquid)	glass capillary	5.00E-09	1.00E-01	excellent	5	medium	wastewater	http://www.allelopathyjournal.org/archives/?Year=2016&Vol=37&Issue=2&Month=3
5	18	small molecule	Ammonium	NH4+ ionophore (solid)	LSG	2.80E-05	5.00E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
6	30	small molecule	N/O radicals	nanoplatinum/nanoceria	Pt electrode	1.00E-08	3.00E-06	medium	1	medium	ocean water	https://pubs.rsc.org/en/Content/ArticleLanding/2017/AY/C7AY01964E#!divAbstract
7	32	small molecule	DO	Pt porphyrin-nTiO2	fiber optic	1.00E-06	5.00E-06	excellent, temp sens	1	High	hydroponic media	https://www.sciencedirect.com/science/article/pii/S0925400514001117
8	32	small molecule	DO	Pt porphyrin	96 well	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	https://www.sciencedirect.com/science/article/pii/S016770121300331X
9	32	small molecule	DO	Pt porphyrin	glass vial	1.00E-06	5.00E-06	excellent, temp sens	45	low	hydroponic media	
10	34	small molecule	H2O2	fractal nPt	Pt electrode	5.00E-09	5.00E-05	excellent	1	high	ocean water	https://www.ncbi.nlm.nih.gov/pubmed/27121177
11	39	small molecule	K+	K+ ionophore (liquid)	glass capillary	1.00E-06	2.50E-01	excellent	2	low	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/24961073
12	41	small molecule	Ca2+	Ca2+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-01	excellent	1	low	Hoaglands media	https://onlinelibrary.wiley.com/doi/full/10.1002/jpln.201700319
13	58	small molecule	acetone	chemosensory proteins-nPt	Pt electrode	5.00E-06	1.00E-05	high	10	low	buffer	https://pubs.rsc.org/en/content/articlelanding/2018/an/c8an00065d#!divAbstract
14	62	small molecule	Nitrate	NO3- ionophore (liquid)	glass capillary	1.00E-06	2.00E-01	excellent	2	medium	wastewater	https://www.ncbi.nlm.nih.gov/pubmed/18985610
15	62	small molecule	Nitrate	NO3- ionophore (solid)	LSG	2.00E-05	1.50E-01	excellent	2	medium	wastewater	https://pubs.acs.org/doi/10.1021/acsami.8b10991
16	108	small molecule	Ag+	Ag+ ionophore (liquid)	glass capillary	1.00E-06	5.00E-02	excellent	2	high	wound dressing	https://link.springer.com/article/10.1007/s11356-014-3058-6
17	111	small molecule	histamine	diamine oxidase-nCu	LSG	6.30E-05	1.00E-03	excellent	2	medium	fermented fish	https://www.mdpi.com/2079-6374/8/2/42
18	147	small molecule	Glutamate	CNT/nPt/GlOx	Pt electrode	1.00E-06	1.00E-03	excellent	2	low	INS1 tissue culture	https://www.sciencedirect.com/science/article/pii/S0165027010001196
19	147	small molecule	Glutamate	CNT/nPt/GlOx	Si biochip	1.00E-06	5.00E-01	excellent	2	low	INS1 tissue culture	https://pubs.rsc.org/en/content/articlelanding/2011/jm/c1jm11561h#!divAbstract
20	154	small molecule	catecholamines	nPt	LSG	5.00E-07	3.00E-03	excellent	2	high	ocean water	
21	154	small molecule	catecholamines	graphene anchored nCuO	LSG	3.00E-07	3.00E-03	high	2	medium	buffer	https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.8b02510
22	176	small molecule	indole acetic acid	fractal nPt	Pt/Ir microwire	1.00E-06	1.00E-03	high	1	high	root growth media	https://link.springer.com/article/10.1007/s00344-017-9688-4
23	181	small molecule	Glucose	nPt/GOx	graphene paper	8.00E-08	1.00E-03	excellent	2	medium	buffer	https://www.ncbi.nlm.nih.gov/pubmed/27209574
24	181	small molecule	Glucose	nPt/GOx	Pt/Ir microwire	1.00E-07	5.00E-06	excellent	1	medium	blood	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166557
25	181	small molecule	Glucose	nPt/nCe/GOx	Pt electrode	1.00E-07	3.00E-06	excellent	1	medium	buffer	https://www.sciencedirect.com/science/article/pii/S0956566314000992

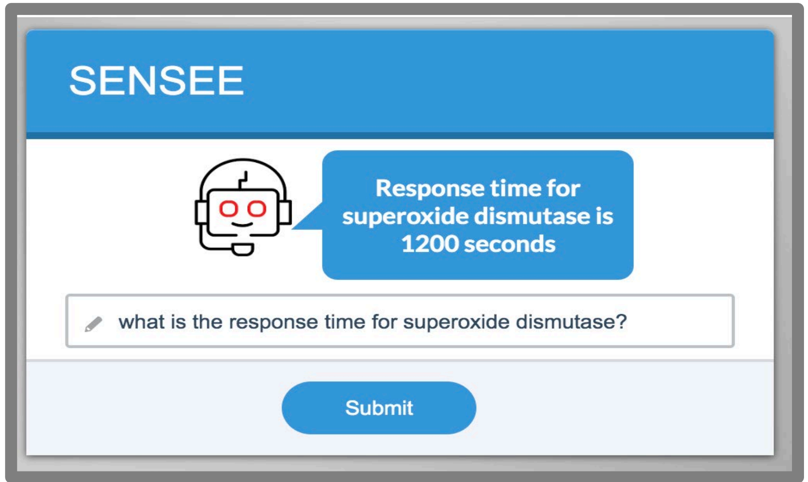
http://bit.ly/SUBSCRIBE-TO-SENSEE

Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

We aim to improve visualization tool



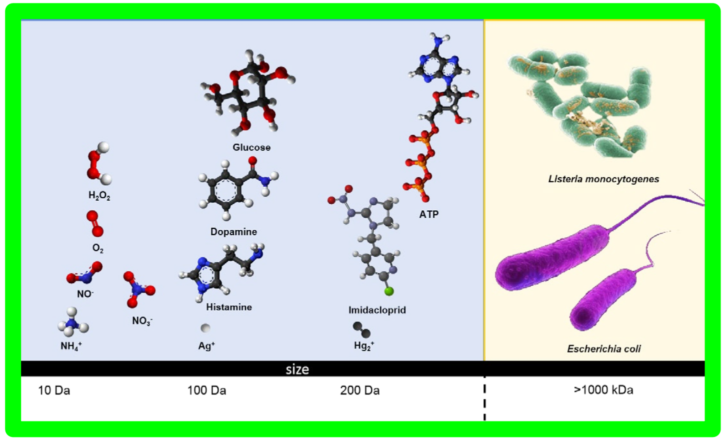


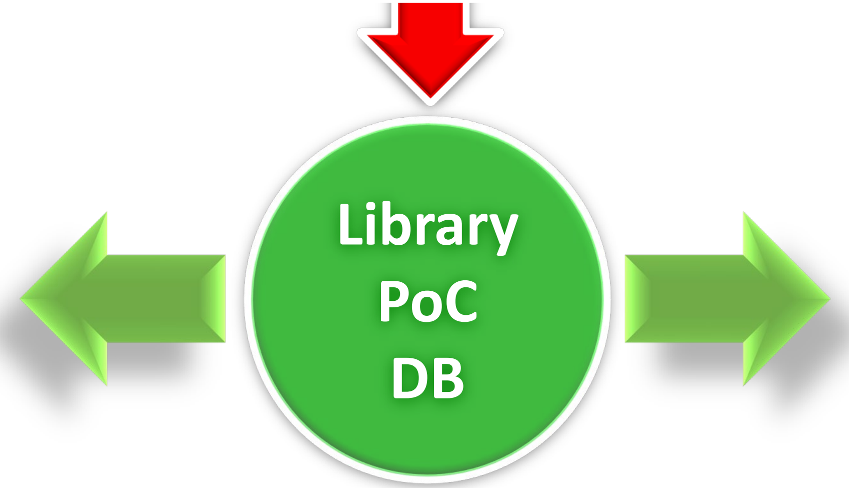


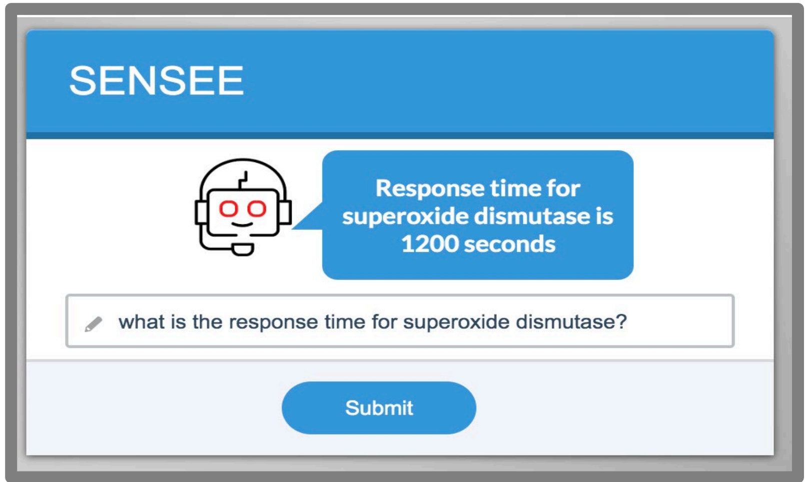
Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

What may follow

- ◆ Deploy SENSEE tool. Go live!
- Automate Feature Engineering
- Ingest Sensor-specific Data
- Use cases for DIDA'S PoC
- Knowledge Graph Algorithms
- Semantic Data Catalogs
- User directed search



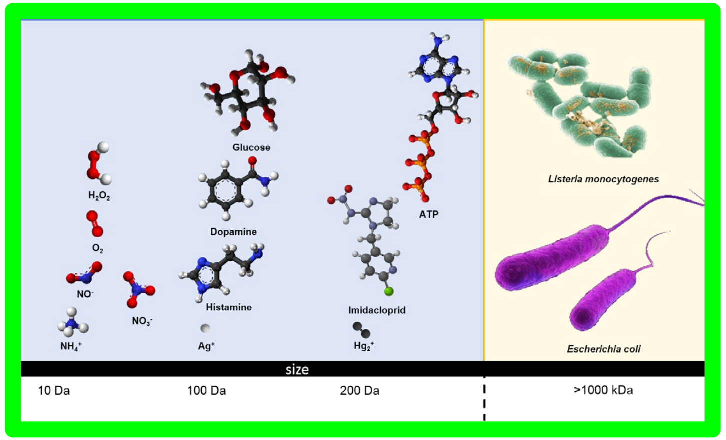


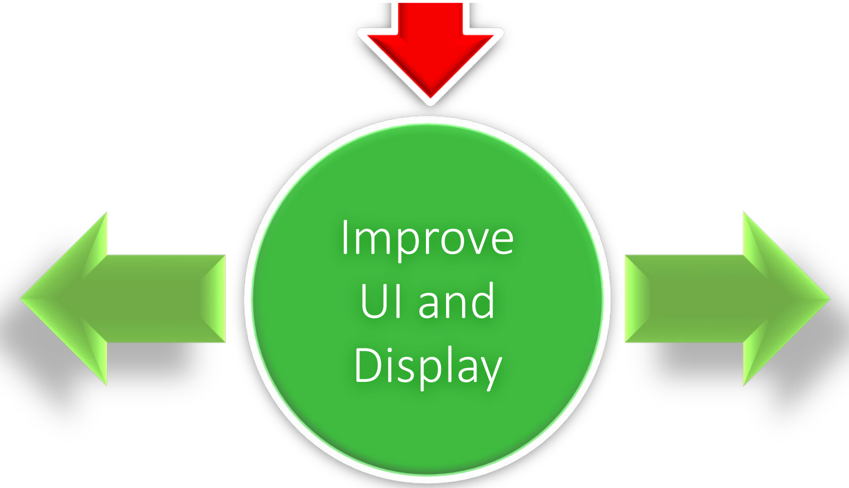


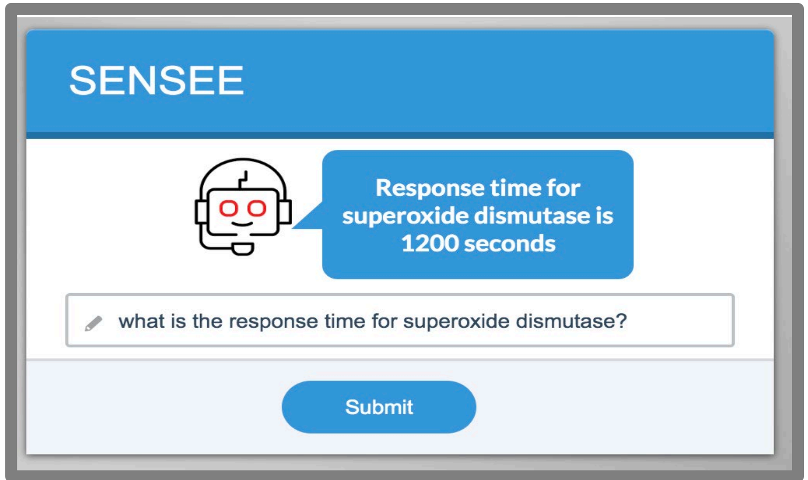
Development of SENSEE 1.0 (**SEN**sor **SE**arch **E**ngine)

Who will use SENSEE 1.0 tool? We anticipate that critical mass of sensor descriptions (categories, attributes) will improve the value of SENSEE 1.0 as a search tool for curated information. Users may be experts in academic and industrial labs. The task of sourcing, uploading, maintaining sensor descriptions may become cost-prohibitive unless a cooperative support structure is implemented to distribute and share the cost of professional services for SENSEE.

- Deploy SENSEE 1.0. Go live!







THE LONGEST JOURNEY

SENSEE

↓

DIDA'S

↓

KIDS

SENSEE 2.0

THE LONGEST JOURNEY

SENSEE

↓

DIDA'S

↓

KIDS

SENSEE 2.0

- Ingest Sensor-specific Data

THE LONGEST JOURNEY

SENSEE

↓

DIDA'S

↓

KIDS

SENSEE 2.0 INGESTS SENSOR-SPECIFIC DATA BASED ON USE CASES

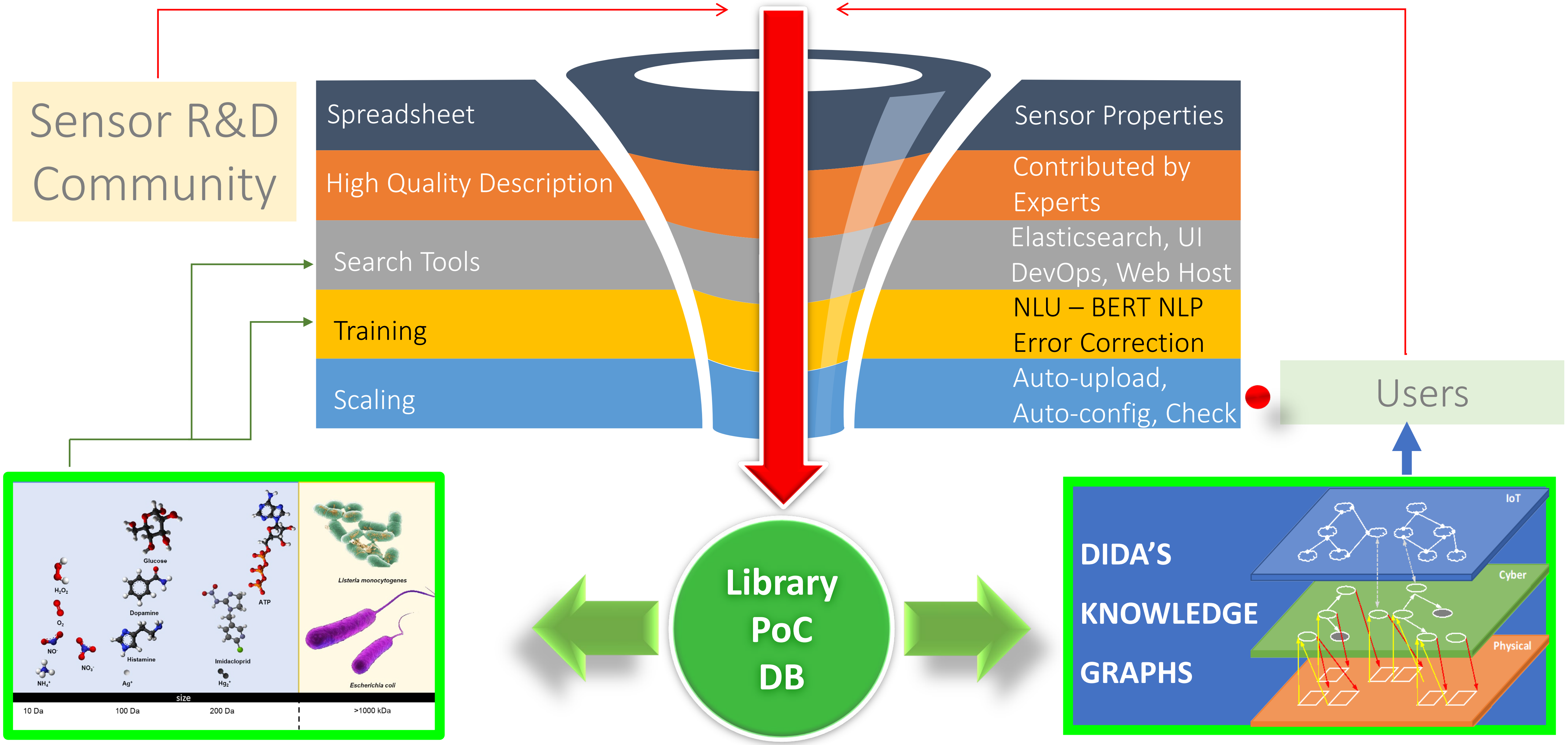
The key performance indicator (KPI) for SENSEE will be a measure of its quality of service (QoS metric) with respect to the delivery of precision responses and value of recommendations. Description of sensor types (categories, attributes) in SENSEE 1.0 may enable end-users to choose sensors relative to use cases. But, without sensor-specific data, ***relative to the use case of the end-user***, the value of SENSEE diminishes. Acquisition of data from sensors in SENSEE 2.0 will be relative to use case. For example, if Comfrey Farms wishes to optimize quality of meat color in its pork product, the outcome (desired color of pork meat) may need to converge and combine data from ammonia sensors (amount of ammonia in the hog environment), homofermentative microbial species in feed (Lactobacillus sp) and colorimetric data from robotic arm involved in meat processing. SENSEE 2.0 aims to acquire end-user case-based sensor data to address problems and questions of pragmatic value. The feasibility of this approach may be challenged by sensor manufacturers (for example, ammonia gas sensor from C2Sense, microbial sensor from Thermo-Fisher and colorimetric sensor from Omron) who may want to aggregate their own data and encrypt data ports and data loggers to prevent data interoperability and distribution. Manufacturer's portals are focused on sensors specific to the manufacturer. SENSEE 2.0 is an open platform, catalyzing synergistic integration of data to synthesize information, with respect to the end-user's problem. The potential for profitability from data fusion followed by synthesis of actionable information, may be an economic incentive for end-users. It may encourage users to support the SENSEE-DIDA'S-KIDS platform approach by uploading sensor data to SENSEE 2.0 directly from their operations. DIDA'S, and in future DIDA'S KIDS, may evolve from data-informed DSS to synthesis of relevant information, followed by the knowledge-informed paradigm in decision science.

http://bit.ly/PARTNER-WITH-PEAS

SENSEE 2.0 and DIDA’S

Data-Informed Decision as a Service

Progress of Development – SENSEE 2.0 and DIDA'S



THE LONGEST JOURNEY

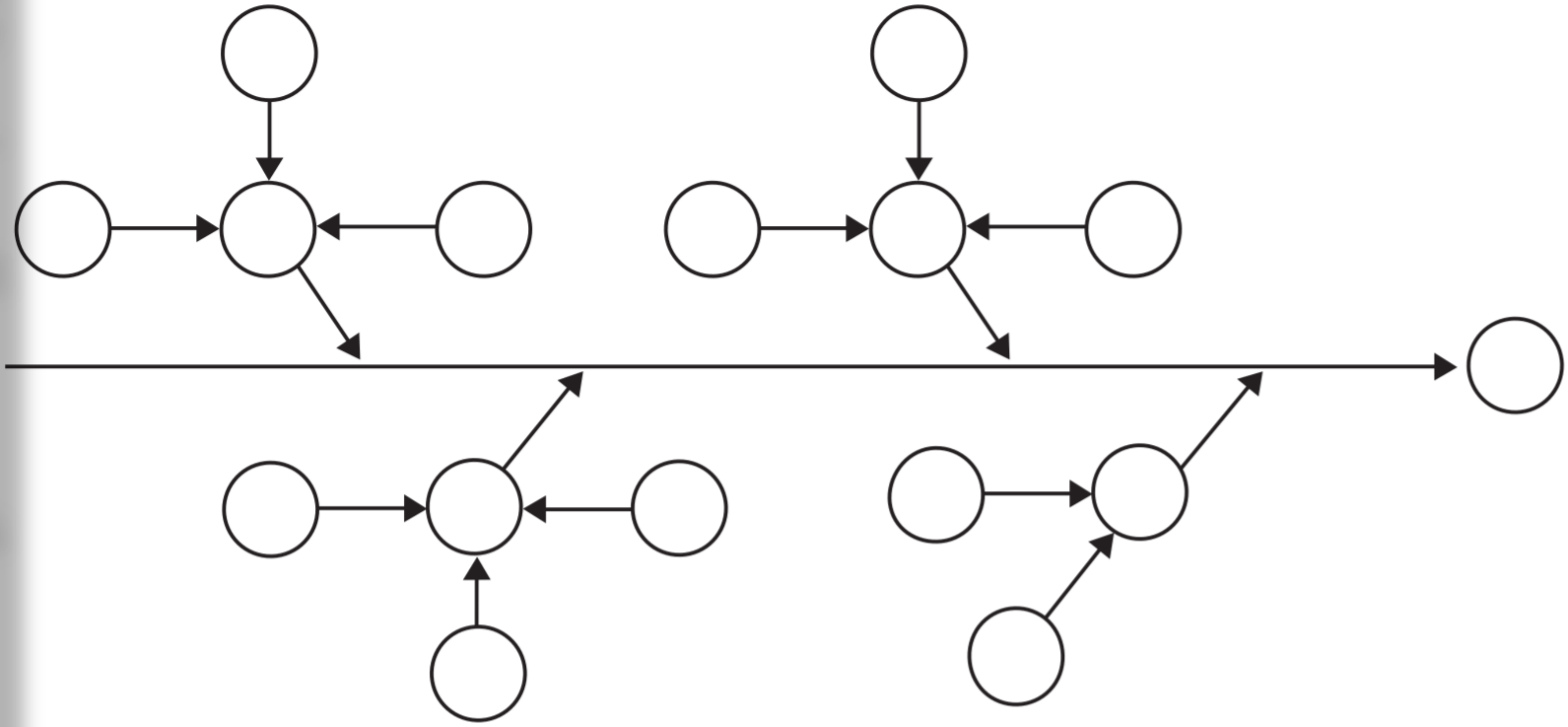
SENSEE

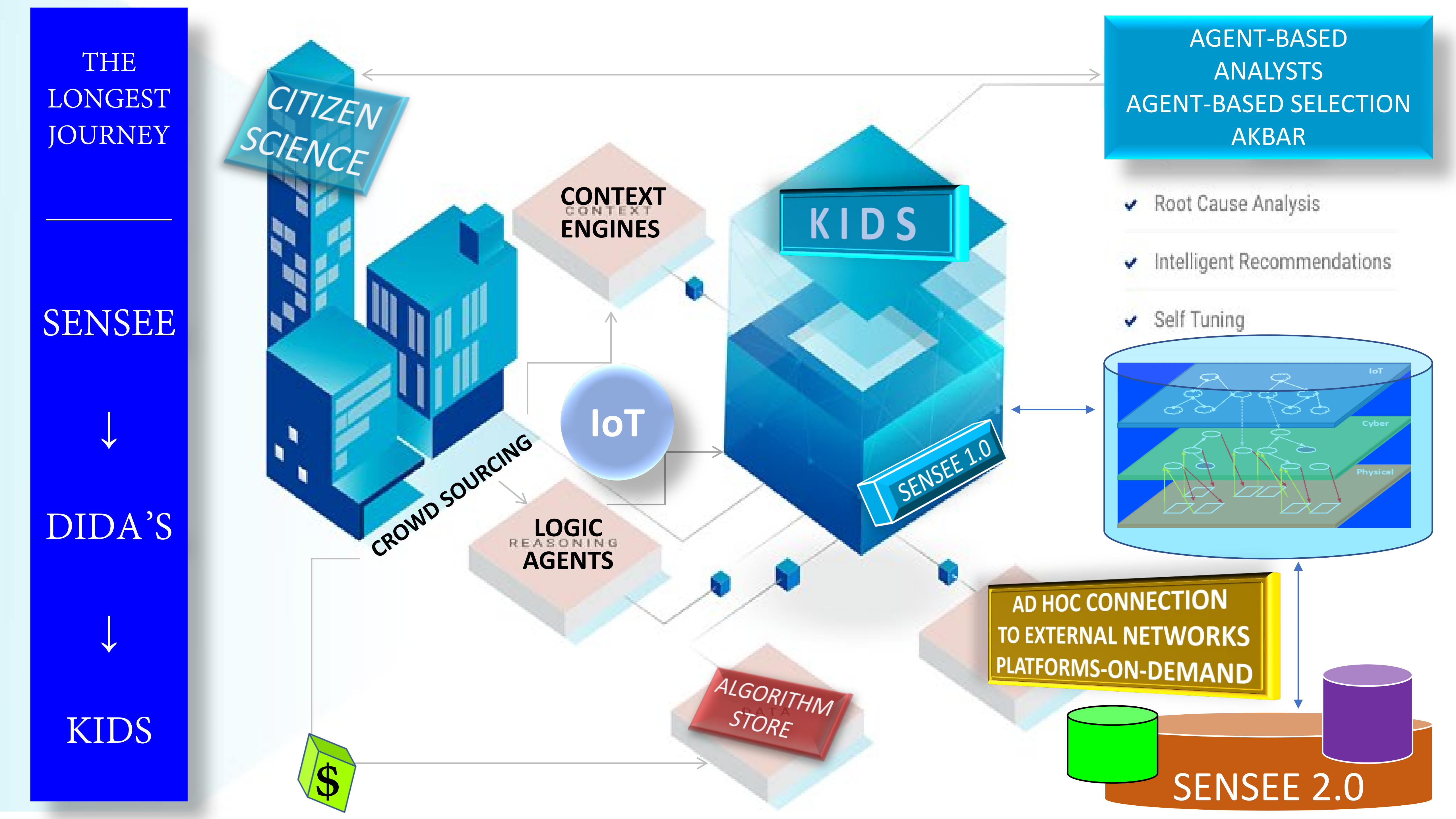
↓

DIDA'S

↓

KIDS




THE LONGEST JOURNEY
SENSEE
↓
DIDA'S
↓
KIDS
CITIZEN SCIENCE
CONTEXT ENGINES
IoT
CROWD SOURCING
LOGIC REASONING AGENTS
KIDS
SENSEE 1.0
ALGORITHM STORE
$
AGENT-BASED ANALYSTS
AGENT-BASED SELECTION
AKBAR
Root Cause Analysis
Intelligent Recommendations
Self Tuning
IoT
Cyber
Physical
AD HOC CONNECTION TO EXTERNAL NETWORKS
PLATFORMS-ON-DEMAND
SENSEE 2.0

Anticipate challenges from manufacturers and users

http://bit.ly/PARTNER-WITH-PEAS

SENSEE

2.0

Unless users allow access to raw data from sensors, the system may be unable to optimize outcomes or minimize risks, for questions which require specific case related data, from relevant sensors. With other general access data, for example, standard protocols for wastewater treatment, it may be possible to offer some degree of information or recommendation but then the value of convergence is limited in its scope.

Anticipated deliverables from SENSEE – Logic Tools?

SENSEE 1.0 (sensor descriptions) and SENSEE 2.0 (sensor-specific data) are tangible pursuits, which can deliver case based solutions, within the scope of [a] data-informed decision support for [b] limited interrelationships in a specific domain [c] restricted to information extraction and recommendation but [d] not approaching the extent of DIDA'S.

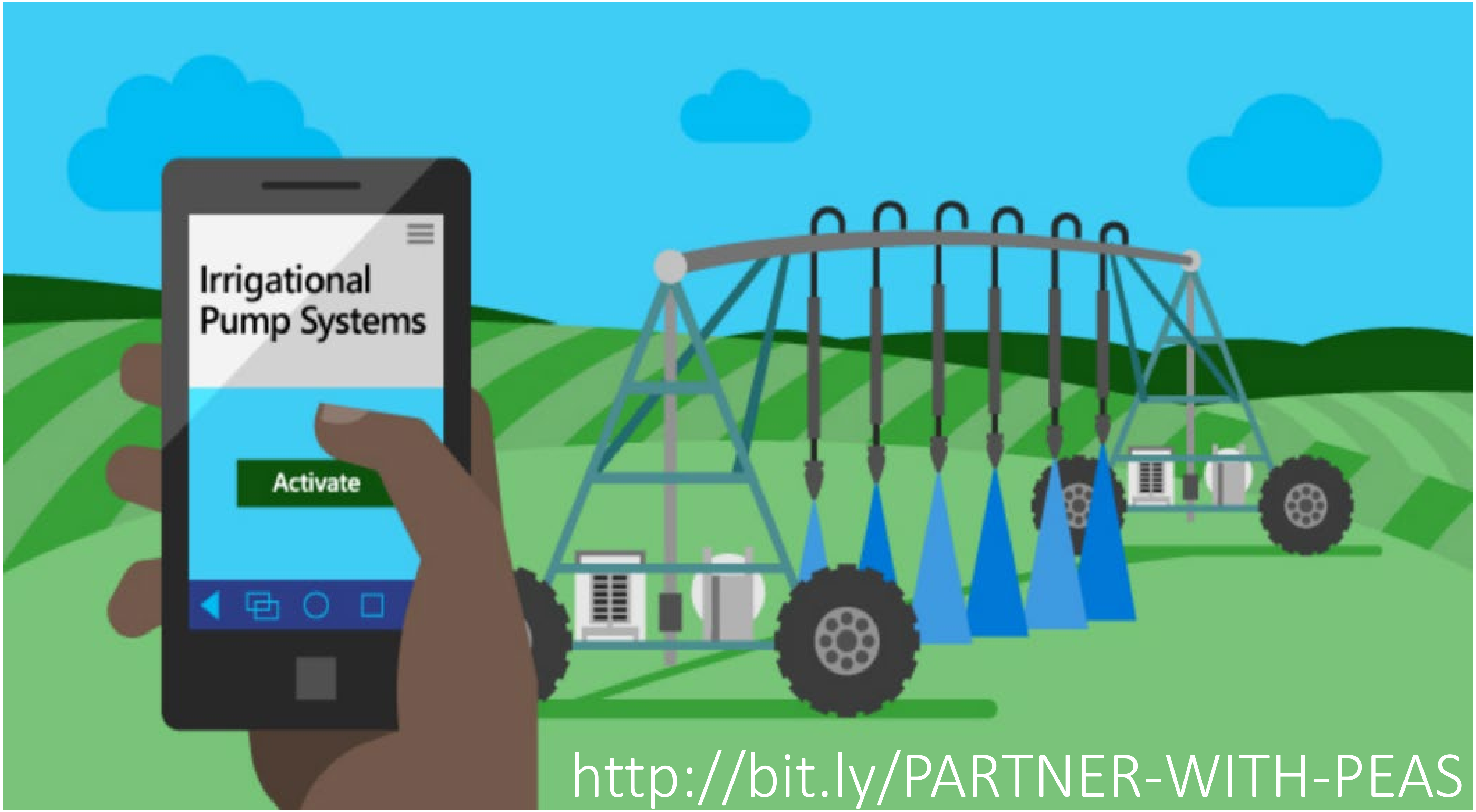
We can use data to create information (thrust of the current state of art with respect to so-called data science) and apply a set of rules and reasoning (logic, dependency, analytics, prior knowledge) to predict ***known*** unknowns, in the form of decision support for humans in the loop (recommendation without discovery or actuation) or venture to relinquish control for partial automation (risk-limited actuation) in an IFTTT (workflow) approach to basic service.

This is a form of data-driven, evidence-driven, ***reasoning*** solution with potential for partially automating workflow. The efficiency gains anticipated from "intelligent" decision support systems lies in our ability to integrate ***logic rules***. DIDA'S KIDS includes this format, as the foundation. Logic rules, if understood (semantics), integrated, optimized, and executed, may be the answer to 80% of the global problems, for a tiny fraction of the cost, which may accelerate market adoption and penetration of digital-agro services. Remaining 20% of issues may require DIDA'S KIDS to create dynamic knowledge composable tools embedded with statistical and mathematical modeling based machine learning solutions. These two approaches may be complimentary for 20% of the problems. But, knowledge tools may not be as critical for 80% of our everyday problems, eg, optimizing and actuating (partial automation) of irrigation water pump systems (control volume and distribution of water) based on soil moisture, salinity, ionic content and weather. Thus, we can focus on **logic tools.**

Anticipated deliverables from SENSEE (SENSEE.ES)

SENSEE 1.0 (sensor descriptions) and SENSEE 2.0 (sensor-specific data) repositories, combined with analytics and artificial reasoning, harks back to “expert systems” which preceded the snake oil sales of AI (artificial intelligence) to the stage where hyped-up “AI systems” exploded to near-extinction (1990’s “winter of AI”). The recent re-invention of AI (2010’s) has catalyzed its re-entry into the den of vipers. SENSEE may stay clear of the foggy panache of AI and focus on delivering expert services, in near real-time, which are profitable for users. An expert service requires we create a framework for an expert system to partially mimic the decision-making ability of human experts, who can solve problems by using data and reasoning aided by prior knowledge. The SENSEE concept of expert service (ES) is not the 1980’s version of expert systems. In SENSEE.ES we will use advanced tools: elasticsearch, NLP, semantic catalogs, graph networks, machine learning, and digital-by-design concepts from the internet of things (IoT), using mobile, agile, standards-based tools to optimize data interoperability, semantic intertoperability, technical interoperability (open platform approach) and, ambitious of all, policy interoperability, to be globally adaptable. In the hands of the human analyst, SENSEE 2.0 is the data source to extract evidence and make informed decisions to act on the evidence (SENSEE 2.0 data). The human analyst supplements this decision making logic using domain expertise and experience in the organization (enterprise, farm, factory) to prescribe analytics and orchestrate any necessary course of action based on the data, processes and reasoning. An Agent-based system (ABS) emulating this “human” step (a part of the **logic tools portfolio** of SENSEE.ES) plus IFTTT (if this then that) type workflow based low level decision-driven partial automation, if combined, may suffice to solve many problems, eg, irrigation water flow.

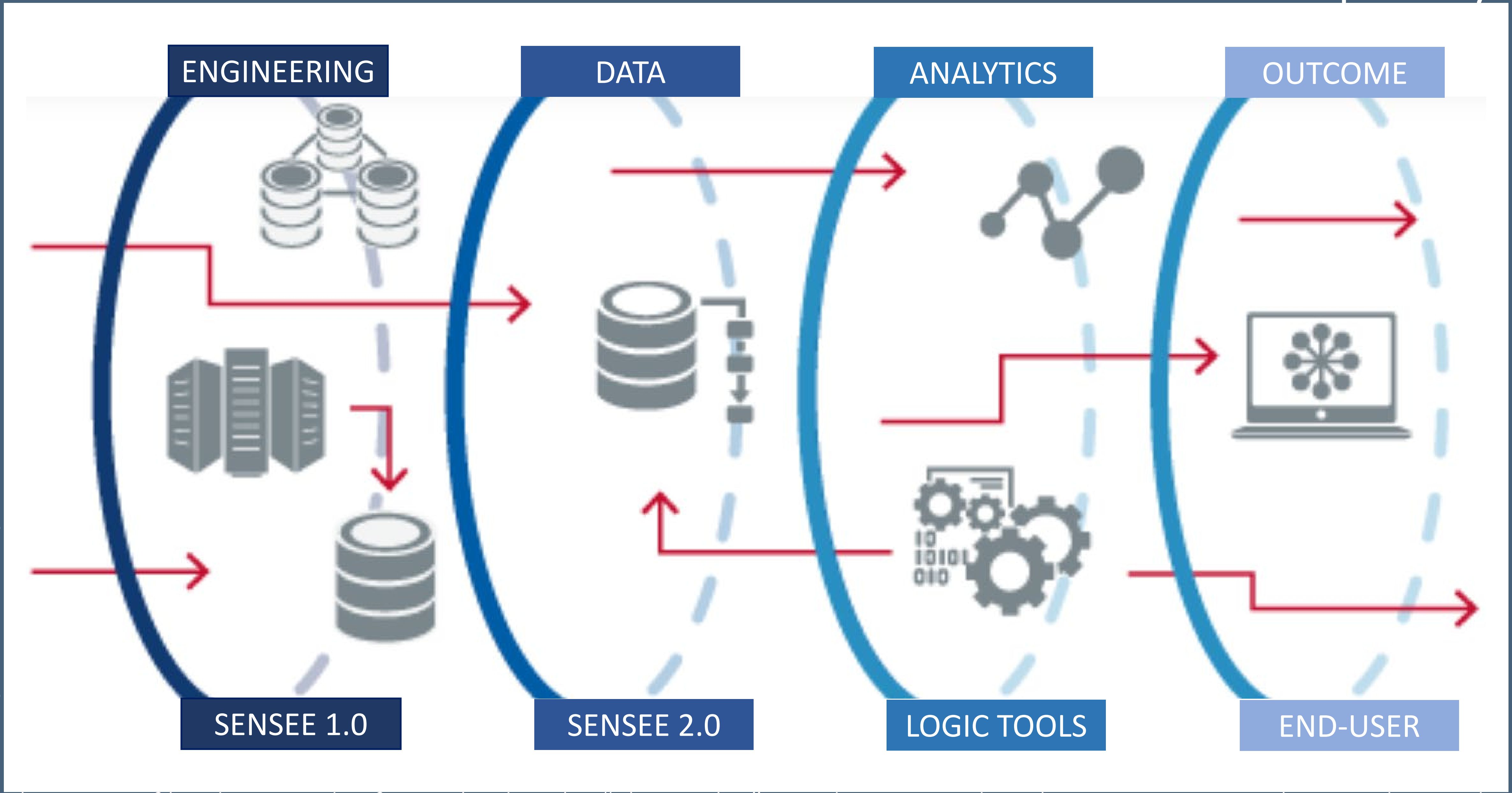
Deliverables from SENSEE – Logic Tools – call it ART ?



https://cloudblogs.microsoft.com/industry-blog/microsoft-in-business/2019/05/15/how-polaris-energy-services-is-transforming-the-agriculture-industry-in-the-cloud/

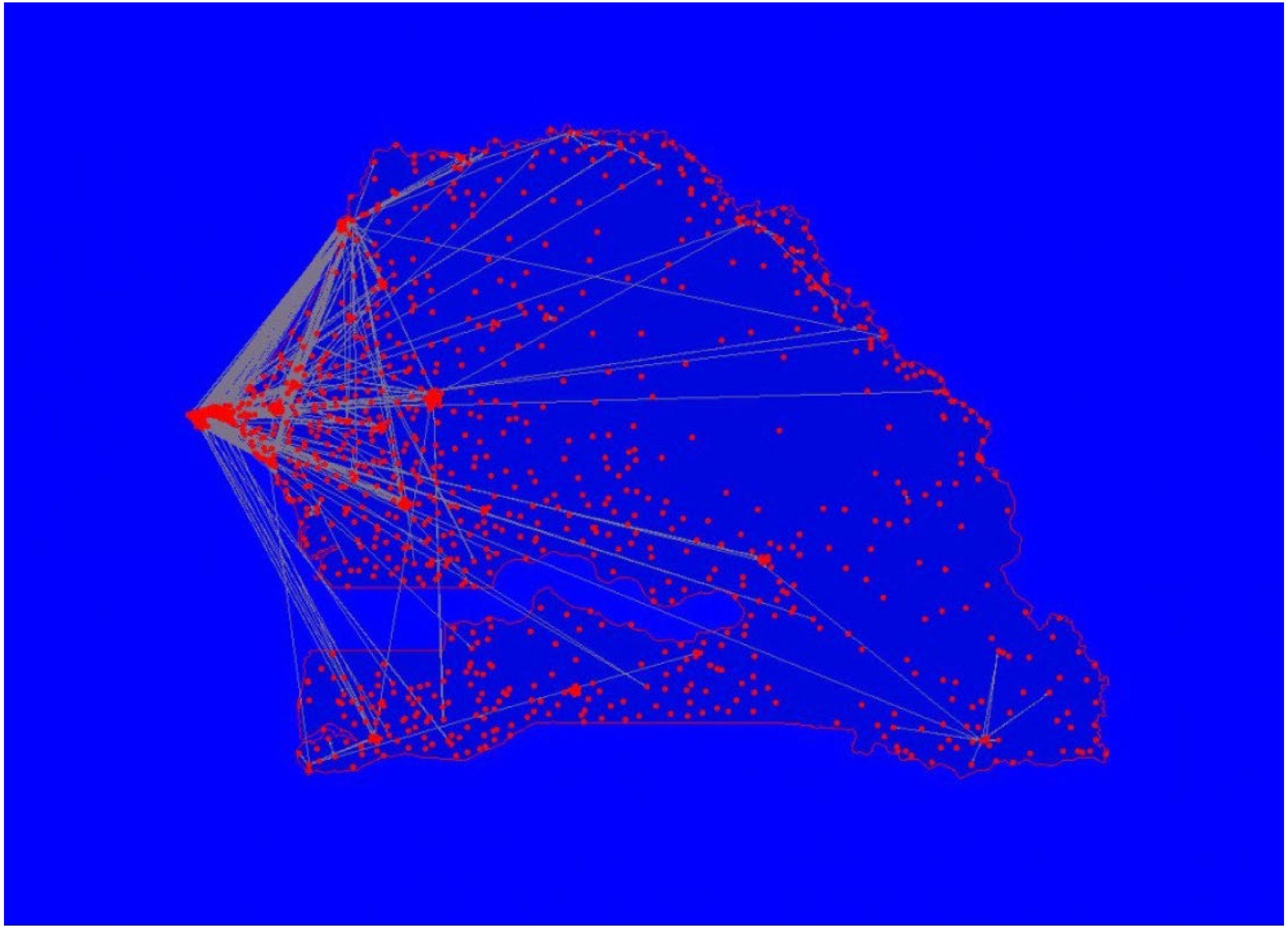
Short-term deliverable from SENSEE? ART of Simplicity

http://bit.ly/PARTNER-WITH-PEAS



The nexus of hardware and software lead to the "Plug-n-Play" paradigm. Extending that synergistic simplicity to data and data-informed decision support (DIDA'S) may evolve into DADA ("Drag and Drop Analytics") and the subset SENSOR DADA.

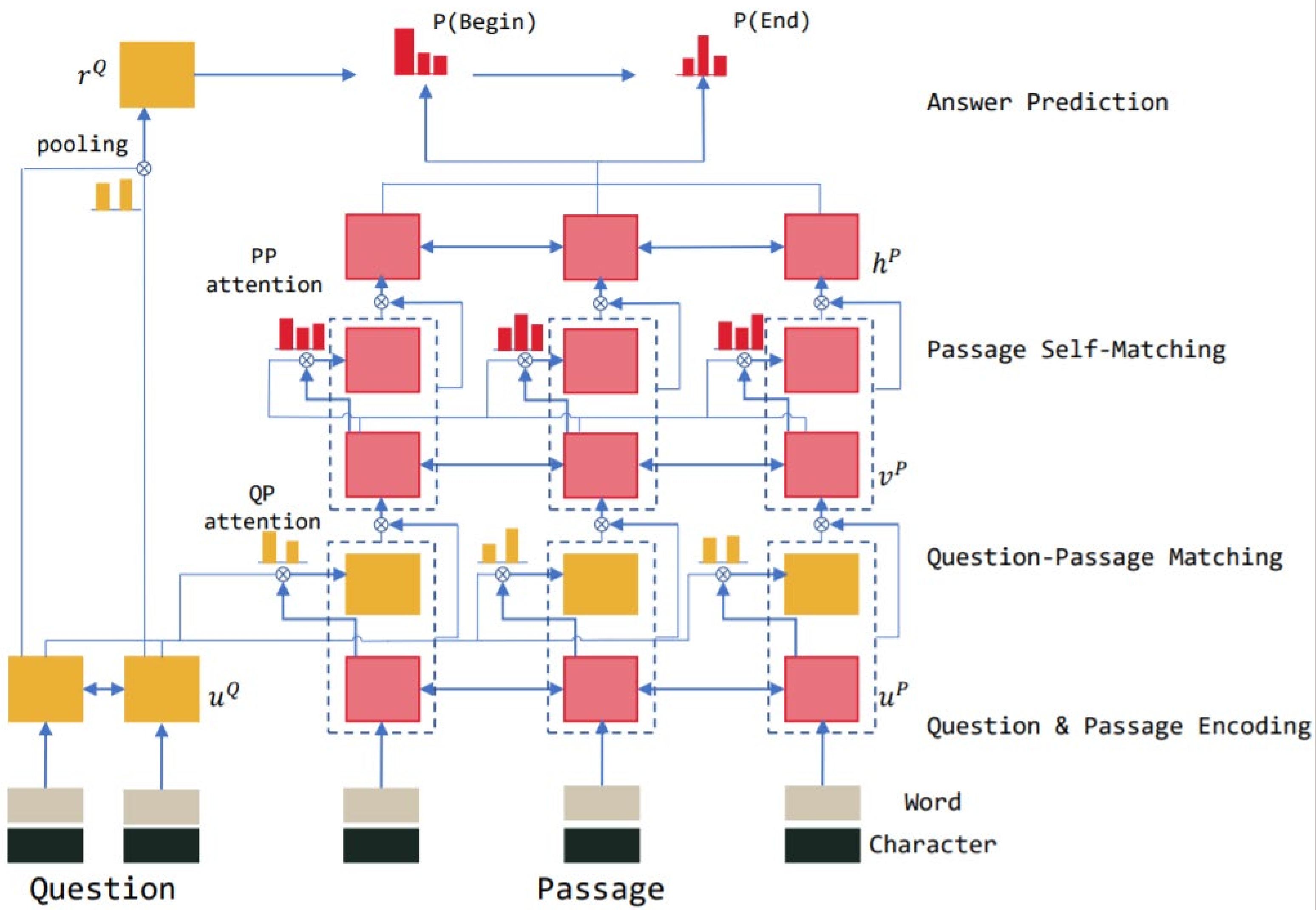
Short-term deliverable from SENSEE? ART of Aggregation



https://arxiv.org/ftp/arxiv/papers/1504/1504.03899.pdf and www.technologyreview.com/s/613987/how-much-electricity-does-a-country-use-just-ask-cell-phone-users

Short-term deliverable from SENSEE? Use natural language

Machine Reading Comprehension (MRC) uses neural network architecture, Reasoning Network (R-Net), to the mimic inferencing process (constrained by subject/predicate optimization/alignment).
https://arxiv.org/pdf/1609.05284.pdf
Another tool is BERT NLP which is also undergoing a series of tests.
https://arxiv.org/pdf/1810.04805.pdf
https://rajpurkar.github.io/SQuAD-explorer/
Convergence of MRC, R-Net, BERT, XLNet, HAN (hierarchical attention networks), etc, with KG (knowledge graphs) may help to mine contextual word embeddings. It may evolve as a tool not only for Q&A but for non-obvious relationship analysis (NORA) and extraction.
https://arxiv.org/pdf/1906.08237.pdf
https://arxiv.org/pdf/1810.06033.pdf
www.nature.com/articles/s41586-019-1335-8



https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf

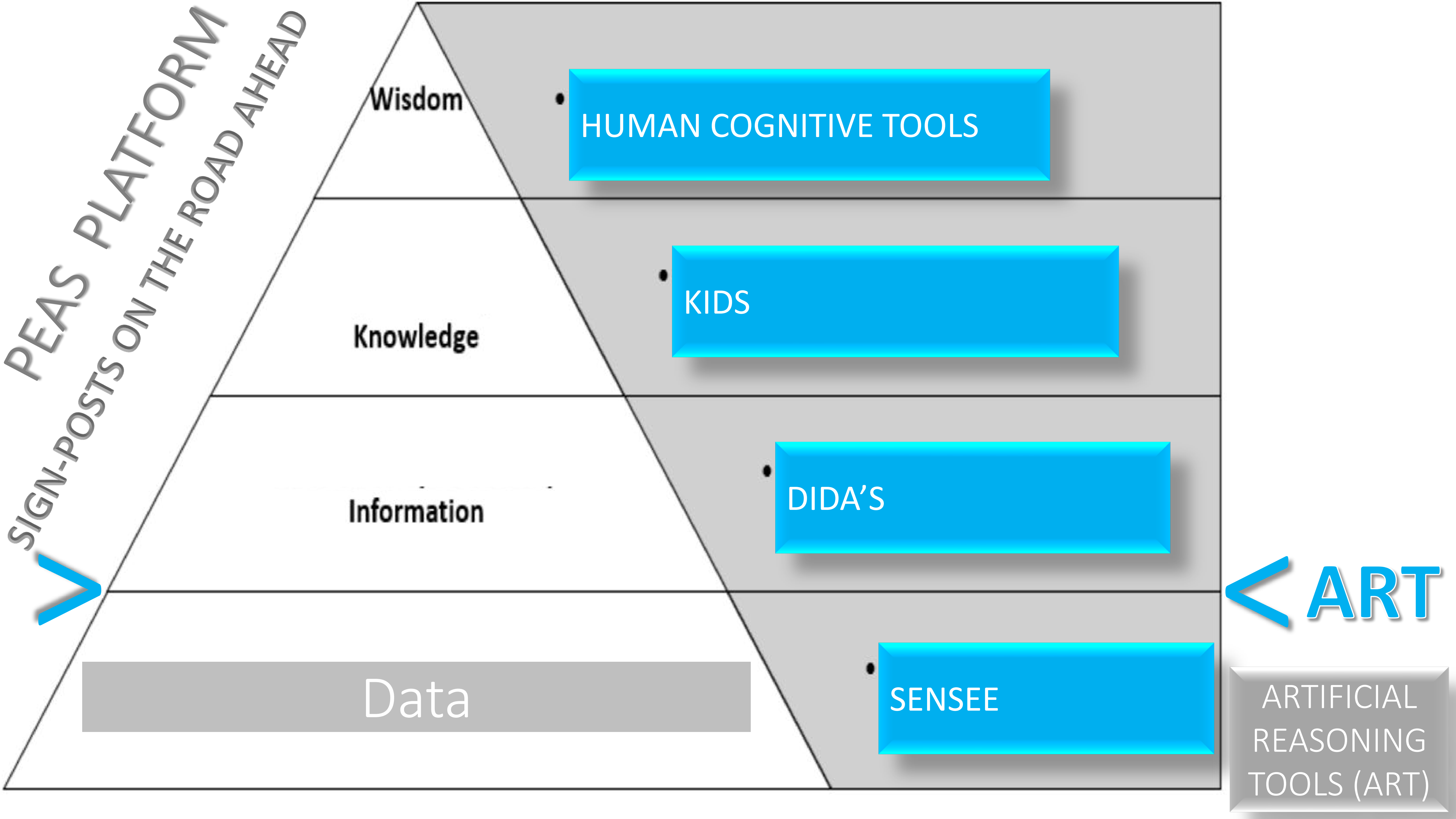
www.microsoft.com/en-us/research/wp-content/uploads/2017/05/r-net.pdf ■ www.microsoft.com/en-us/ai/ai-lab-experiments?activetab=pivot1:primaryr7

Deliverables from SENSEE – Logic Tools – call it ART ?

ART

Artificial Reasoning Tools

SENSEE leads us to ART, a logical middle ground that may deliver decision tools, as partial solutions for problems bounded by domains (not too expansive in scope) before DIDA'S KIDS.


PEAS PLATFORM
SIGN-POSTS ON THE ROAD AHEAD
Wisdom
Knowledge
Information
Data
HUMAN COGNITIVE TOOLS
KIDS
DIDA'S
SENSEE
< ART
ARTIFICIAL REASONING TOOLS (ART)

But, is knowledge still the key performance indicator?

- P
- E
- A
- S

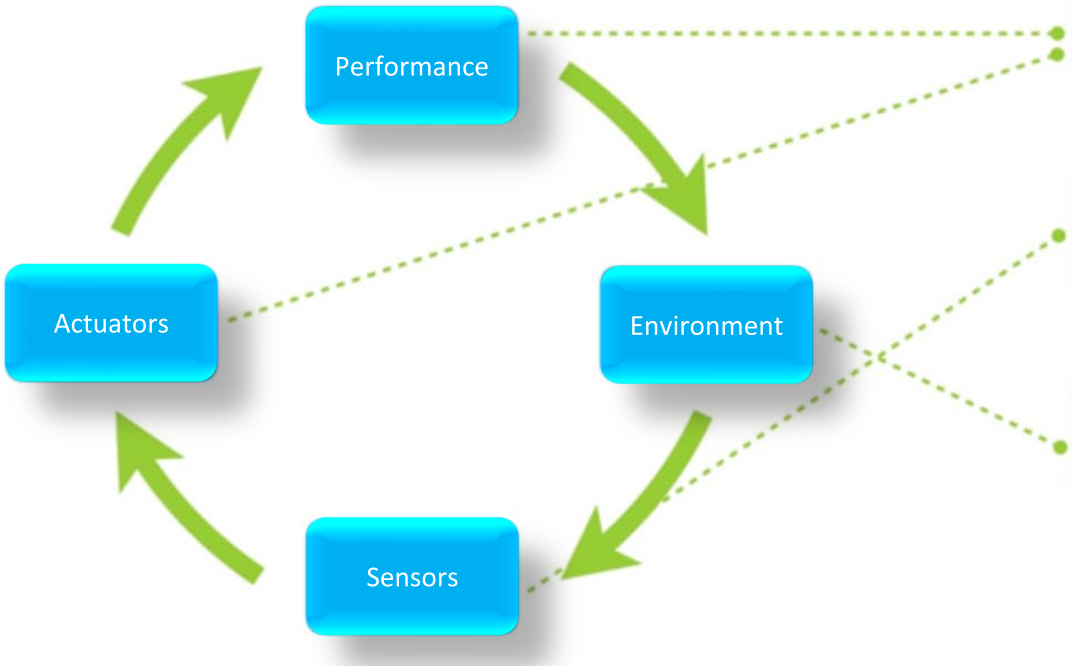
- Performance

KIDS

- Sensors

Knowledge is the ultimate key performance indicator

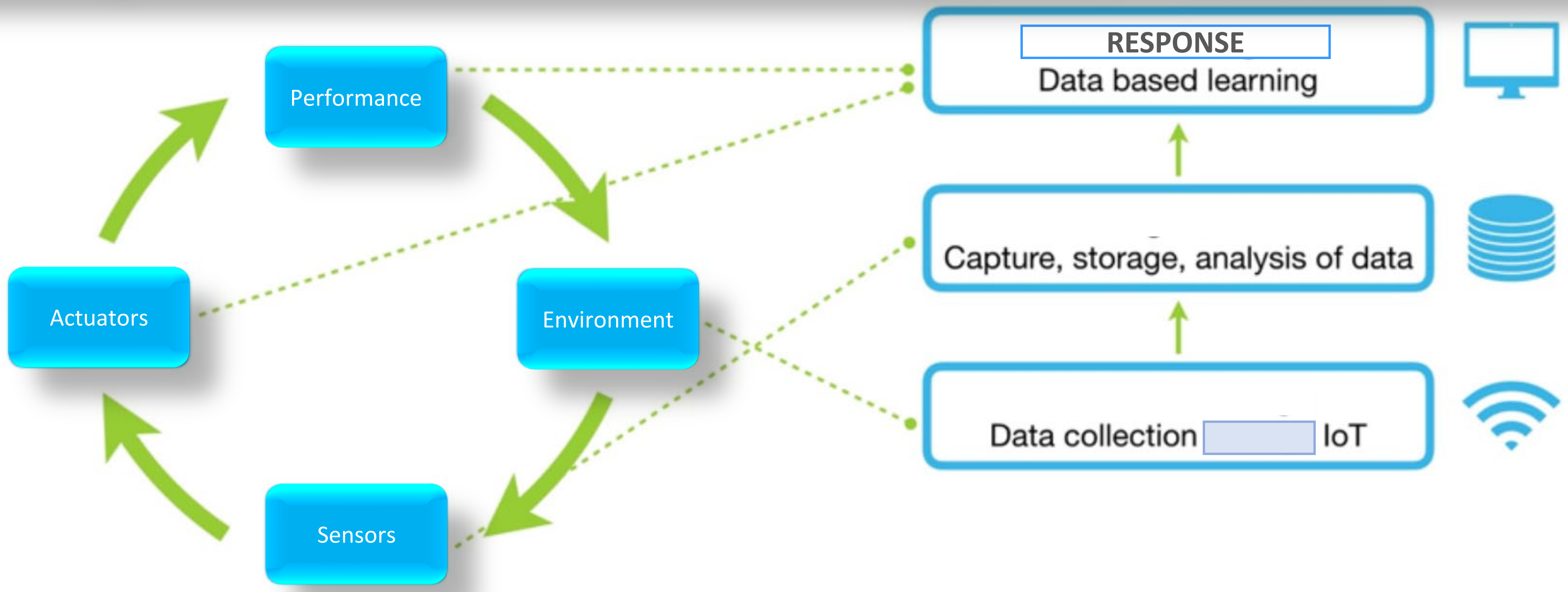
http://bit.ly/PARTNER-WITH-PEAS



https://courses.edx.org/asset-v1:ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_intelligent_agents_new__1_.pdf

Knowledge is the ultimate key performance indicator

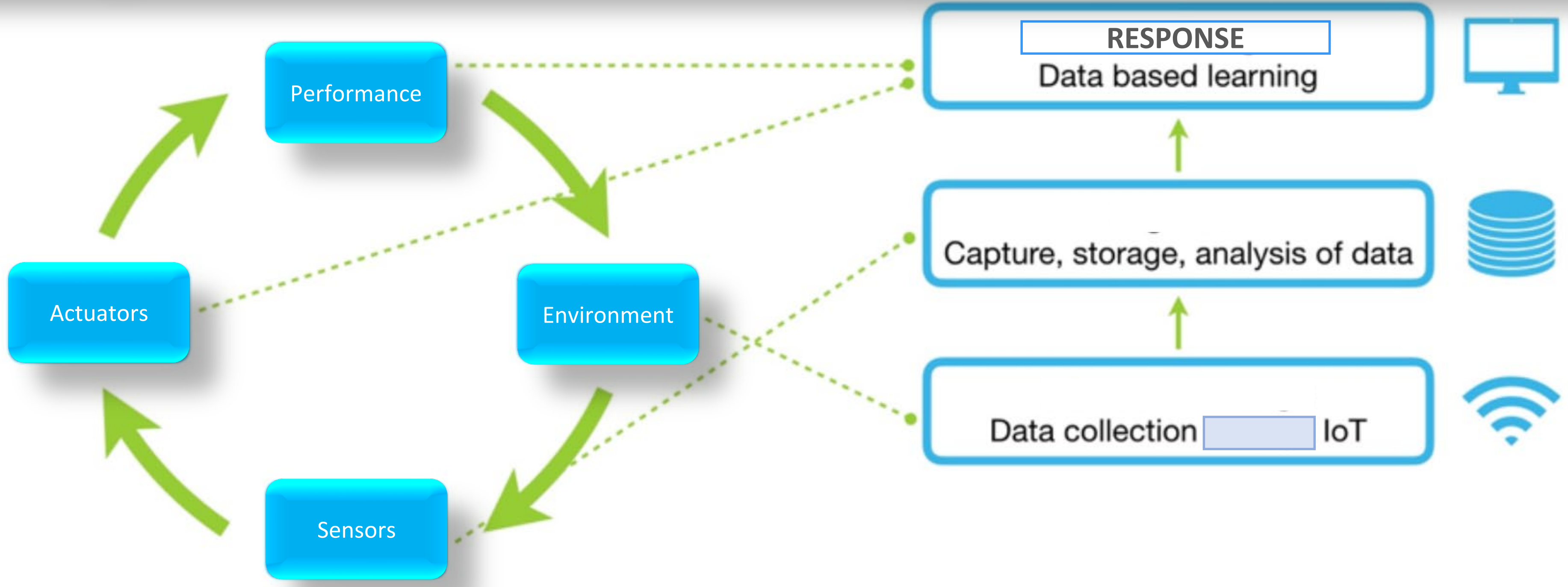
http://bit.ly/PARTNER-WITH-PEAS

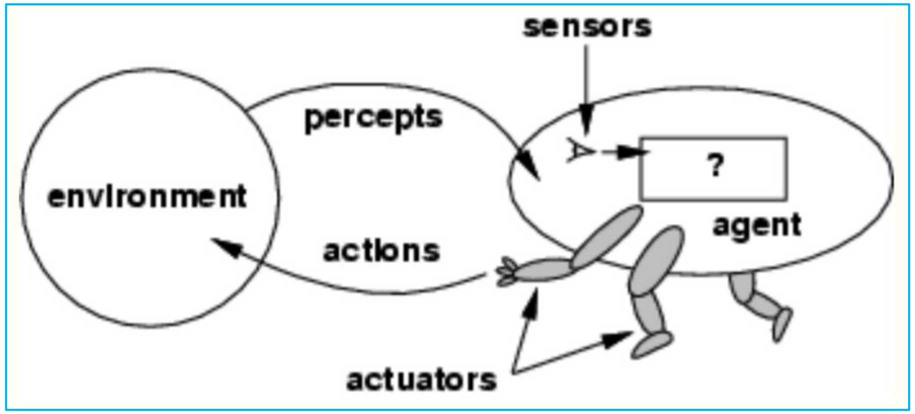


https://courses.edx.org/asset-v1:ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_intelligent_agents_new__1_.pdf

Knowledge is the ultimate key performance indicator

http://bit.ly/PARTNER-WITH-PEAS





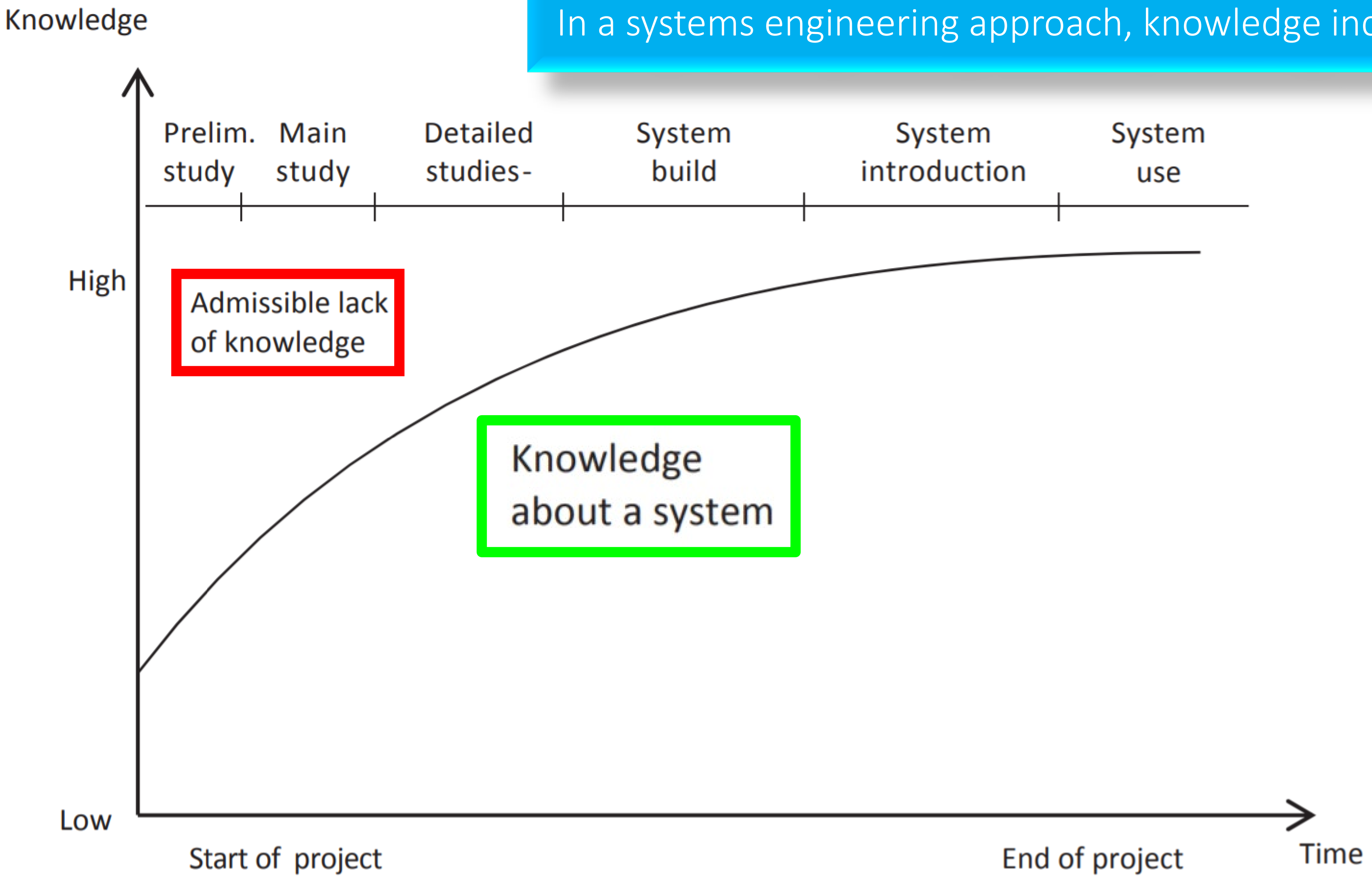
https://www.ics.uci.edu/~welling/teaching/ICS171fall10/Agents171Fall10.pdf

https://courses.edx.org/asset-v1:ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_intelligent_agents_new__1_.pdf

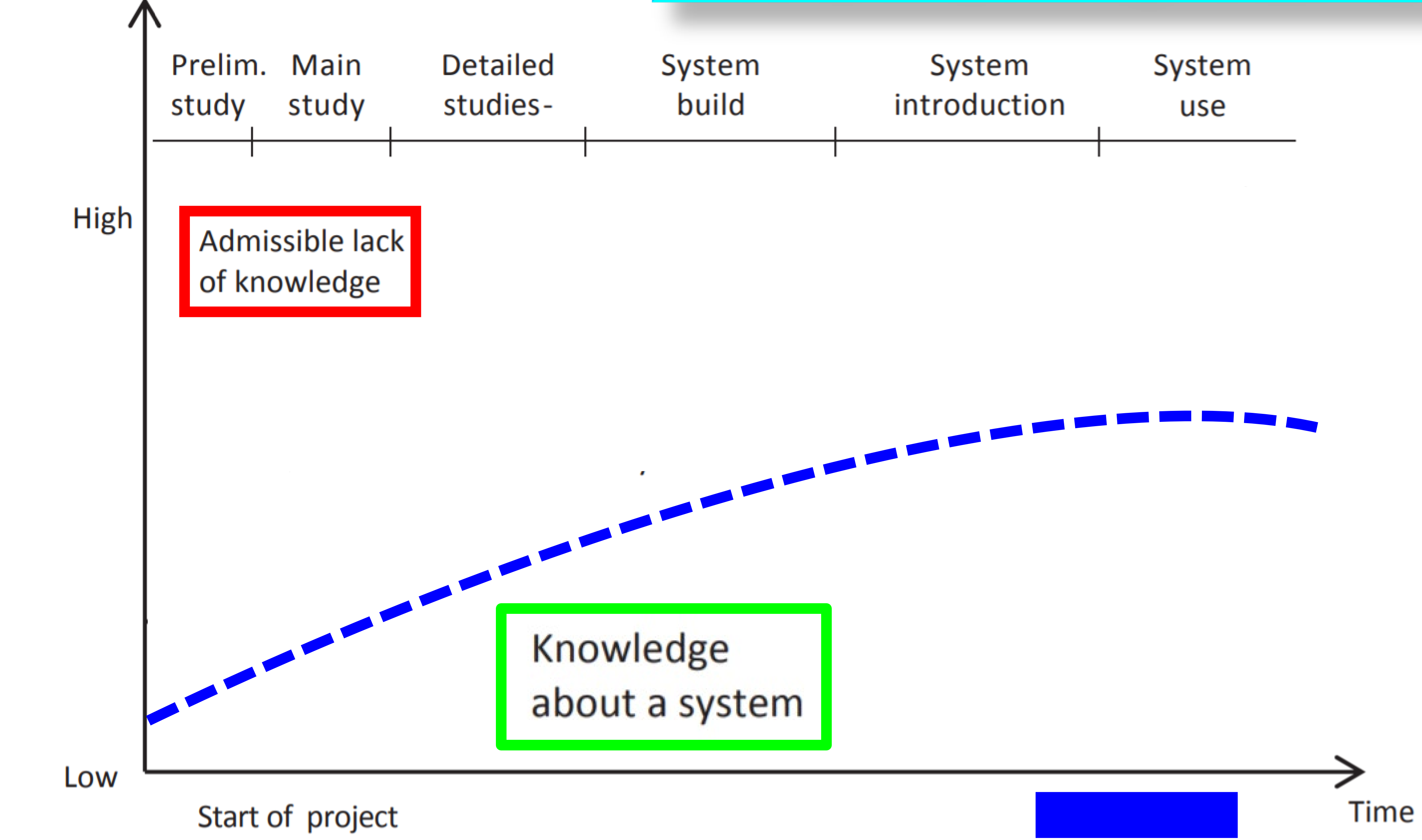
Have we gained ***knowledge*** from data and decisions?

An open question, for the long run …

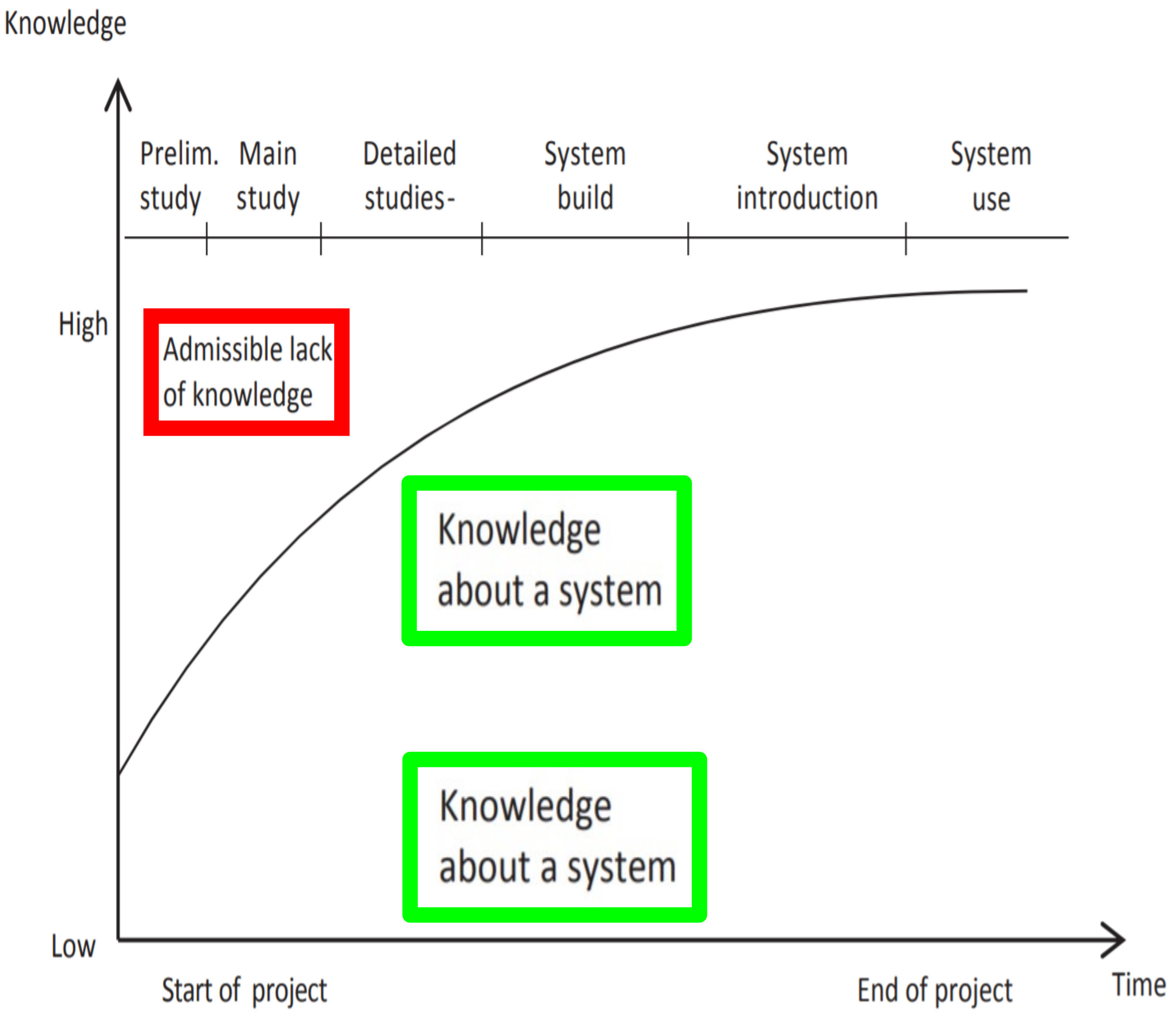
In a systems engineering approach, knowledge increases.

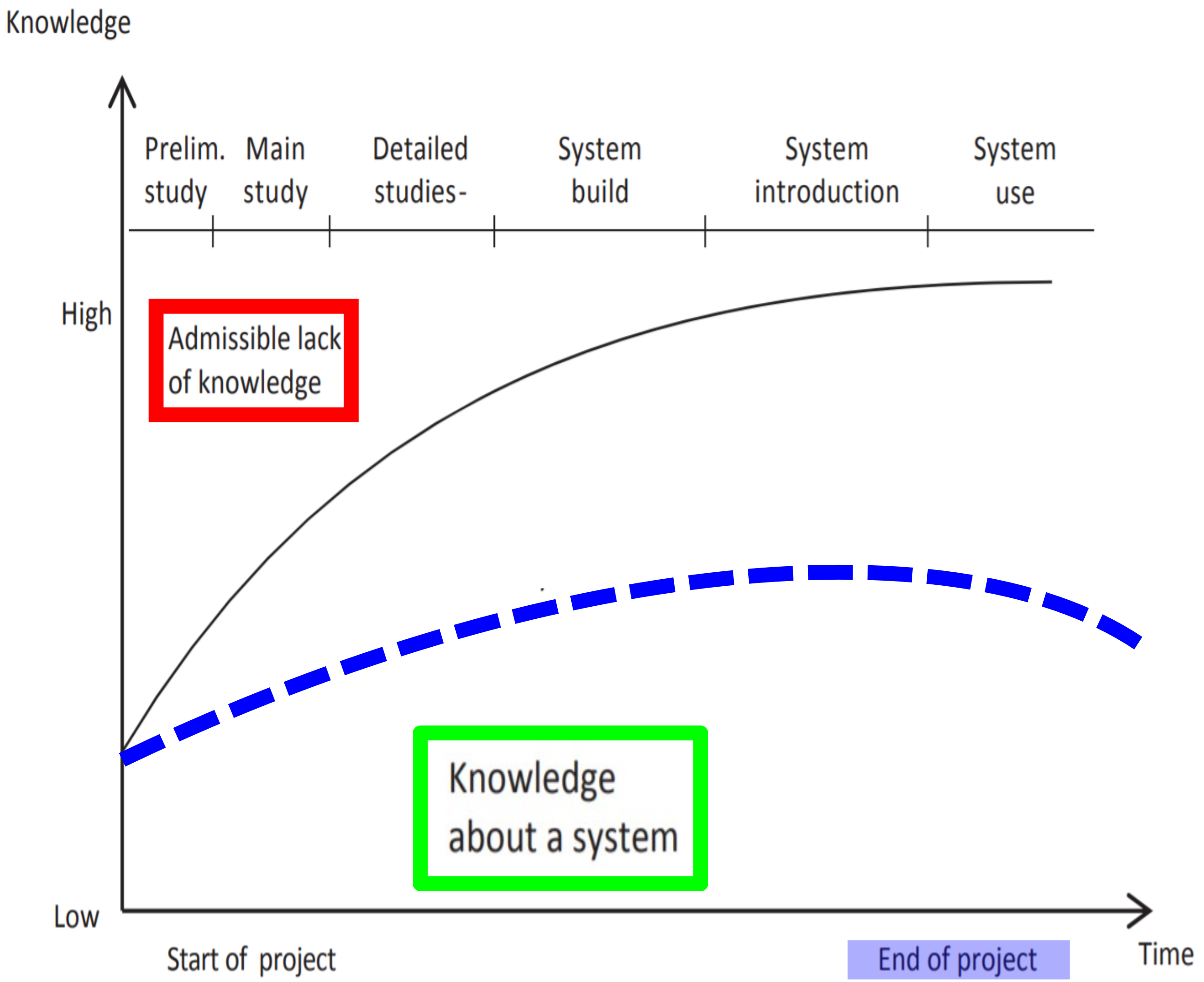


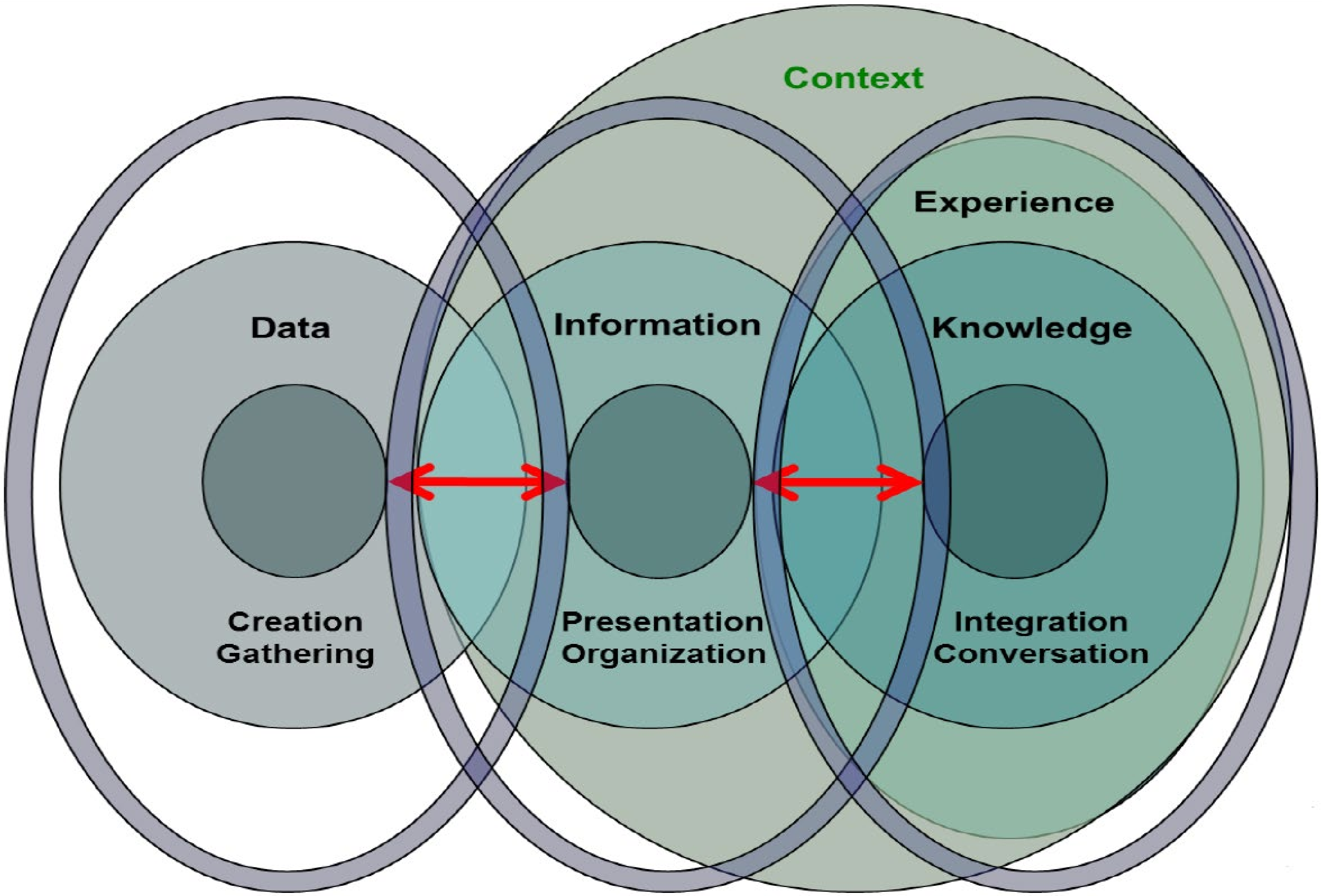
In the real world, knowledge increase may be bit sluggish.



Have we gained knowledge from data and decisions?

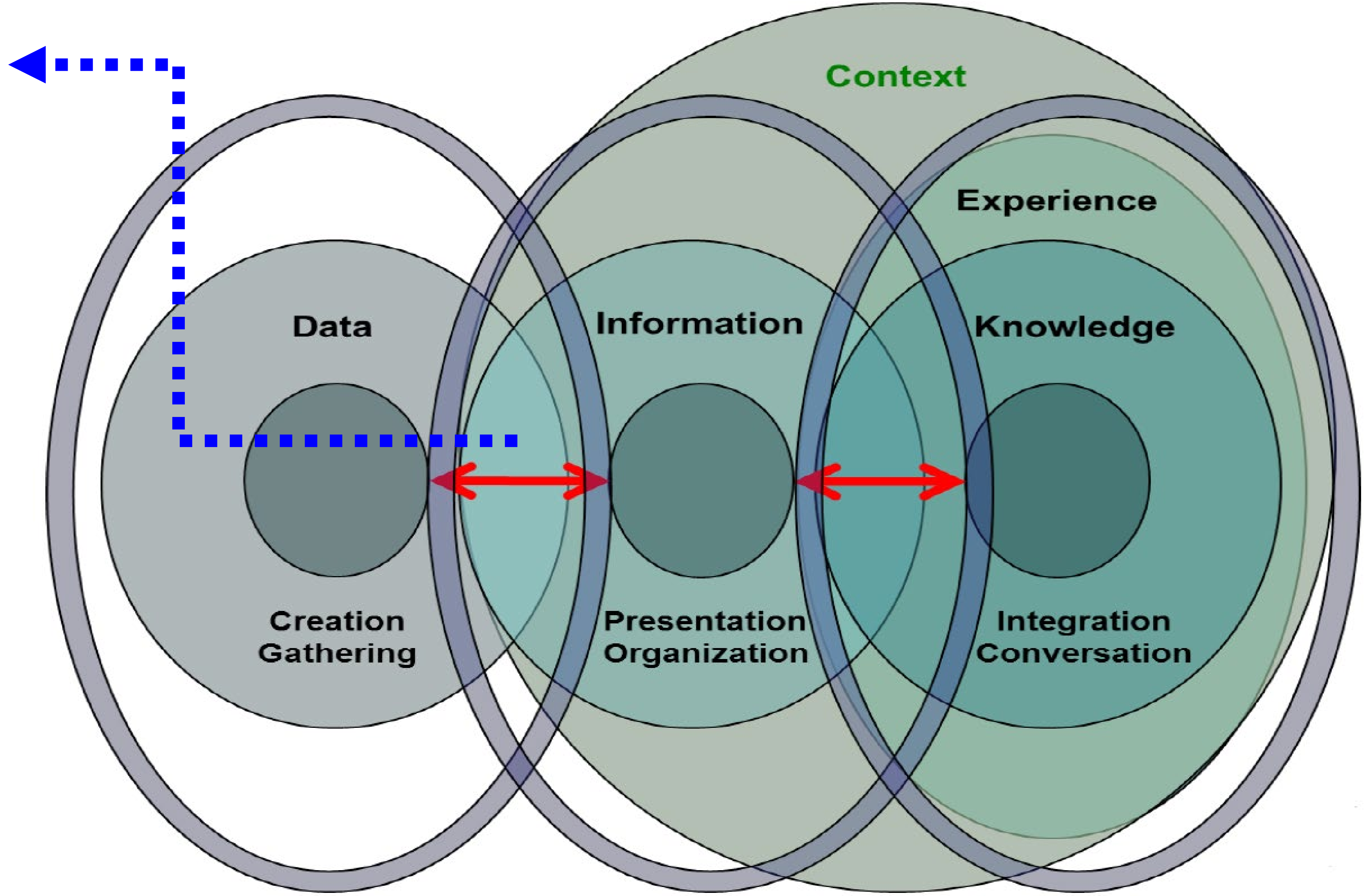


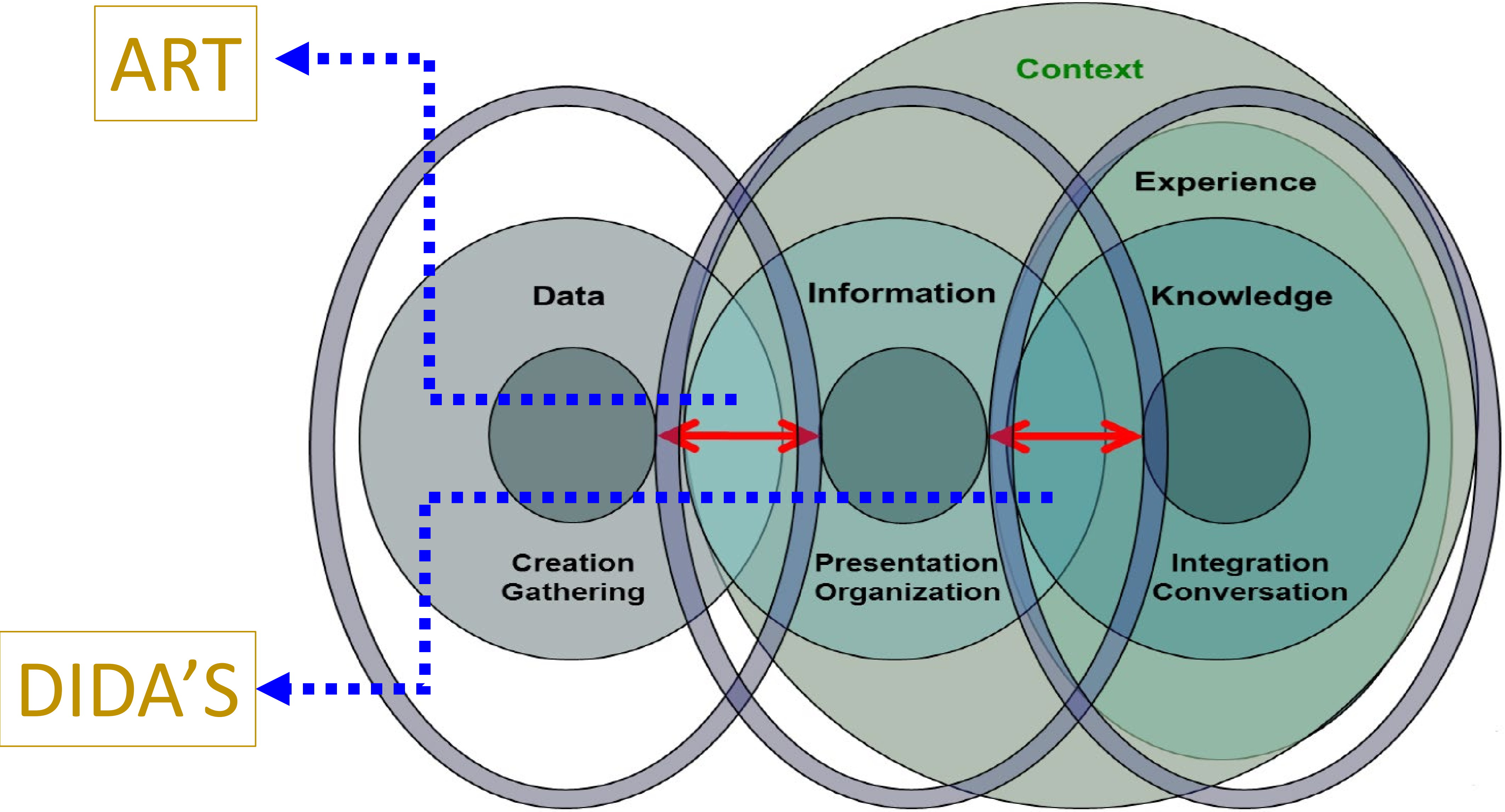



Context
Experience
Data
Information
Knowledge
Creation
Gathering
Presentation
Organization
Integration
Conversation

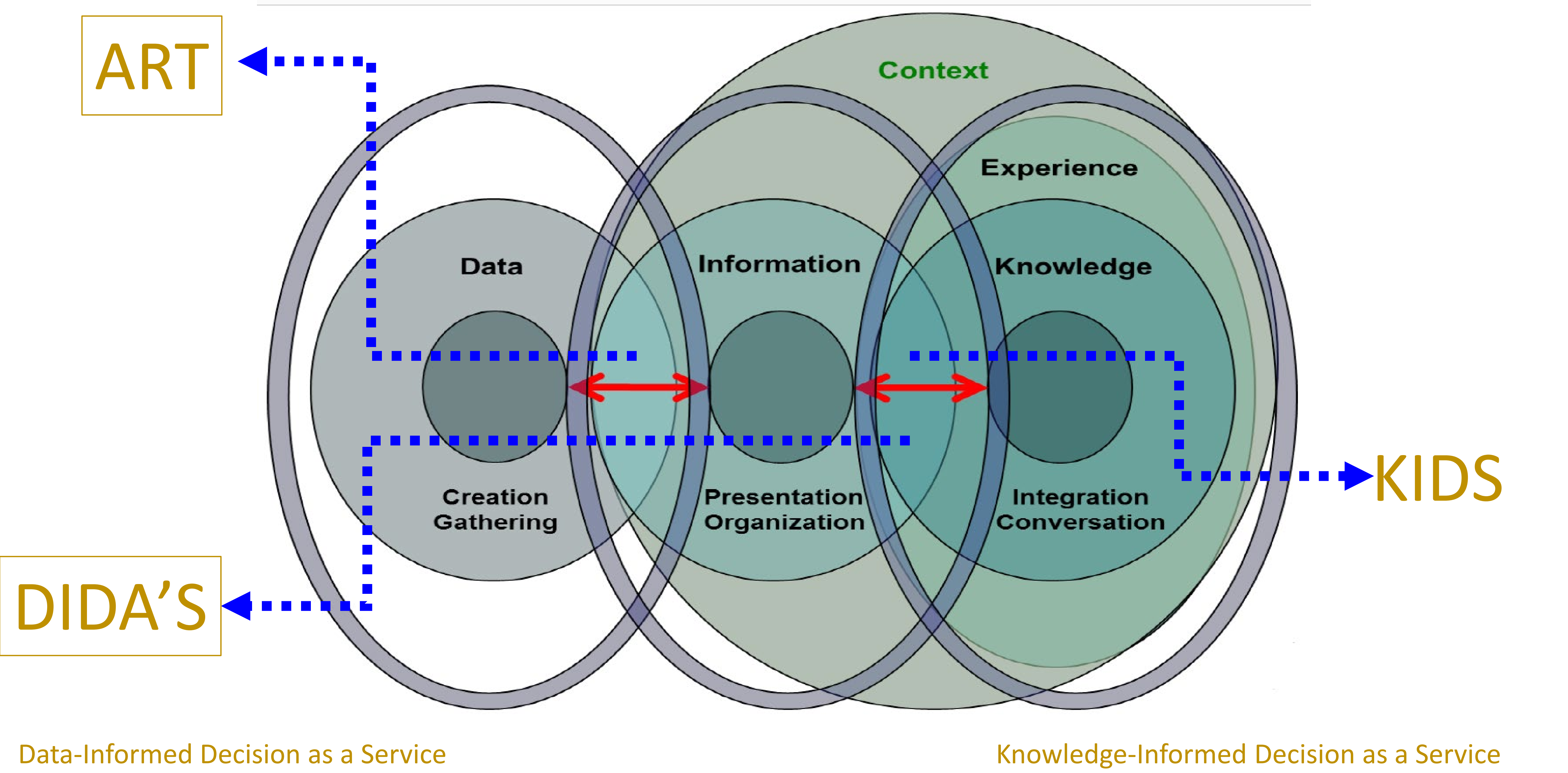
Artificial Reasoning Tools (ART)

ART



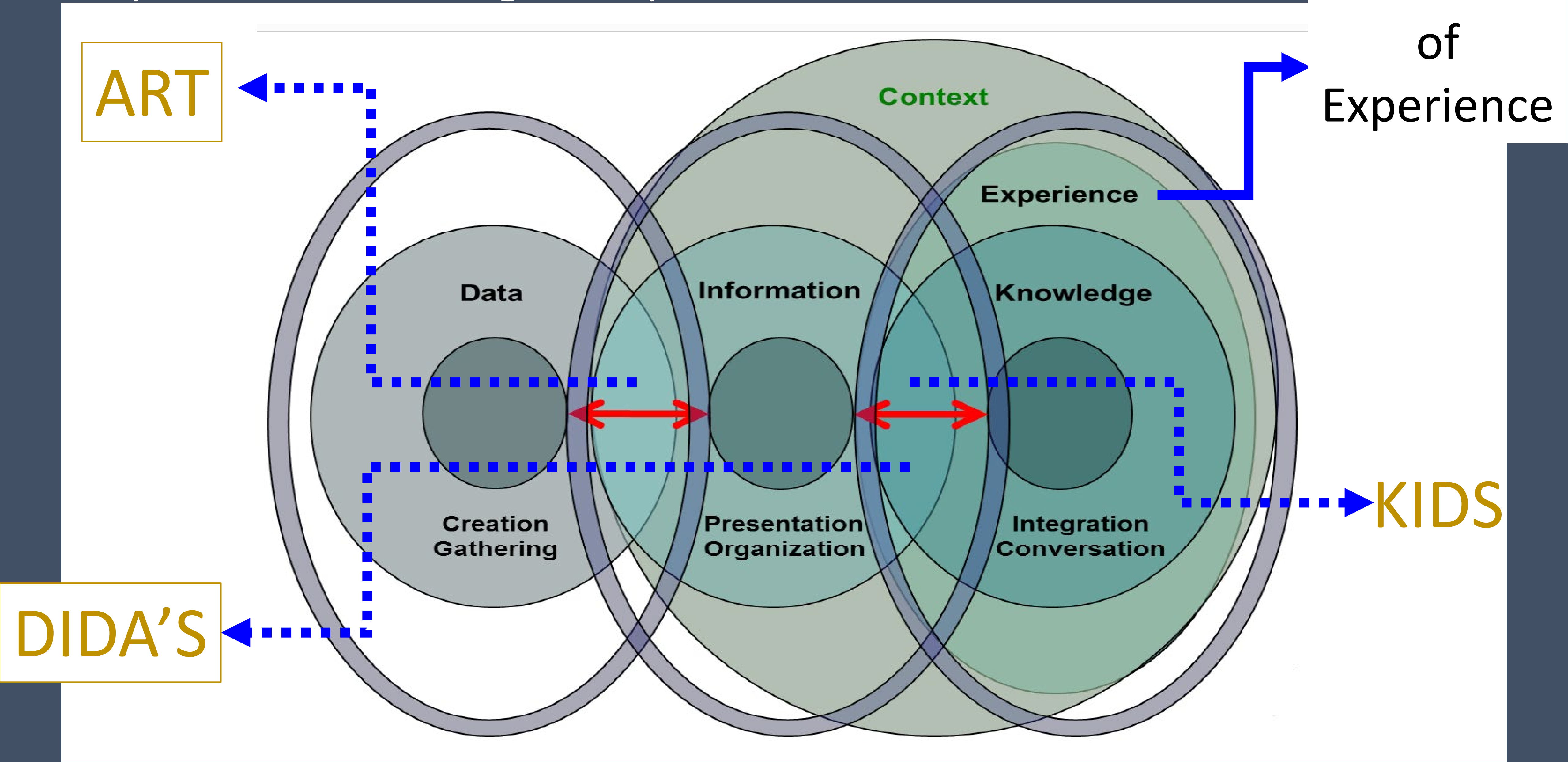


Data-Informed Decision as a Service



ART
Context
Experience
Data
Information
Knowledge
Creation
Gathering
Presentation
Organization
Integration
Conversation
KIDS
DIDA'S
Data-Informed Decision as a Service
Knowledge-Informed Decision as a Service

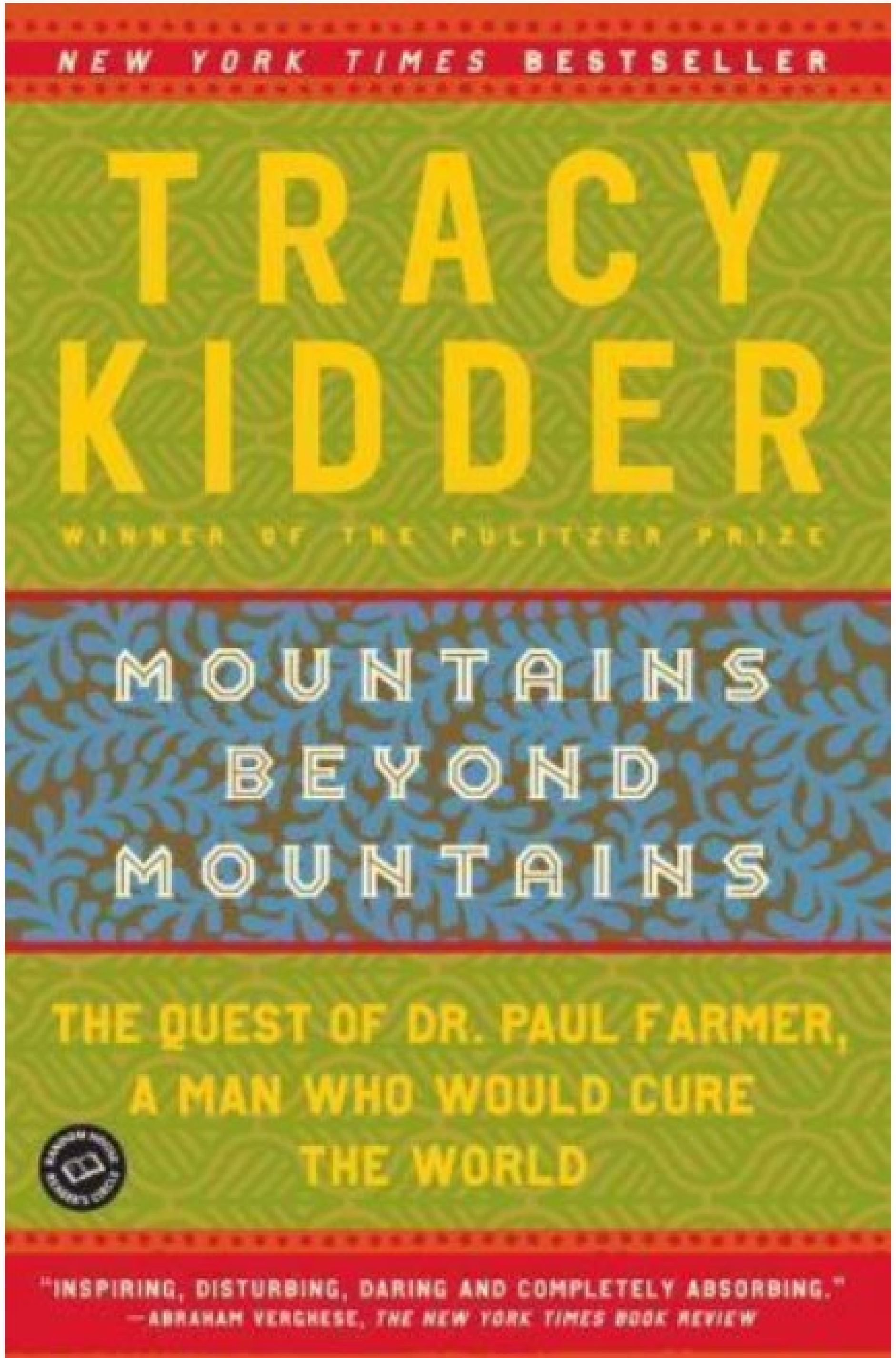
Beyond knowledge, experience

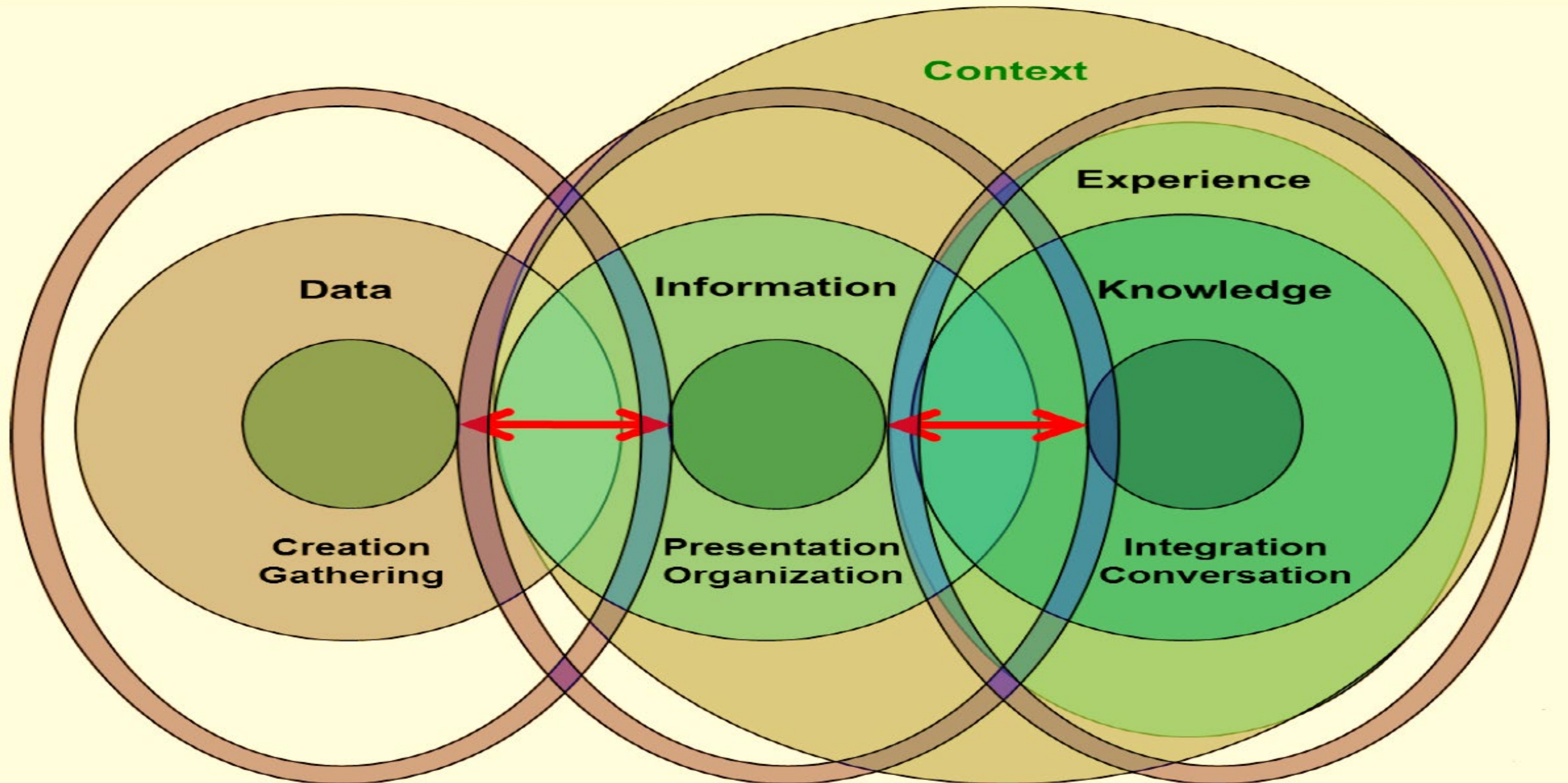


Cartoon by Jim Hendler

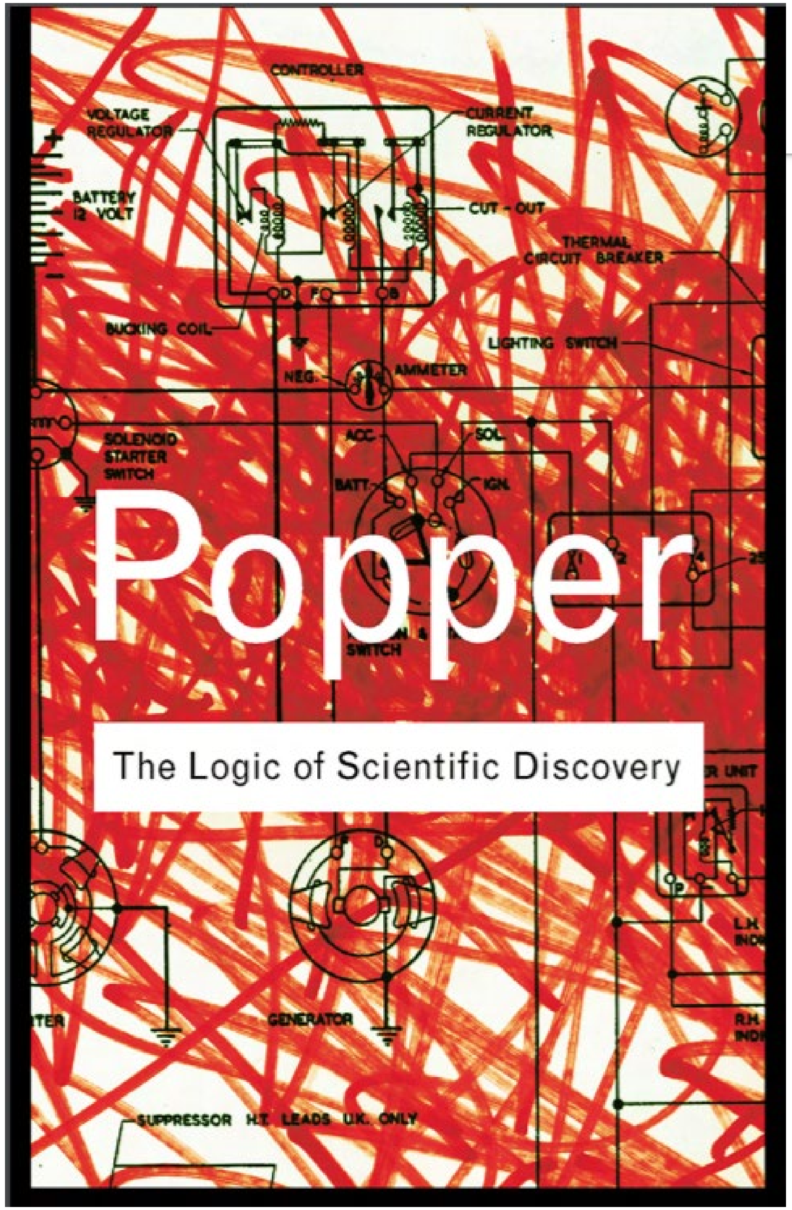
Beyond knowledge, experience

mountains beyond mountains





Beyond knowledge & beyond the grasp of artificial reasoning – Realm of Experience



But how is the system that represents our world of experience to be distinguished? The answer is: by the fact that it has been submitted to tests, and has stood up to tests. This means that it is to be distinguished by applying to it that deductive method which it is my aim to analyse, and to describe.

'Experience', on this view, appears as a distinctive *method* whereby one theoretical system may be distinguished from others; so that empirical science seems to be characterized not only by its logical form but, in addition, by its distinctive *method*. (This, of course, is also the view of the inductivists, who try to characterize empirical science by its use of the inductive method.)

The theory of knowledge, whose task is the analysis of the method or procedure peculiar to empirical science, may accordingly be described as a theory of the empirical method—*a theory of what is usually called 'experience'*.

http://strangebeautiful.com/other-texts/popper-logic-scientific-discovery.pdf

Elusive Quest for Knowledge
Advanced integration of information, data, decisions

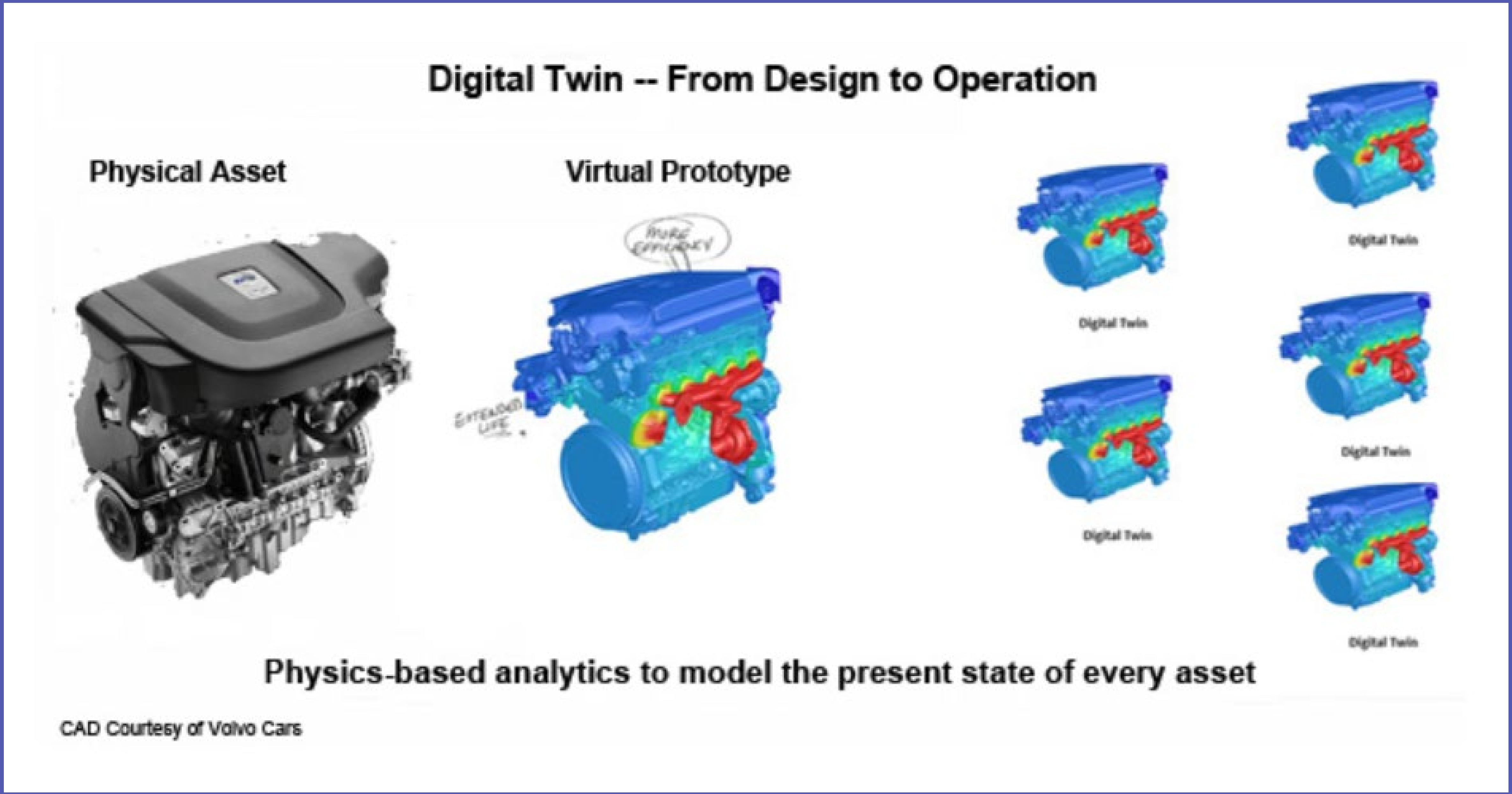
- KIDS may need TWINS

- KIDS may include SARA

KIDS need TWINS

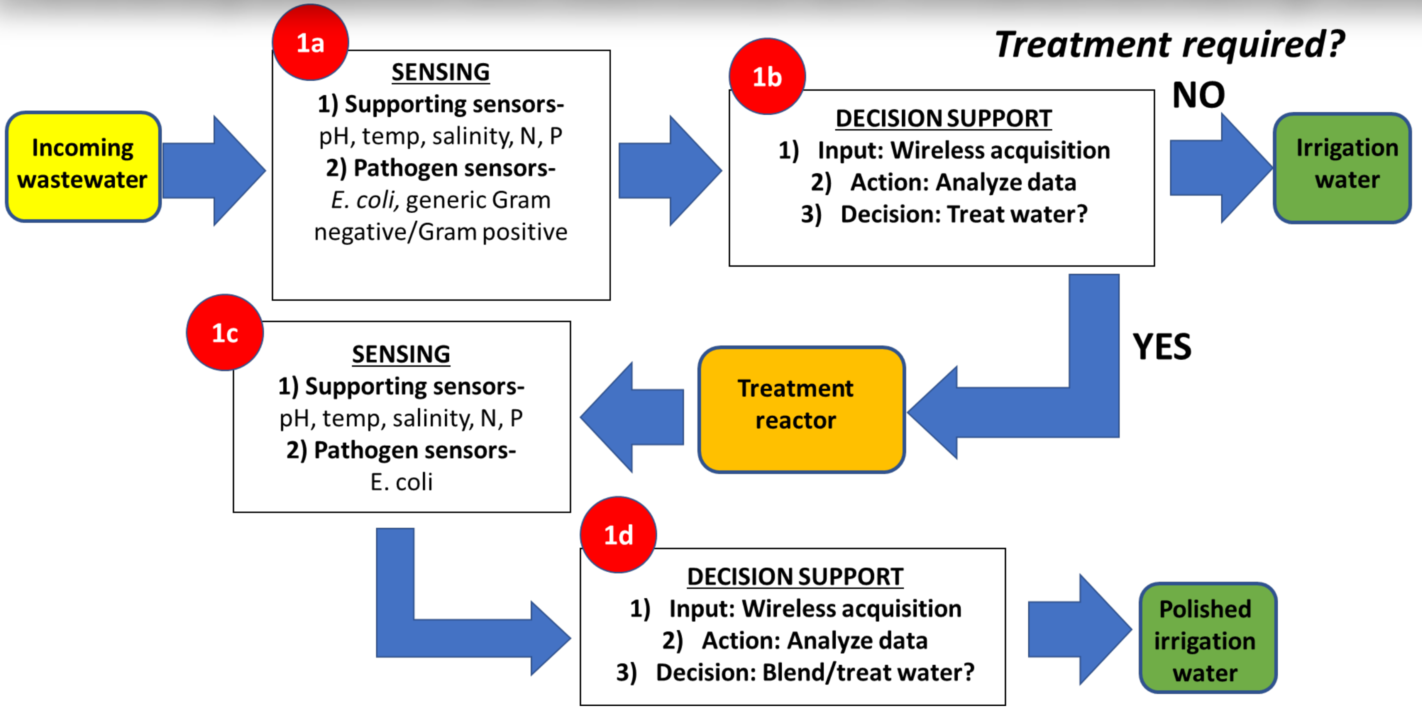
Adding PLM capacity in the form of Digital Twins to DIDA'S and KIDS

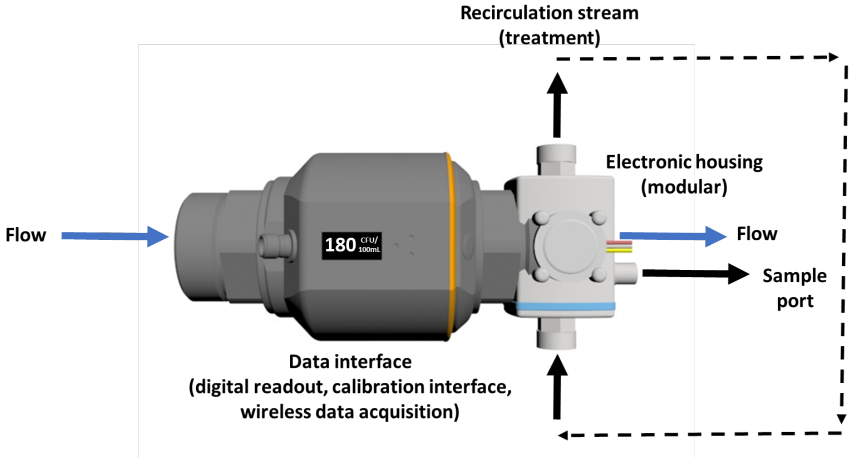
PLM sensor data – machine tools, pumps, pipelines, cold storage

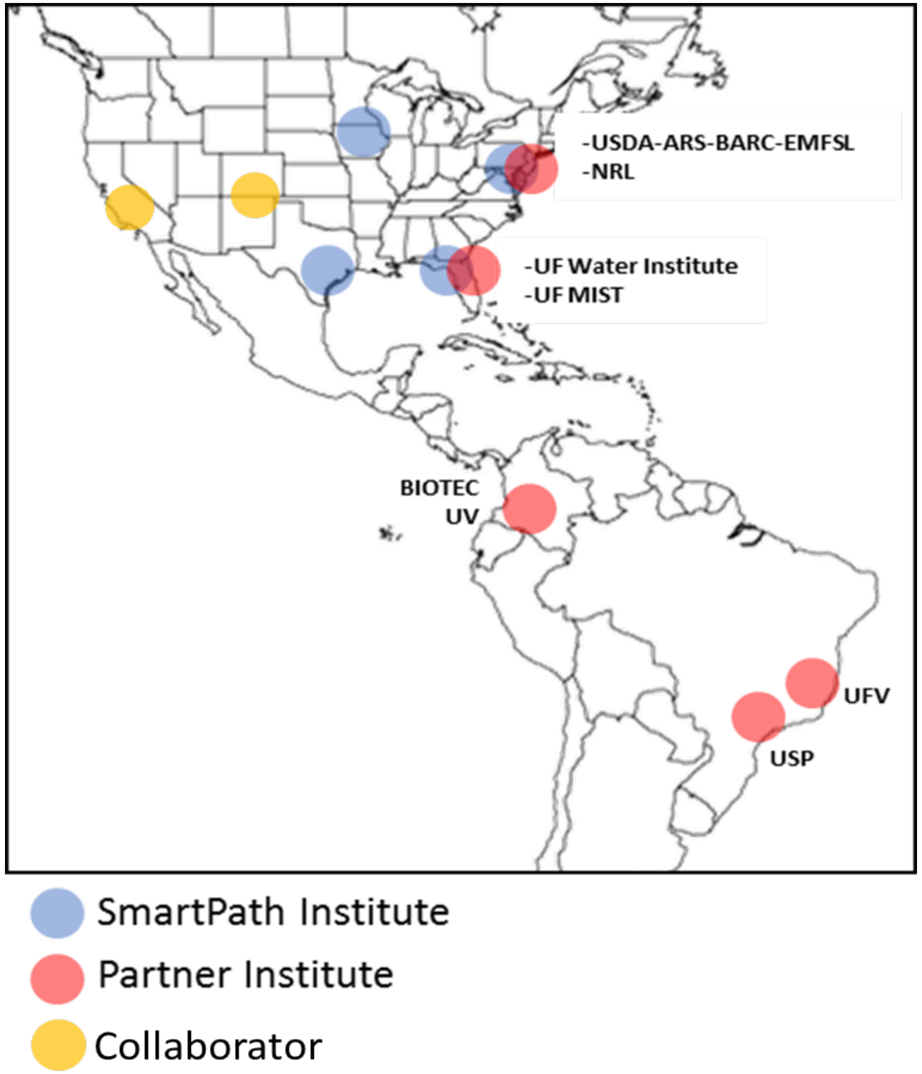


https://www.ansys.com/blog/digital-twin-pump

Use case: KIDS integrated with DIGITAL TWINS may improve the ecosystem in terms of machinery lifecycle in the agroecosystem





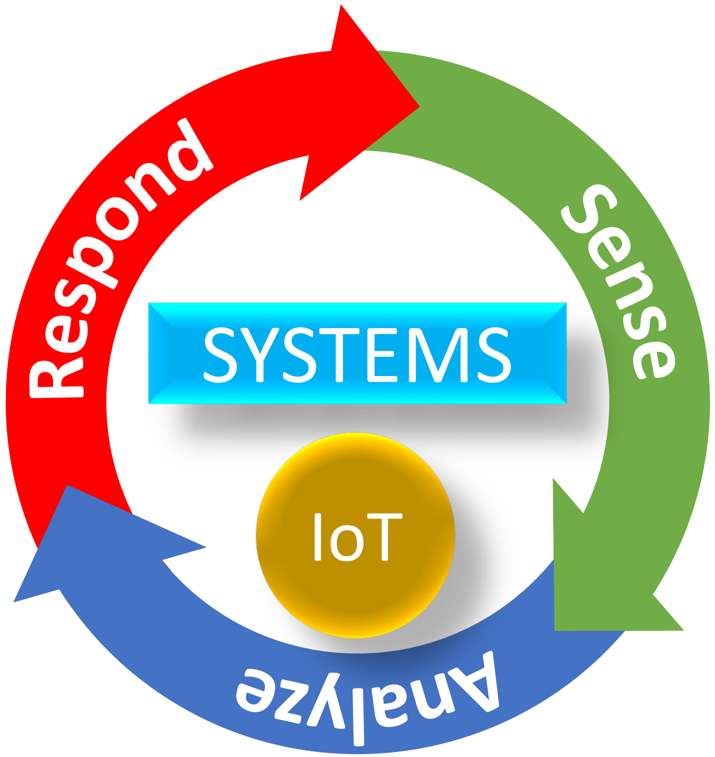


KIDS to include SARA

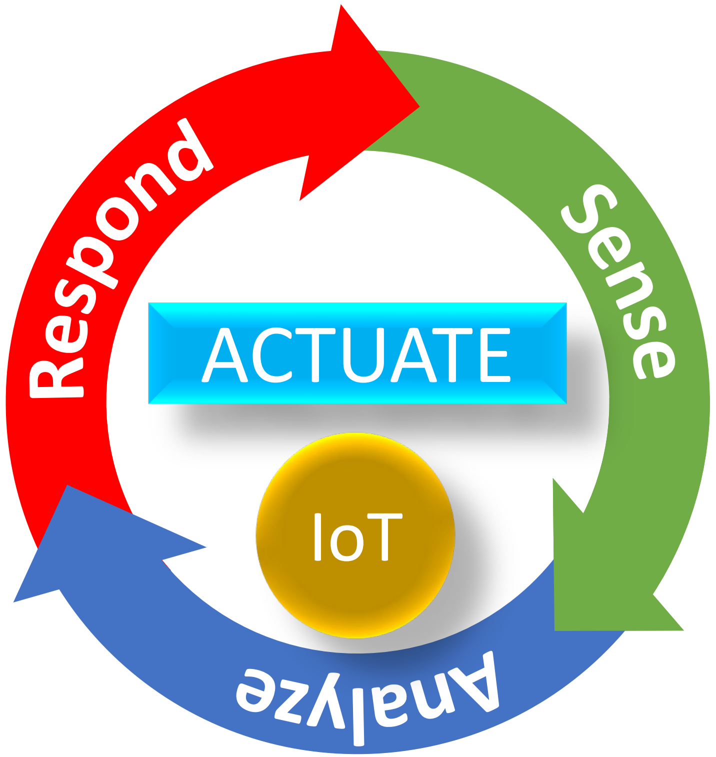
From sense, analyze, response, systems (SARS)

to sense, analyze, response, actuate (SARA)

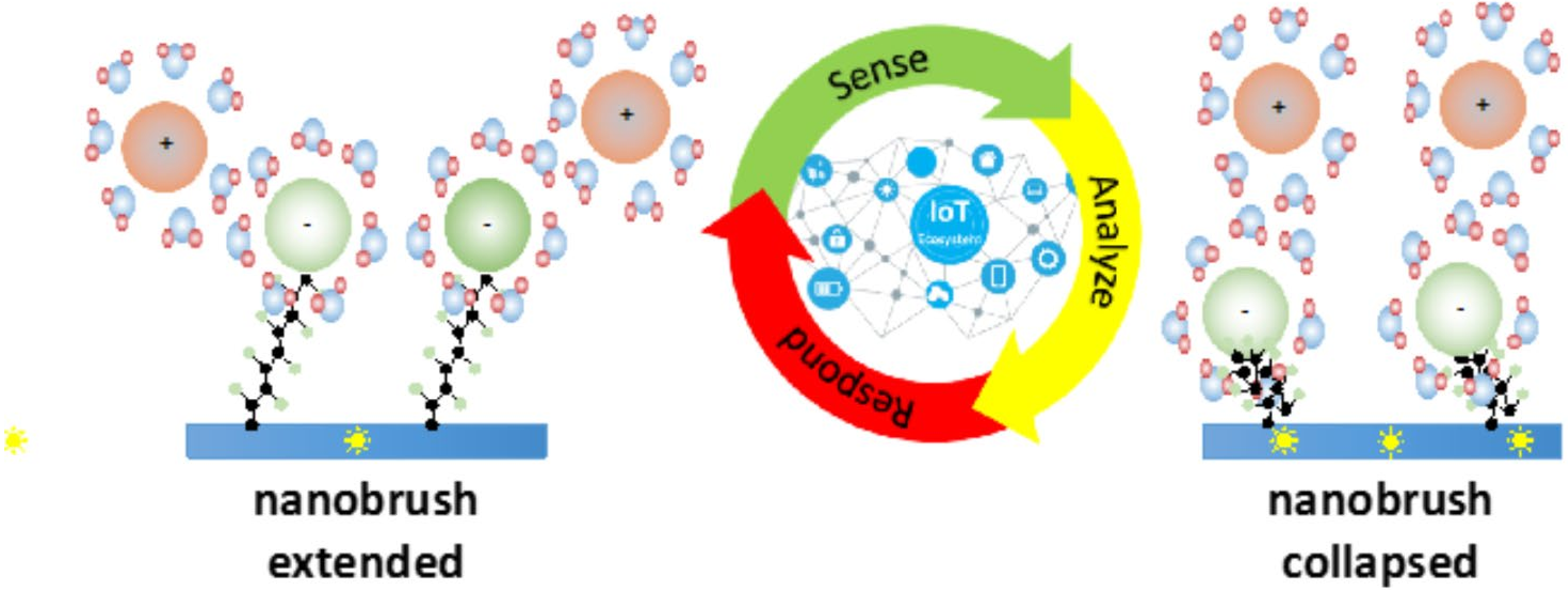
DIDA’S includes Sense, Analyze, Response, Systems (SARS)



KIDS to include Sense, Analyze, Response, Actuate (SARA)



Atoms to Bits - Sense, Analyze, Response, Actuate (SARA) Sensors using Smartphone



Nanomaterials can be monitored and controlled with smartphone

Future of digital transformation for the agro-ecosystem and emergence of digital products for traditional agri-businesses.

https://emclamor.wixsite.com/mclamorelab

Conventional Wisdom Questions Growth from Digital Transformation:

Conventional Wisdom Questions Growth from Digital Transformation:

NASA is opening the space station to $35,000-a-night visits. A tourist who paid Russia $30 million to get there a decade ago says it's a 'seismic shift.'

Dave Mosher 2h

NASA plans to open up the International Space Station to tourists and, to an even larger degree, commercial activity. NASA; Alyssa Powell/Business Insider

Conventional Wisdom ?

BUSINESS INSIDER

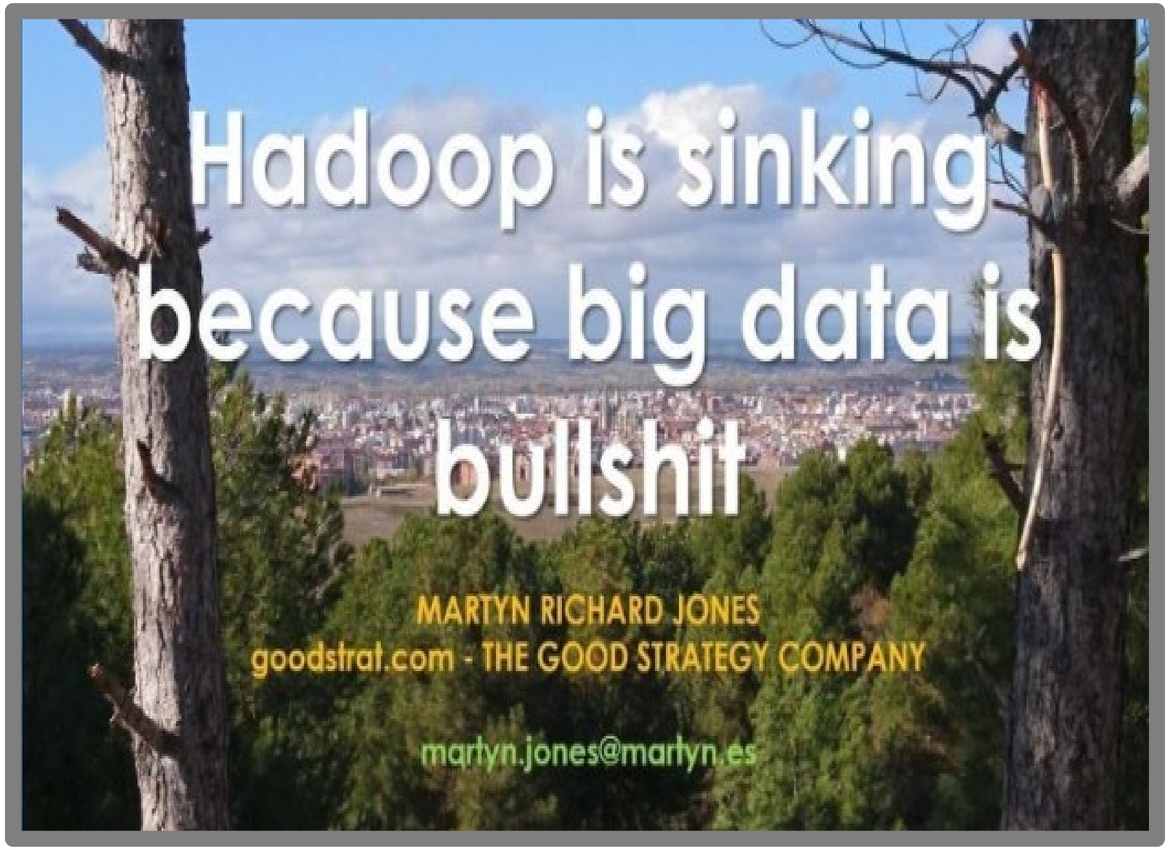
TECH | FINANCE | POLITICS | STRATEGY | LIFE | ALL

BI PRIME | INTELLIGENCE

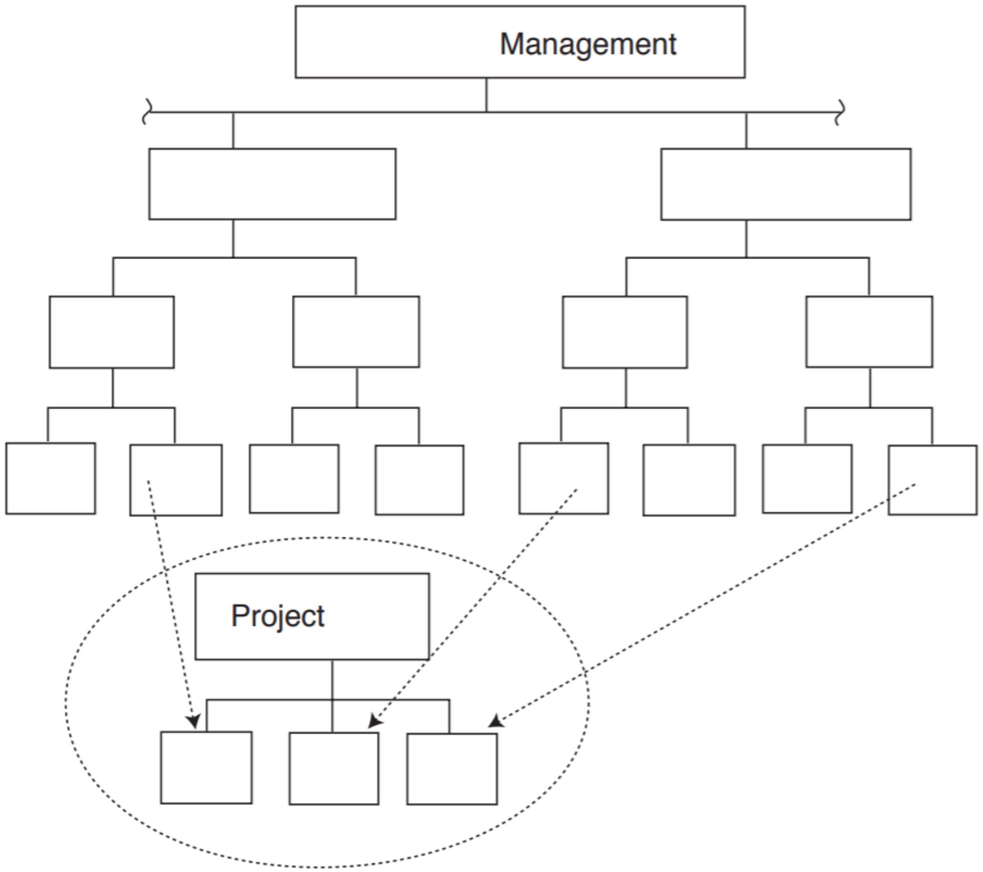
NASA is opening the space station to $35,000-a-night visits. A tourist who paid Russia $30 million to get there a decade ago says it's a 'seismic shift.'

Dave Mosher 2h

NASA plans to open up the International Space Station to tourists and, to an even larger degree, commercial activity. NASA; Alyssa Powell/Business Insider

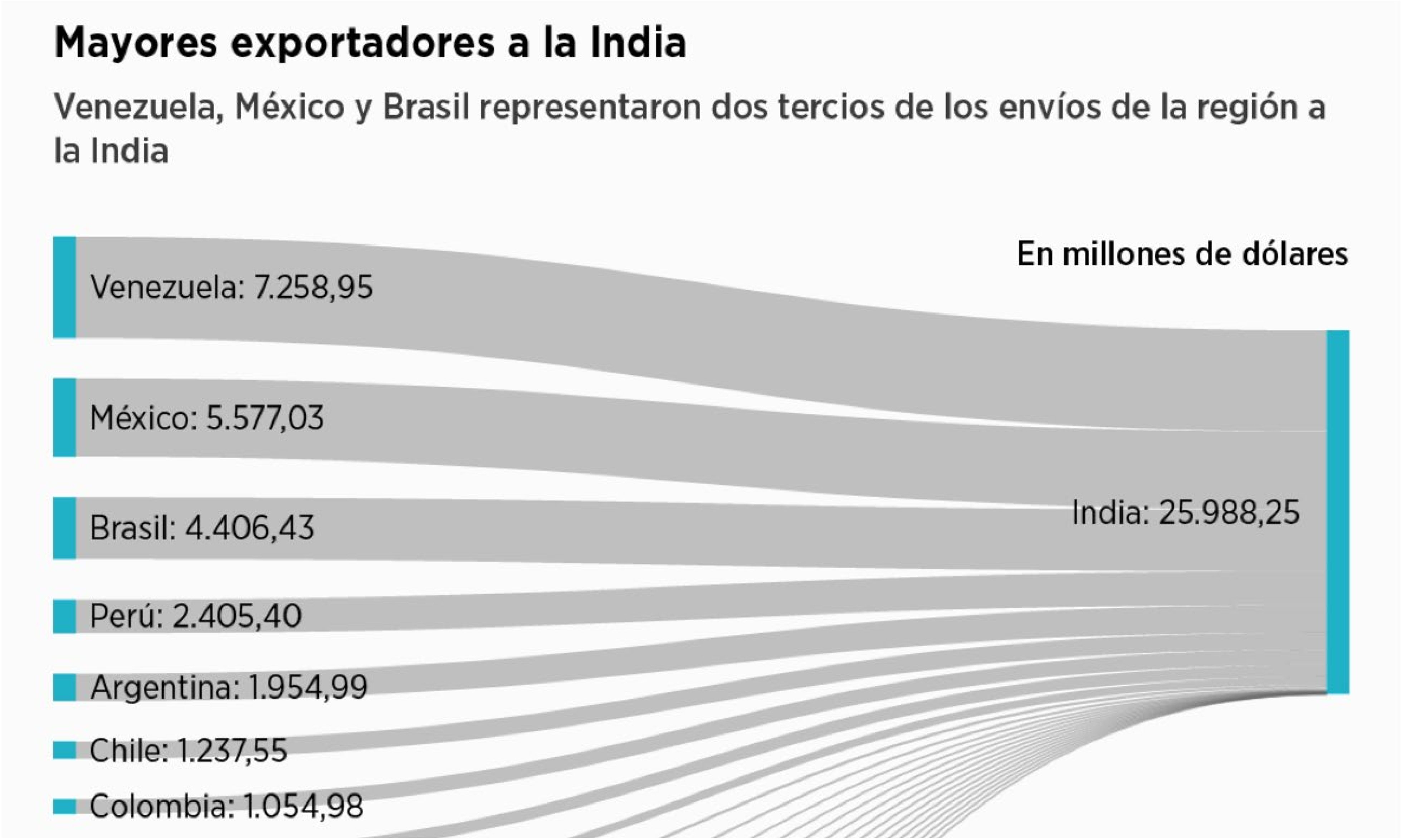


CONVENTIONAL WISDOM



Think different. Think non-linear. Think outside the box. Think beyond boundaries.

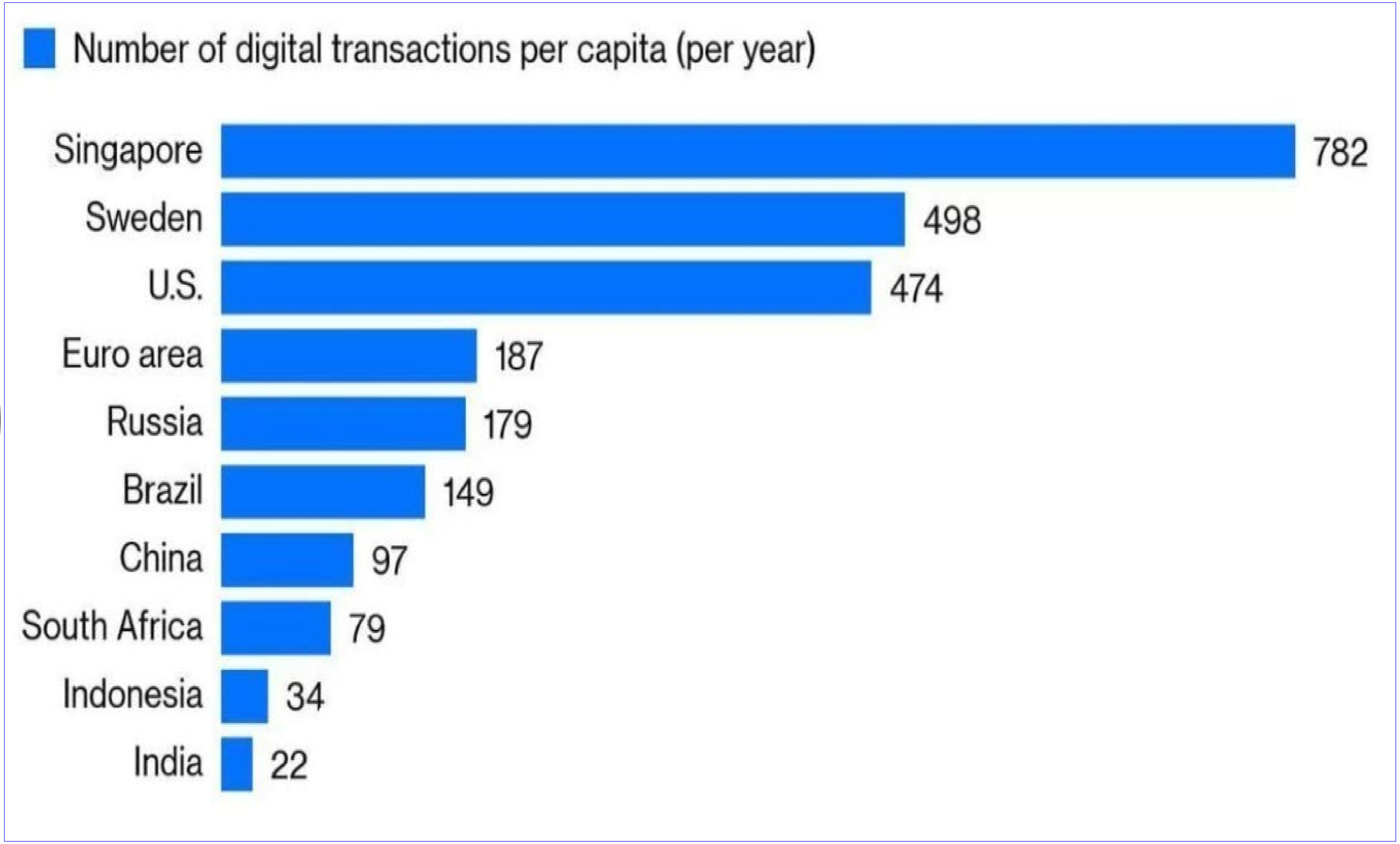
Exports from Latin America to India was US$2 billion in 2000. In 2018, it exceeded US$25 billion.



https://www.iadb.org/en/improvinglives/can-india-become-latin-americas-next-trade-frontier

Is there a market for digital ART products in traditional agri-business?

To obtain the volume of potential digital transactions, multiply the number shown with the population of the country.



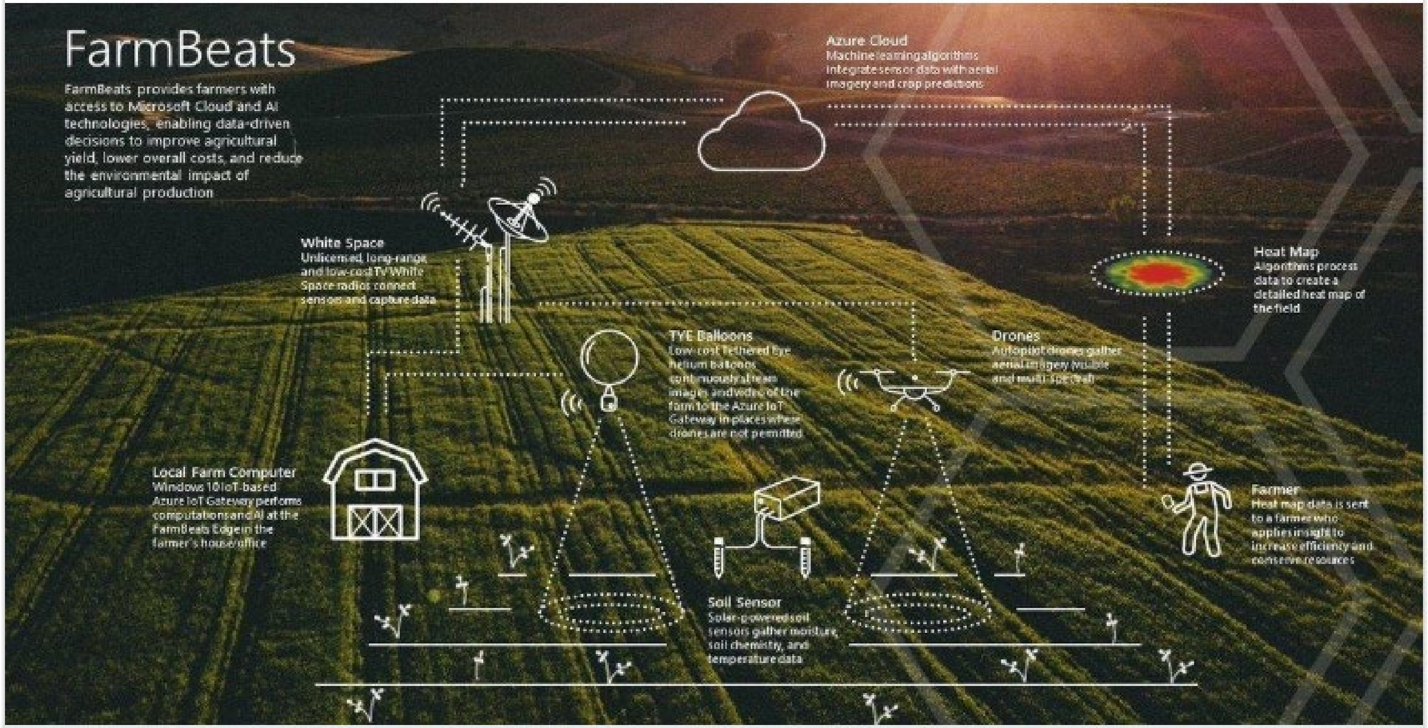
https://www.washingtonpost.com/business/india-going-cashless-could-be-a-model-for-the-world/2019/06/05/d8fec830-87ee-11e9-9d73-e2ba6bbf1b9b_story.html

http://bit.ly/PARTNER-WITH-PEAS

Ranveer Chandra
Chief Scientist, Azure Global at Microsoft

Tom Keane
Corporate Vice President of Azure Global - Microsoft Azure

Microsoft FarmBeats program uses Azure to connect agricultural devices and generate data intended to help farms transform business.



http://bit.ly/SUBSCRIBE-TO-SENSEE

http://bit.ly/PARTNER-WITH-PEAS

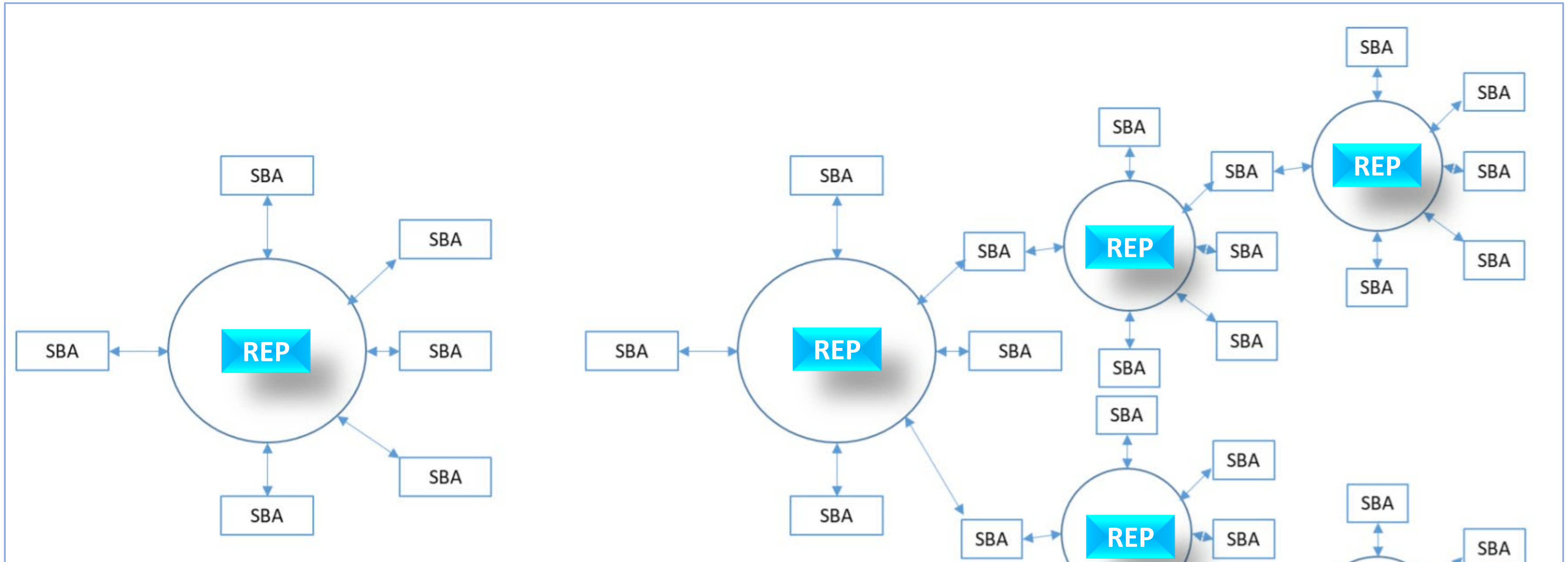
http://bit.ly/SUBSCRIBE-TO-SENSEE

Low hanging fruits in digital-agro

Prevention of food waste, global public health, water and food safety

Minimalistic approach combining SENSEE in a basic DIDA'S platform tasked with simple information arbitrage on mobile devices.

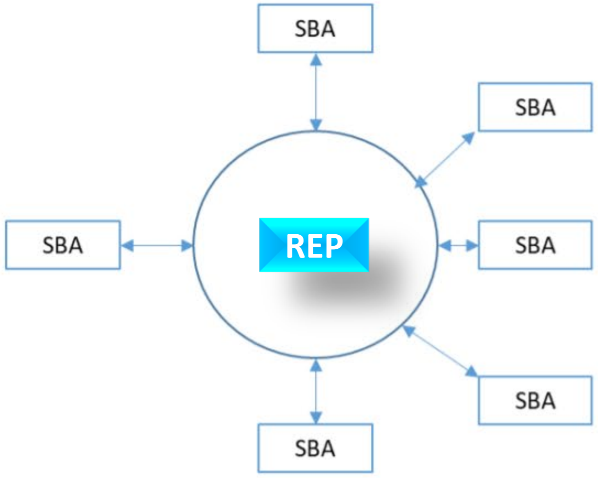
Intelligent Information Arbitrage

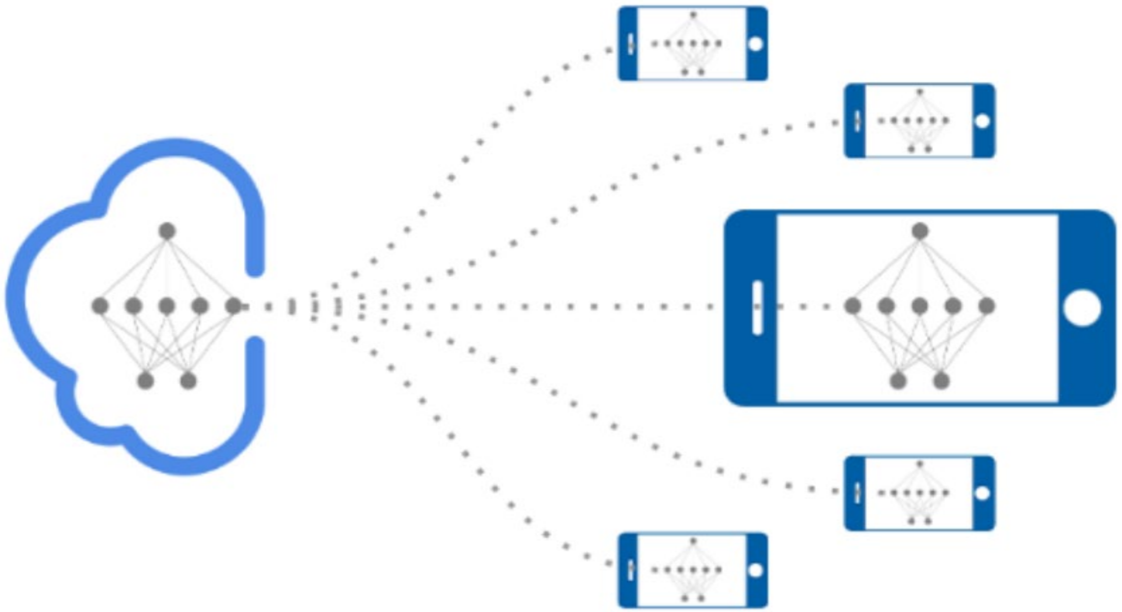


REP – Repository (Regional Data)
example: Sensor DB “SENSEE”
Sensor search engine (SENSEE)
SBA – “SERVICE BOT” as Agents

Locally distributed repositories [REP] containing data and questions (relevant to users and growers by crop or environment) can be globally connected using an Agent-based system

Intelligent Information Arbitrage

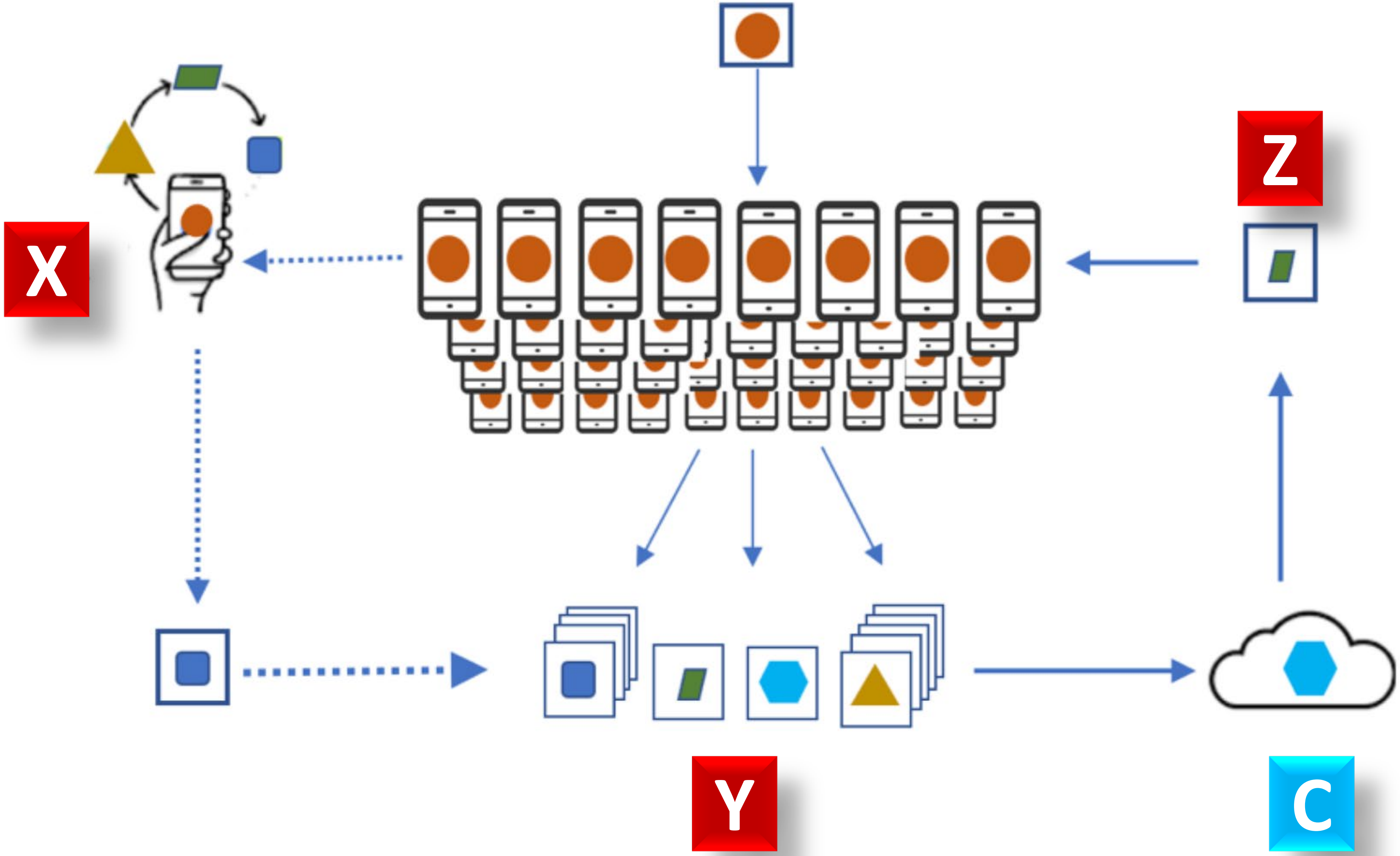




REP – Repository (Regional Data)
example: Sensor DB "SENSEE"
Sensor search engine (SENSEE)
SBA – "SERVICE BOT" as Agents

Locally distributed repositories [REP] containing data and questions (relevant to users and growers by crop or environment) can be globally connected and consumed by any system

Intelligent Information Arbitrage



https://medium.com/@SantanuB/a-beginners-guide-to-federated-learning-b29e29ba65cf

Intelligent Information Arbitrage

Z

X

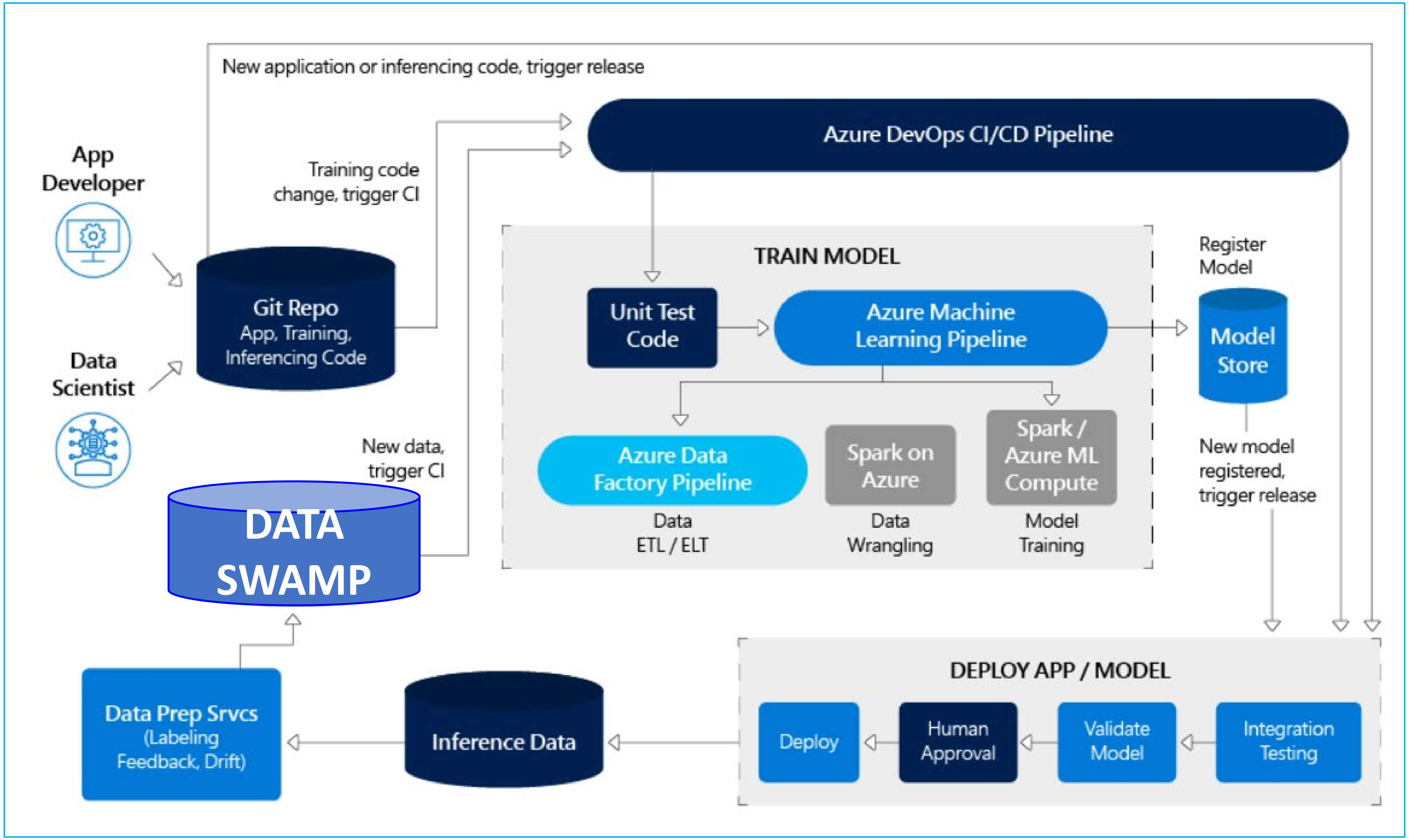
The value of the REP concept (SENSEE) may be enhanced by coupling publish-subscribe modes with crowd-sourced data adoption/dissemination. User "X" may update data, recommend tools or techniques or share outcomes/outputs (for example, growers can share photographs of infected produce or sinfully delicious tomatoes). Thus, local user personalization (point X, in the crowd) is sent/stored to the analytical platform (engine Y). An emerging consensus from contributed data (for example, improved technique or data with incorrect units or better use of a tool) is sent to cloud C for expert evaluation and critical analysis. Verified change Z is communicated to all subscribers, globally. This process repeats, to enhance open models and enrich common goals for public goods, using distributed data from crowdsourcing (users, farmers, growers, scientists, engineers, academics, politicians) but deploying a neutral/trusted analytical evaluator (cloud C) to deconstruct/reconstruct, aggregate/disaggregate data and models, to serve the best interest of the system. It may prevent data pollution, act to neutralize cyberthreats and stop, if possible, attacks perpetrated by GAN (general adversarial network) as infectious agents. This suggestion draws from "federated learning models" commonly used by financial institutions and banks to train fraud detection models without sharing their sensitive customer data. Popular frameworks now include TensorFlow Federated, an open source framework by Google for experimenting on decentralized data. PySyft is a open source library that is built on top of PyTorch for encrypted, privacy preserving deep learning. Federated AI Technology Enabler (FATE) is an open-source project initiated by Webank's AI group to provide a secure computing framework to support the Federated AI ecosystem. Despite the hype whipped up by the glib snake oil salesmen of AI, there is value in this approach, if and when rationally analyzed, for specific purposes, using bonafide tools, which may be customized for specific applications and are based on rigorous mathematics and statistics. It may be useful for SITS and its ecosystem to explore these advanced tools of the future and enterprise solutions around federated learning and other secure computation techniques across different verticals. At present, the primary deployment challenge may be the computational constraint of edge devices (smartphone, tablet) to perform local training, cloud consultations and inferencing. However, smartphones and IoT devices are increasingly equipped with GPUs or sufficient computing hardware to run CNN/RNN and other AI models at the edge to augment near-real time "intelligent" decision support systems, at the point of use. REP/SENSEE may be the SmartPath/SITS approach to harvest these ideas and convert them into actionable transactions that can help the ag industry in the pursuit of food.

Y

C

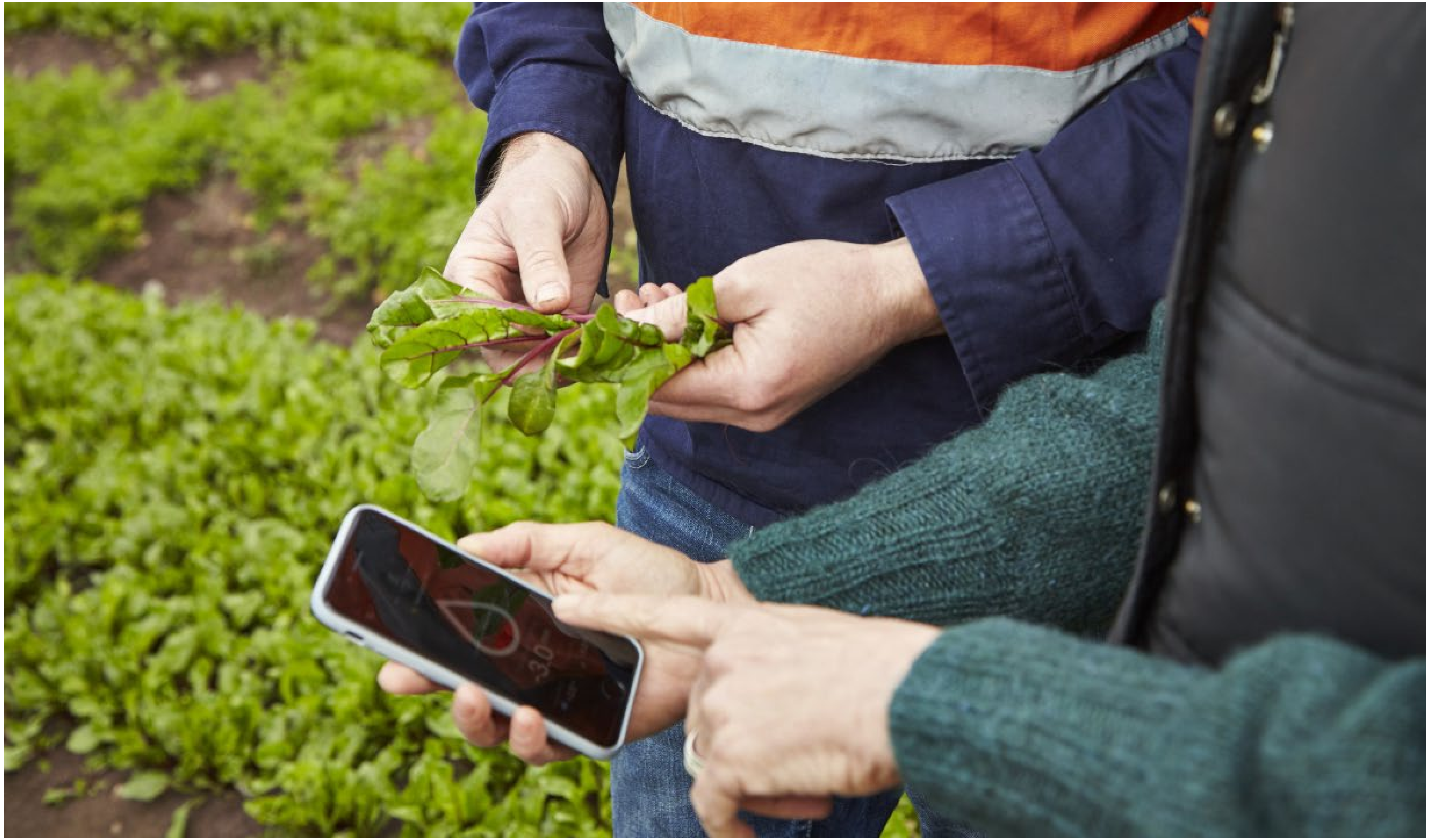
https://medium.com/@SantanuB/a-beginners-guide-to-federated-learning-b29e29ba65cf

Intelligent Information Arbitrage



https://azure.microsoft.com/en-us/blog/breaking-the-wall-between-data-scientists-and-app-developers-with-azure-devops/

Intelligent Information Arbitrage

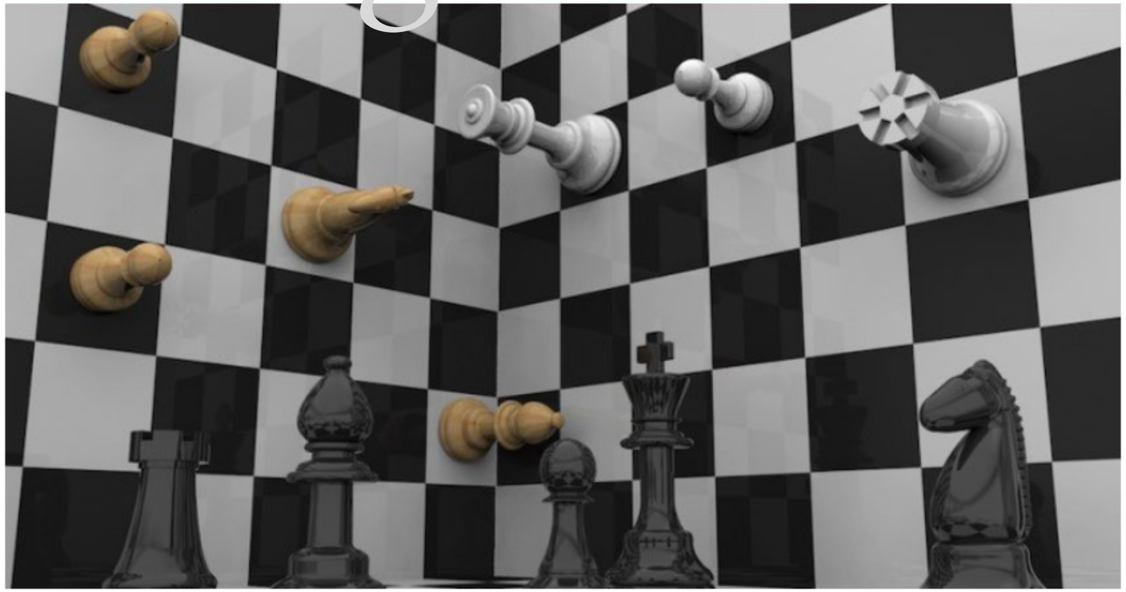


HARVEST and INVEST in UNSTRUCTURED DATA and (MERGE DATA IN CONTEXT FOR INTELLIGENT) DATA ANALYTICS

Intelligent Information Arbitrage

WATER v PEOPLE

Change the rules



SENSEE inside

https://www.who.int/news-room/detail/18-06-2019-1-in-3-people-globally-do-not-have-access-to-safe-drinking-water-–-unicef-who

1 in 3 people globally do not have access to safe drinking water – UNICEF, WHO

New report on inequalities in access to water, sanitation and hygiene also reveals more than half of the world does not have access to safe sanitation services.

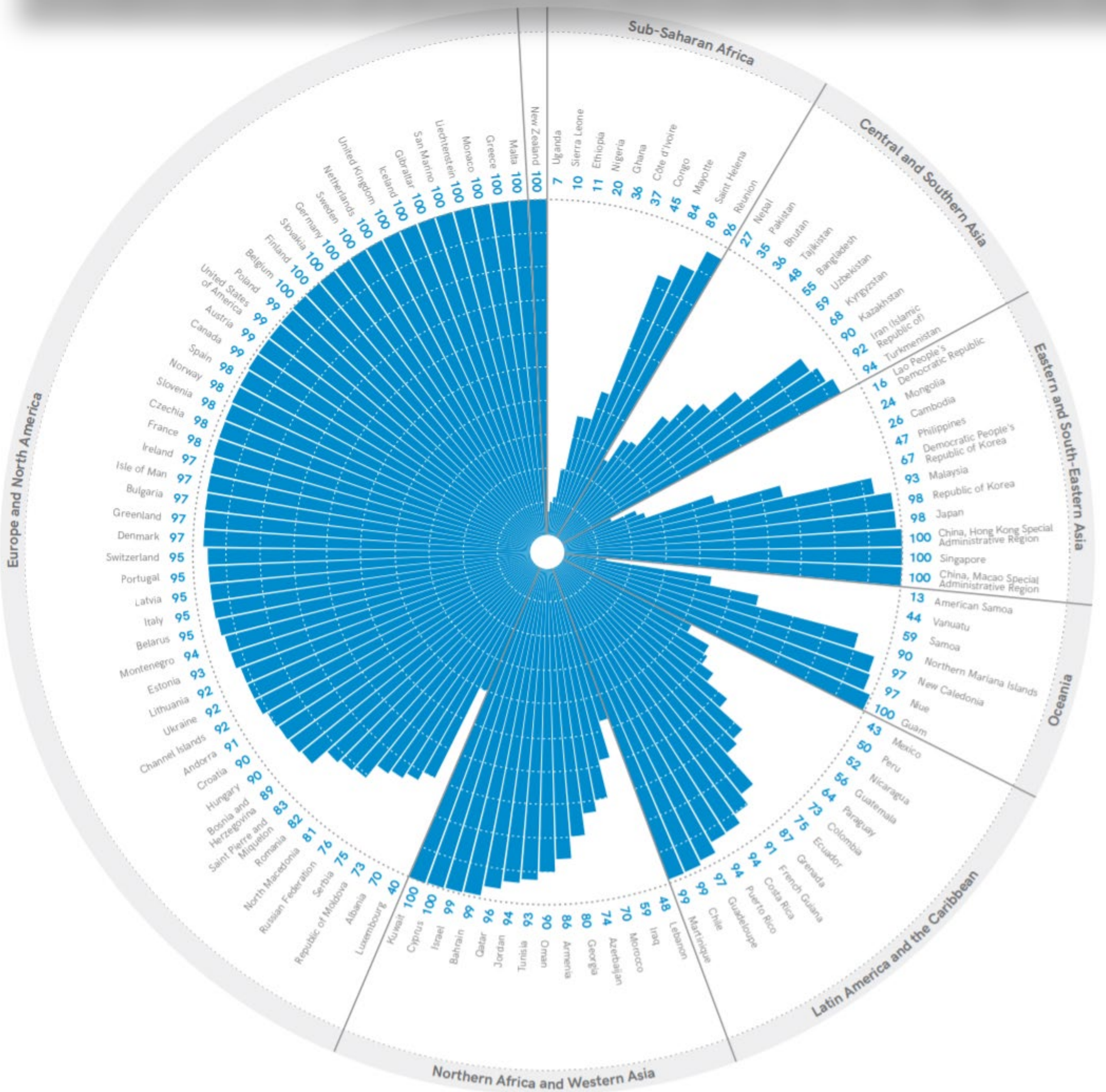
18 June 2019 | News release | New York, Geneva

Billions of people around the world are continuing to suffer from poor access to water, sanitation and hygiene, according to a new report by UNICEF and the World Health Organization. Some 2.2 billion people around the world do not have safely managed* drinking water services, 4.2 billion people do not have safely managed sanitation services, and 3 billion lack basic** handwashing facilities.

Nada Osseiran
Communications Officer
WHO
Mobile: +41 79 445 1624

www.who.int/water_sanitation_health/publications/jmp-2019-full-report.pdf?ua=1

Proportion of population with safely managed water (left) and sanitation (right) services, 2017 (%)



(L) Fig 51 (water) and (R) Fig 69 (sanitation) from www.who.int/water_sanitation_health/publications/jmp-2019-full-report.pdf?ua=1


gettyimages
Gina Pricope

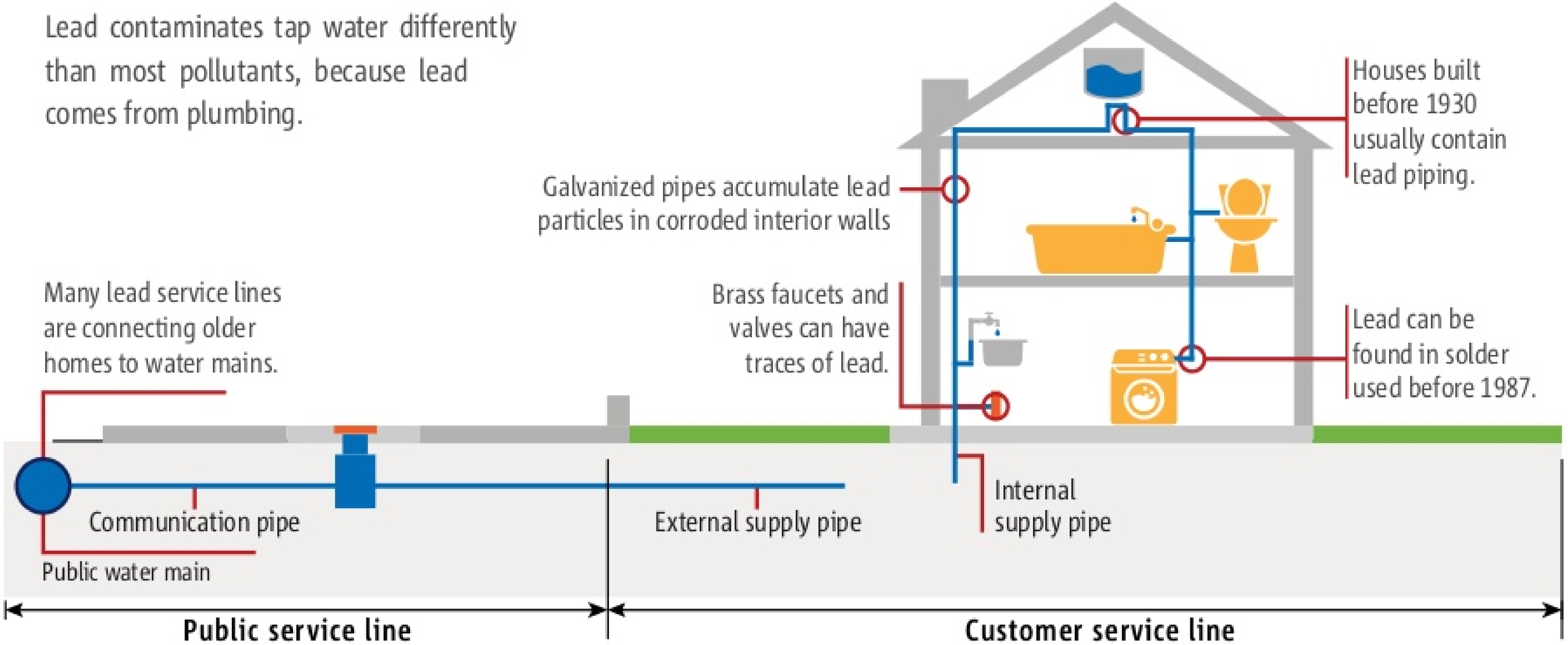
Pb

HHS Public Access
Author manuscript
J Public Health Manag Pract. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:
J Public Health Manag Pract. 2019 ; 25(Suppl 1 LEAD POISONING PREVENTION): S84–S90. doi: 10.1097/PHH.0000000000000871.

The Flint Water Crisis: A Coordinated Public Health Emergency Response and Recovery Initiative

THE PIPES



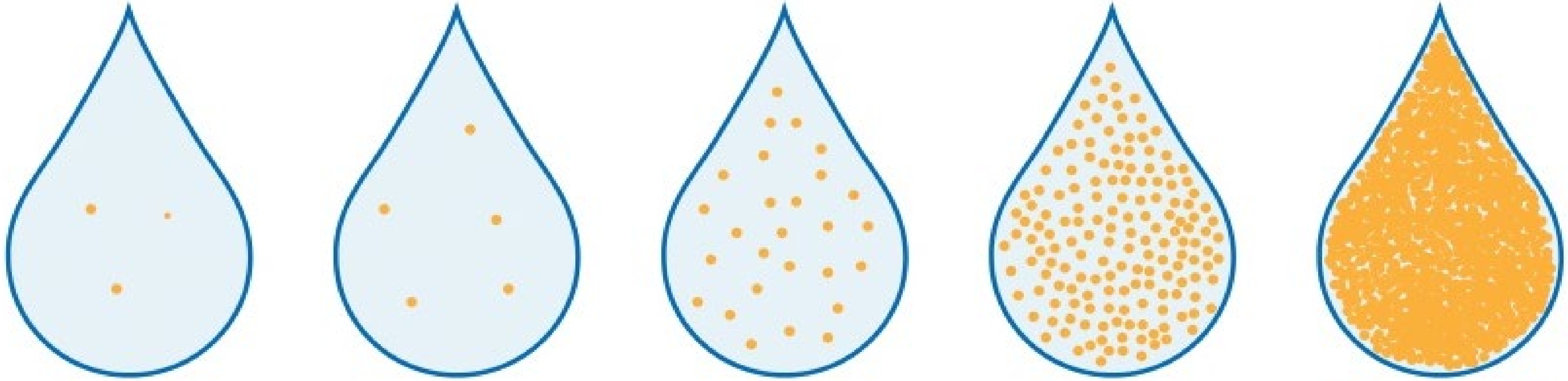
Summing up: water from Flint River moved through lead pipes, picking up the toxin as it went, and spread it throughout the population.

https://www.slideshare.net/lbuckfire/the-flint-michigan-water-crisis-causes-effects

FLINT, MICHIGAN WATER CRISIS - **LEAD POISONING FACTS AND FIGURES**

THE KEY PROBLEM

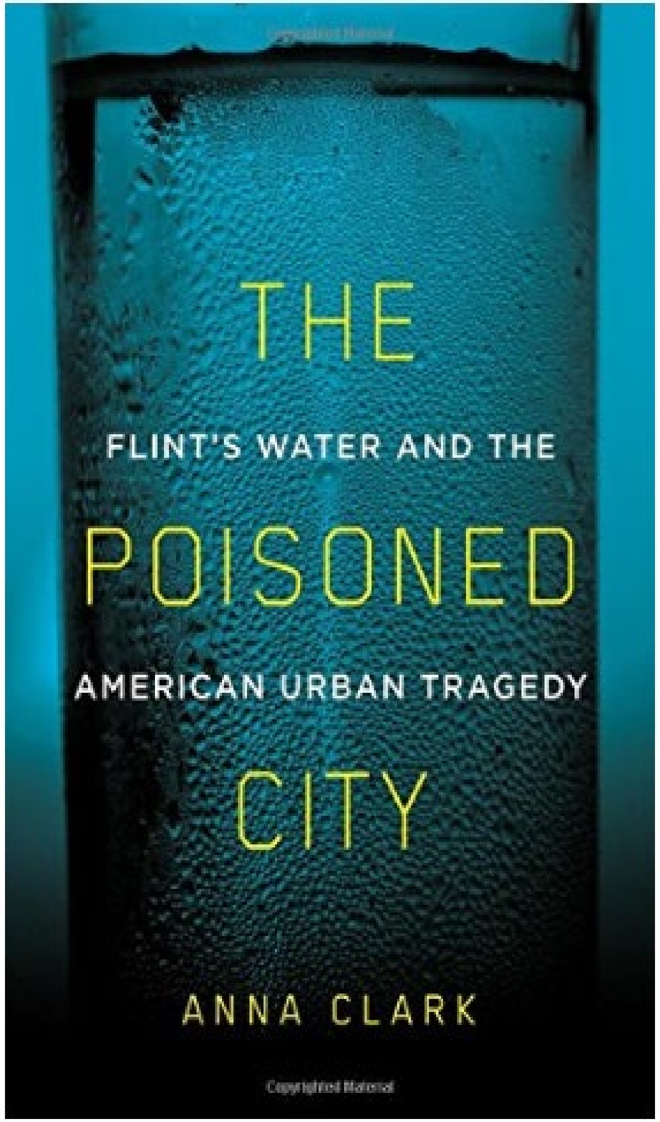
Water from the Flint River is highly corrosive (its water has about 8 times more chloride (Cl–) in it than Detroit water) to iron and lead. Unfortunately, these pipe materials are widely used throughout Flint.



2.3 parts per billion	5 parts per billion	27 parts per billion	158 parts per billion	13,000 parts per billion
Detroit water	Cause for concern	Flint water	Highest level found in Virginia Tech's Flint sample	Highest level found in Flint

And, if these pipes are exposed to corrosive water, or if water sits too long inside them, the lead could be released and may end up coming out of the tap.

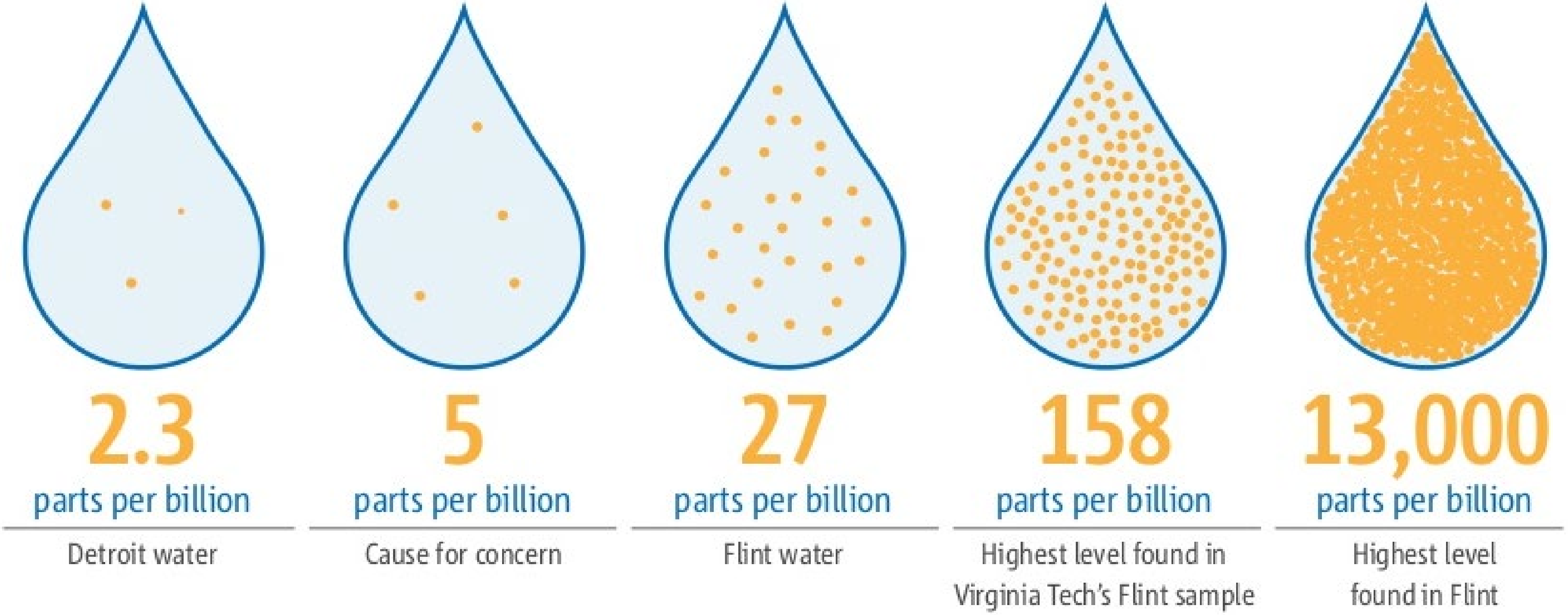
https://www.slideshare.net/lbuckfire/the-flint-michigan-water-crisis-causes-effects



FLINT, MICHIGAN WATER CRISIS - **LEAD POISONING FACTS AND FIGURES**

THE KEY PROBLEM

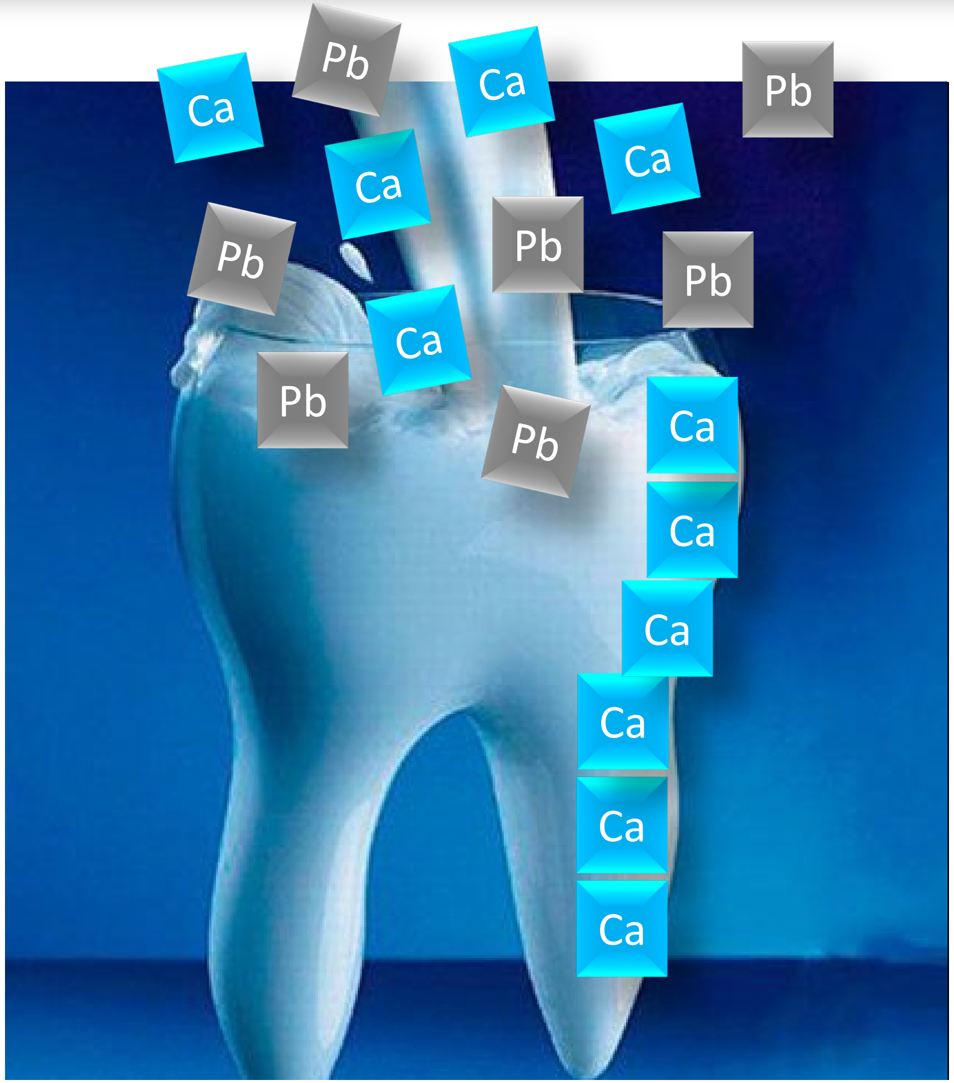
Water from the Flint River is highly corrosive (its water has about 8 times more chloride (Cl–) in it than Detroit water) to iron and lead. Unfortunately, these pipe materials are widely used throughout Flint.



And, if these pipes are exposed to corrosive water, or if water sits too long inside them, the lead could be released and may end up coming out of the tap.

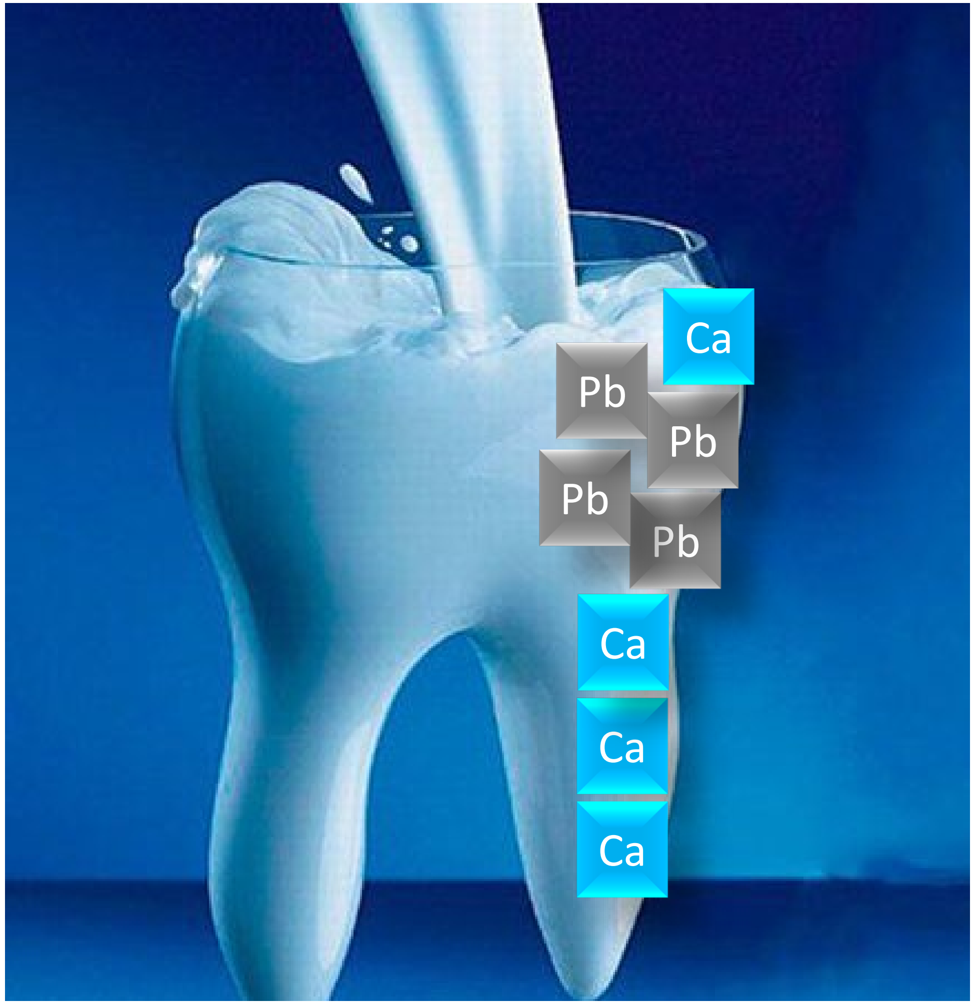
https://www.slideshare.net/lbuckfire/the-flint-michigan-water-crisis-causes-effects

Lead (Pb) accumulates in teeth and bones



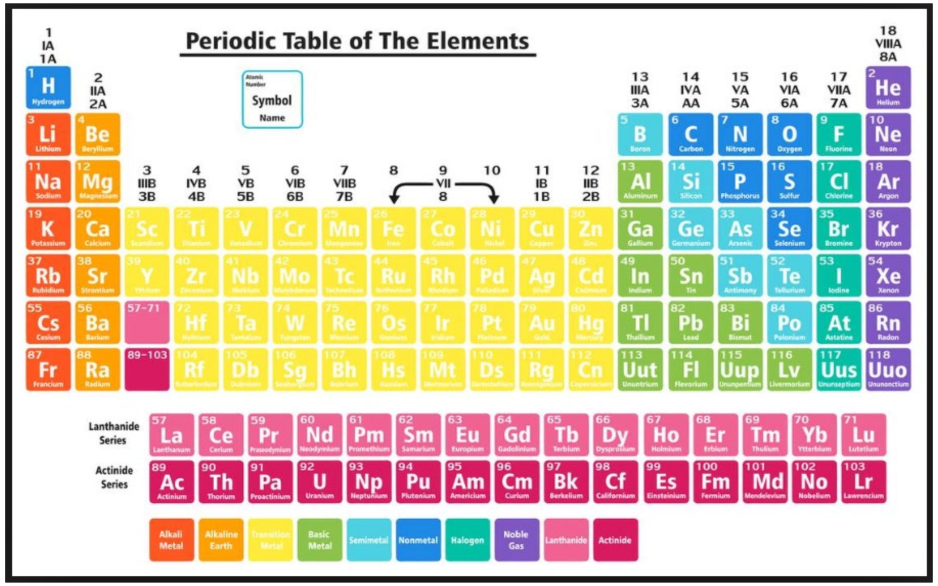
More lead exposure

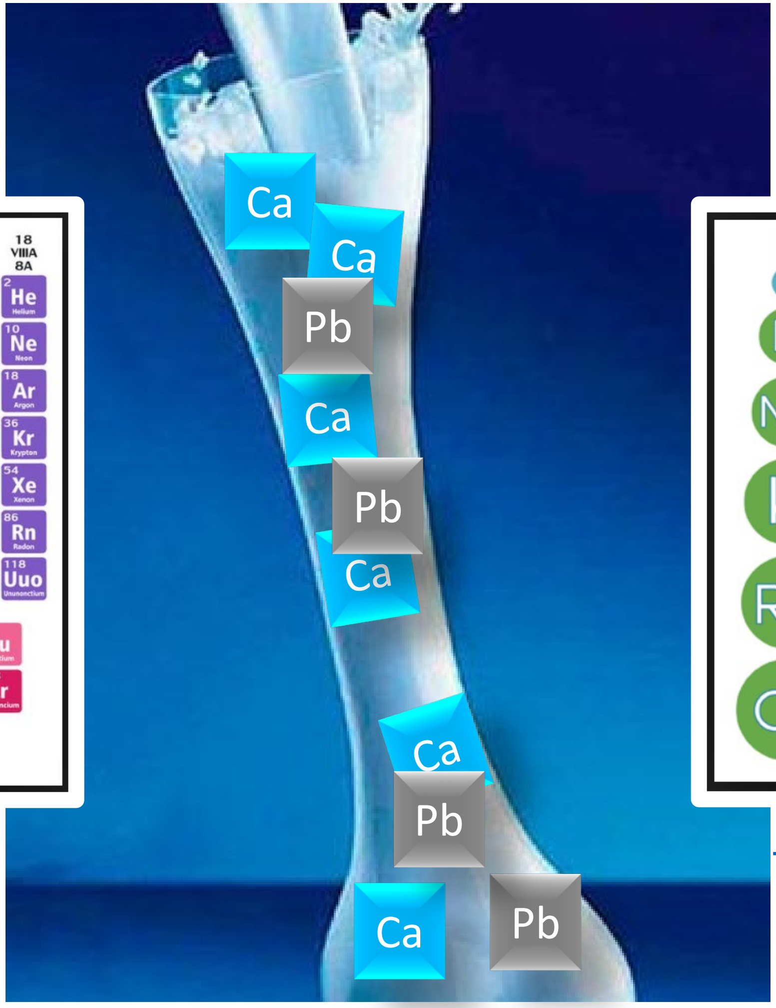
More lead is stored

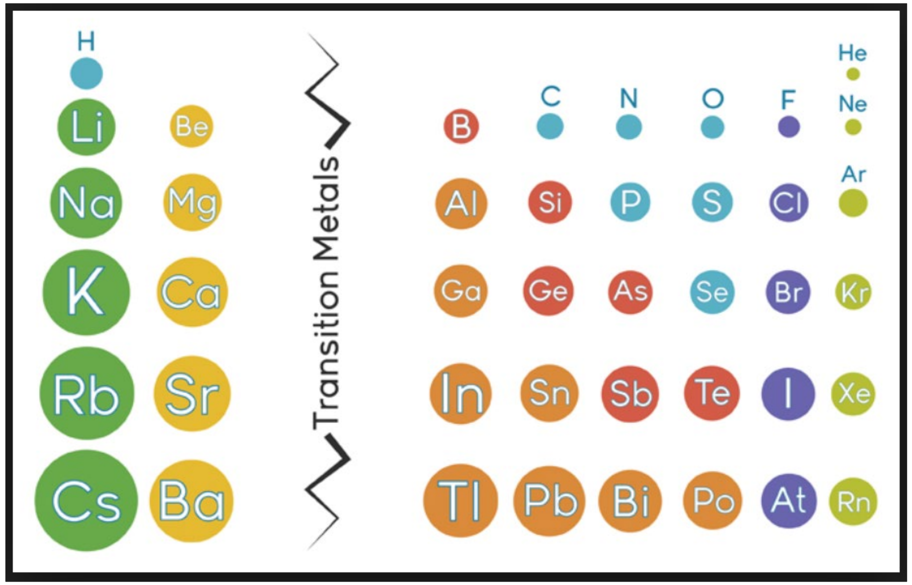


https://edu.glogster.com/glog/calcium/2ar2hn030do?=glogpedia-source

Lead (Pb) accumulates in teeth and bones



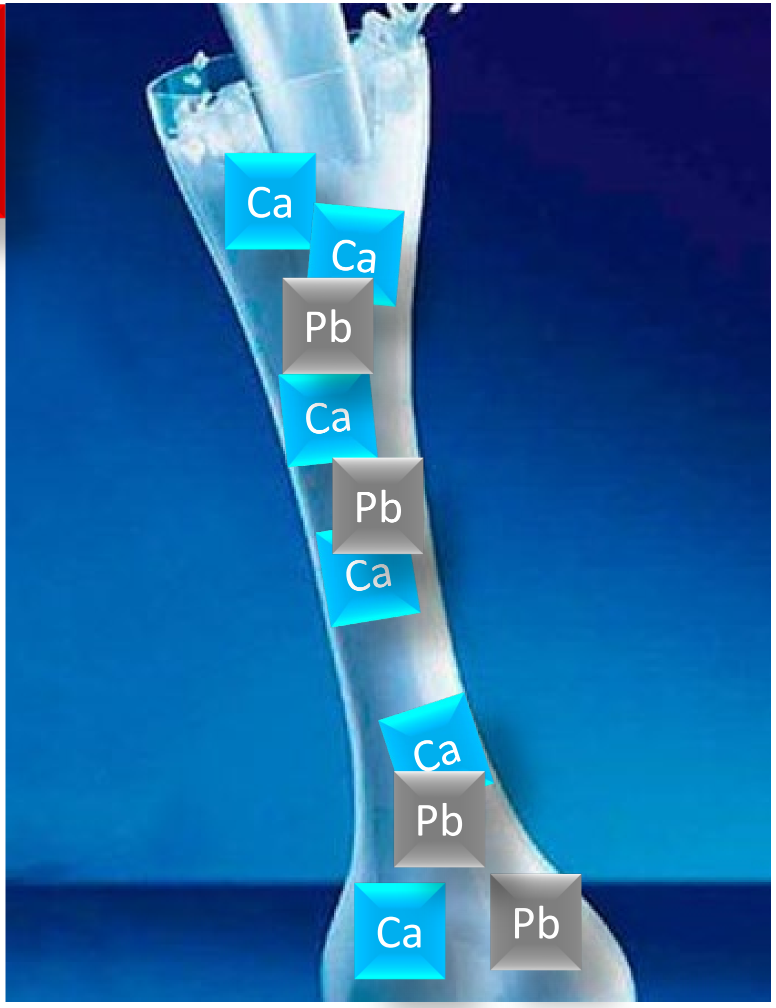




https://chemistryonline.guru/ionic-radius/

Lead (Pb) accumulates in teeth and bones

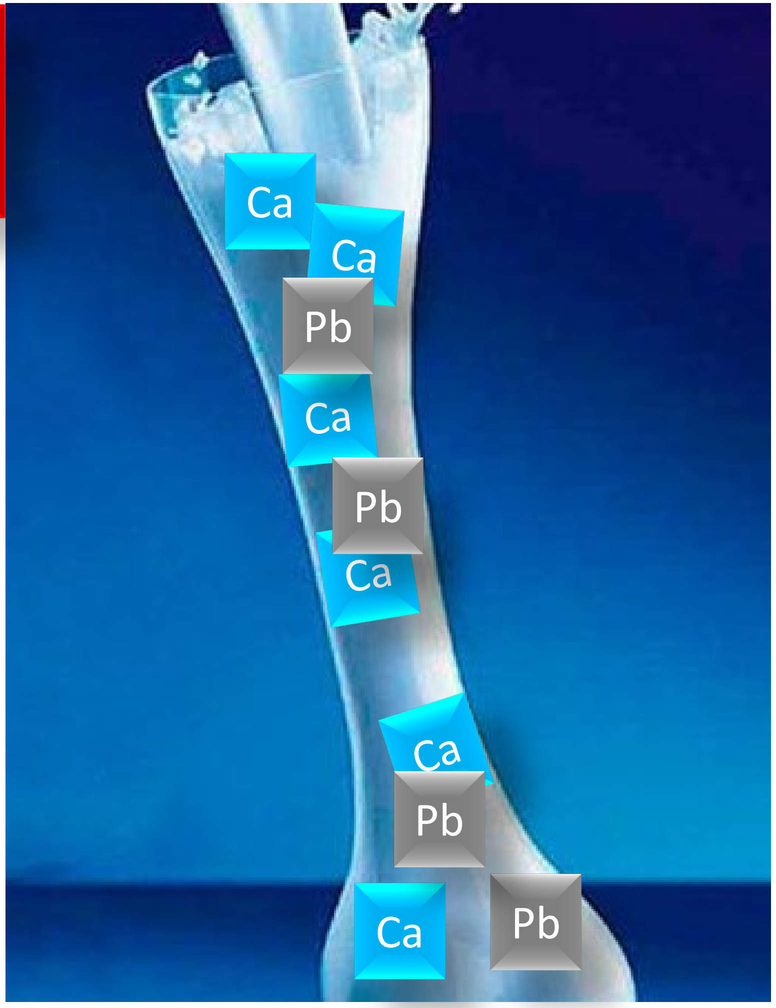
Osteoporosis



What happens to the **Pb** in the bones?

Lead (Pb) concentration increases in the blood

Osteoporosis



Pb Pb ? Pb Pb

What happens to the **Pb** in the bones?

Childhood Lead Poisoning causes dementia in adults

Pb

Environmental Health Perspectives
https://ehp.niehs.nih.gov/doi/10.1289/ehp.9716

The Association between Blood Lead Levels and Osteoporosis Adults—Results from the Third National Health and Nutrition Examination Survey (NHANES III)

James R. Campbell and Peggy Auinger

Published: 1 July 2007 | https://doi.org/10.1289/ehp.9716 | Cited by: 35

78% of the U.S. population (1970s) had blood lead levels ≥ 10 µg/dL. Bone is a repository for 90–95% of the total body burden of lead and harbors it for years after initial exposure. Thus, a high proportion of adult Americans may currently have elevated bone lead levels. With this many who were exposed to lead when younger, and the morbidity associated with osteoporosis, it is important to investigate whether an association exists between lead exposure and osteoporosis in humans. Our objective was to conduct a secondary analysis to explore an association between lead exposure and osteoporosis in U.S. adults.

Pb from water accumulates in teeth and bones. When Pb leaches out of bones, it may contribute to osteoporosis in adult life. Increased amount of Pb in blood may also contribute to dementia, Alzheimer's and neurotoxicity.

www.ncbi.nlm.nih.gov/pmc/articles/PMC3567843/pdf/nihms367232.pdf

NIH Public Access
Author Manuscript
Curr Alzheimer Res. Author manuscript; available in PMC 2013 February 08.

Published in final edited form as:
Curr Alzheimer Res. 2012 June ; 9(5): 563–573.

Alzheimer's Disease and Environmental Exposure to Lead: The Epidemiologic Evidence and Potential Role of Epigenetics

Kelly M. Bakulski[1], **Laura S. Rozek**[1,2], **Dana C. Dolinoy**[1], **Henry L. Paulson**[3], and **Howard Hu**[1,4,5,*]

[1]University of Michigan, School of Public Health, Department of Environmental Health Sciences

[2]University of Michigan, Medical School, Department of Otolaryngology

[3]University of Michigan, Department of Neurology

[4]University of Michigan, Department of Epidemiology

[5]University of Michigan, Medical School, Department of Internal Medicine

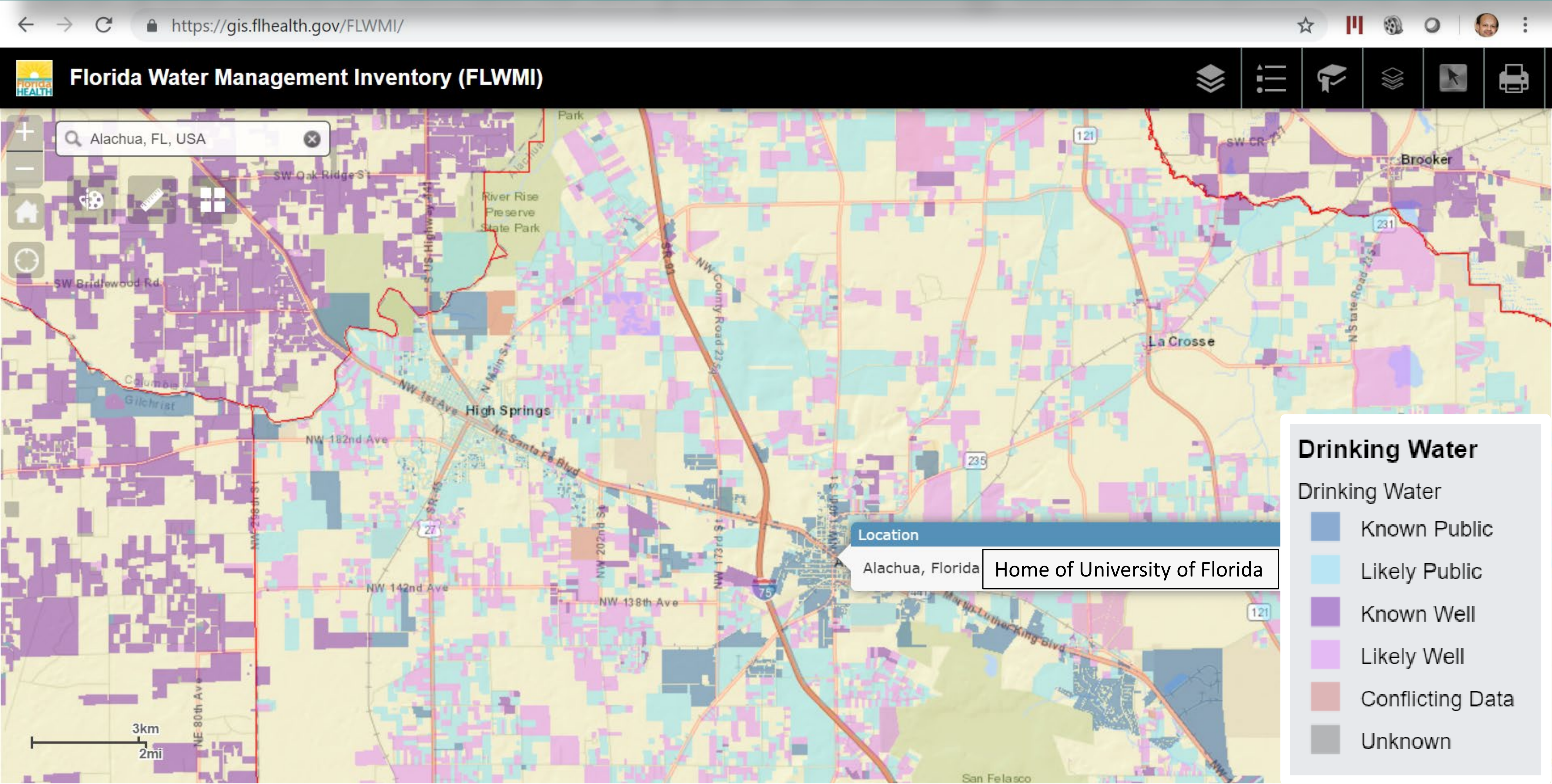
Abstract

Several lines of evidence indicate that the etiology of late-onset Alzheimer's disease (LOAD) is complex, with significant contributions from both genes and environmental factors. Recent research suggests the importance of epigenetic mechanisms in defining the relationship between environmental exposures and LOAD. In epidemiologic studies of adults, cumulative lifetime lead (Pb) exposure has been associated with accelerated declines in cognition. In addition, research in animal models suggests a causal association between Pb exposure during early life, epigenetics, and LOAD. There are multiple challenges to human epidemiologic research evaluating the relationship between epigenetics, LOAD, and Pb exposure. Epidemiologic studies are not well-suited to accommodate the long latency period between exposures during early life and onset of Alzheimer's disease. There is also a lack of validated circulating epigenetics biomarkers and retrospective biomarkers of Pb exposure. Members of our research group have shown bone Pb is an accurate measurement of historical Pb exposure in adults, offering an avenue for future epidemiologic studies. However, this would not address the risk of LOAD attributable to early-life Pb exposures. Future studies that use a cohort design to measure both Pb exposure and validated epigenetic biomarkers of LOAD will be useful to clarify this important relationship.

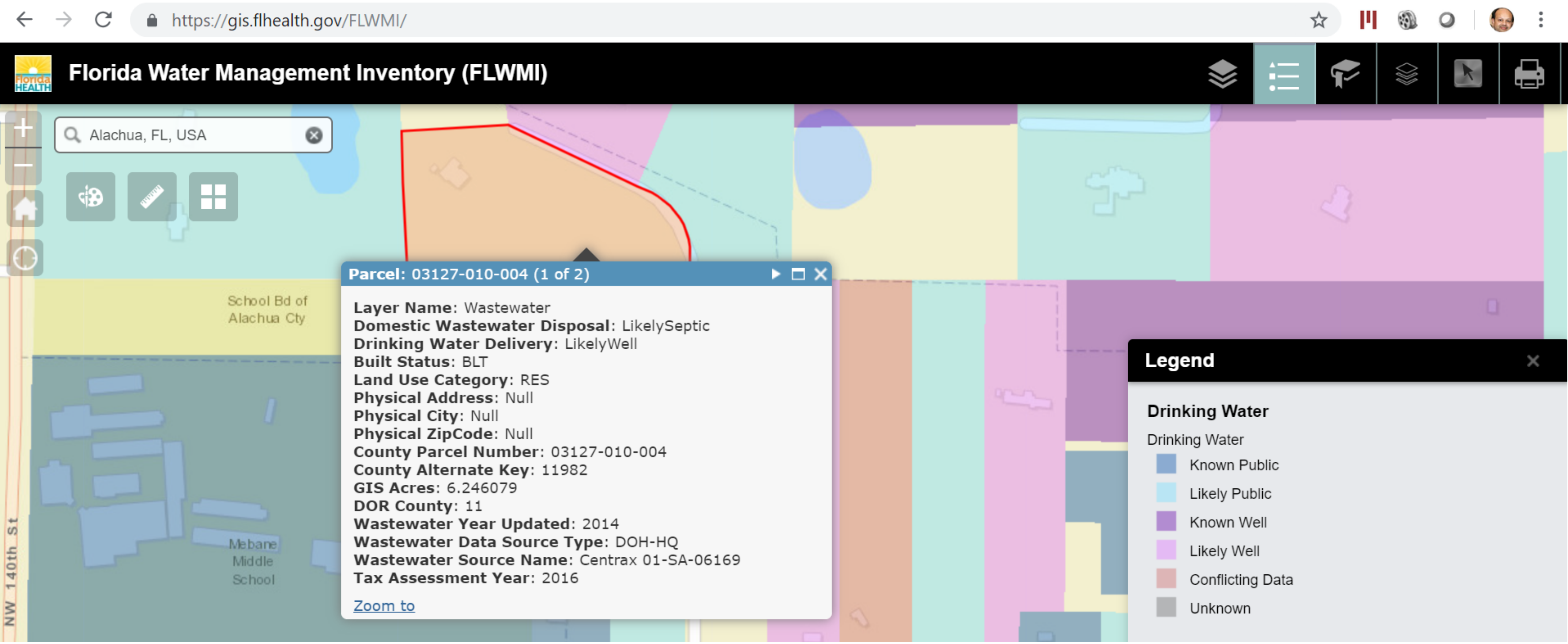
Do you know what is in your drinking water?

https://www.cdc.gov/nceh/lead/publications/nceh_prevent_childhood_lead_poisoning_508.pdf

Do you know what is in your drinking water from the well?



Drinking water from wells – schools near University of Florida



Drinking water from water treatment plant in Gainesville, FL

Gainesville Regional Utilities (GRU) - Murphree WTP

EWG's drinking water quality report shows results of tests conducted by the water utility and provided to the Environmental Working Group by the Florida Department of Environmental Protection, as well as information from the U.S. EPA Enforcement and Compliance History database (ECHO). For the latest quarter assessed by the EPA (July 2018 - September 2018), **tap water provided by this water utility was in compliance with federal health-based drinking water standards.**

WHAT ABOUT LEAD? | WANT TO FILTER THESE CONTAMINANTS OUT?

3 contaminants detected above health guidelines

5 other detected contaminants

Includes chemicals detected in 2015 for which annual utility averages exceeded an EWG-selected health guideline established by a federal or state public health authority; chemicals detected under the EPA's Unregulated Contaminant Monitoring Rule (UCMR 3) program in 2013 to 2015, for which annual utility averages exceeded a health guideline established by a federal or state public health authority; radiological contaminants detected between 2010 and 2015.

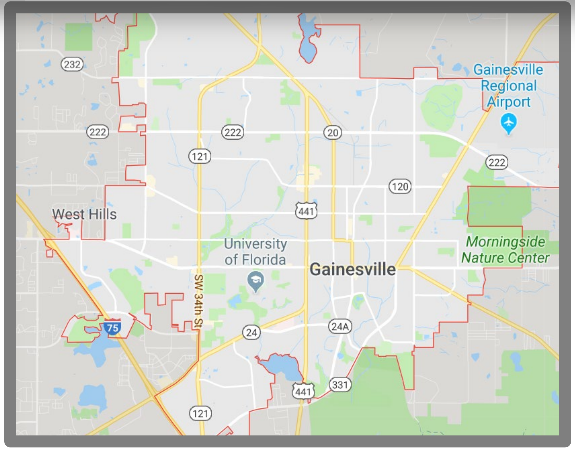
- Chromium (hexavalent) *cancer* +
- Radiological contaminants *cancer* +
- Total trihalomethanes (TTHMs) *cancer* +

WANT TO FILTER THESE CONTAMINANTS OUT?

Pollution sources

Click on each pollution source to see from which source contaminants come.

Agriculture | Industry | Treatment byproducts | Runoff & sprawl | Naturally occuring



Gainesville Regional Utilities (GRU) - Murphree WTP

EWG's drinking water quality report shows results of tests conducted by the water utility and provided to the Environmental Working Group by the Florida Department of Environmental Protection, as well as information from the U.S. EPA Enforcement and Compliance History database (ECHO). For the latest quarter assessed by the EPA (July 2018 - September 2018), **tap water provided by this water utility was in compliance with federal health-based drinking water standards.**

WHAT ABOUT LEAD? | WANT TO FILTER THESE CONTAMINANTS OUT?

3 contaminants detected above health guidelines

5 other detected contaminants

Includes chemicals detected in 2015 for which annual utility averages were lower than an EWG-selected health guideline established by a federal or state public health authority; chemicals detected under the EPA's Unregulated Contaminant Monitoring Rule (UCMR 3) program in 2013 to 2015, for which annual utility averages were lower than an EWG-selected health guideline established by a federal or state public health authority.

- Chlorate +
- Chromium (total) +
- Haloacetic acids (HAA5) +
- Strontium +
- Vanadium +

WANT TO FILTER THESE CONTAMINANTS OUT?

www.ewg.org/tapwater/system.php?pws=FL2010946 https://www.nrdc.org/stories/flint-water-crisis-everything-you-need-know

February 12, 2019

In an agreement filed by the parties of the 2017 settlement, the city of Flint committed to using a data-driven approach to locate the remaining lead pipes delivering drinking water to residents' homes. The city will use a statistical model—already proven effective in earlier efforts—to guide its selection of homes for service line excavations in 2019. This approach will increase efficiency and help ensure all remaining lead pipes are identified and removed.

https://www.nrdc.org/stories/flint-water-crisis-everything-you-need-know

NRDC

OUR WORK OUR EXPERTS OUR STORIES GET INVOLVED ABOUT US

OUR STORIES › GUIDE

Flint Water Crisis: Everything You Need to Know

After officials repeatedly dismissed claims that Flint's water was making people sick, residents took action. Here's how the lead contamination crisis unfolded—and what we can learn from it.

November 08, 2018 | Melissa Denchak

https://www.nih.gov/news-events/news-releases/baby-teeth-link-autism-heavy-metals-nih-study-suggests

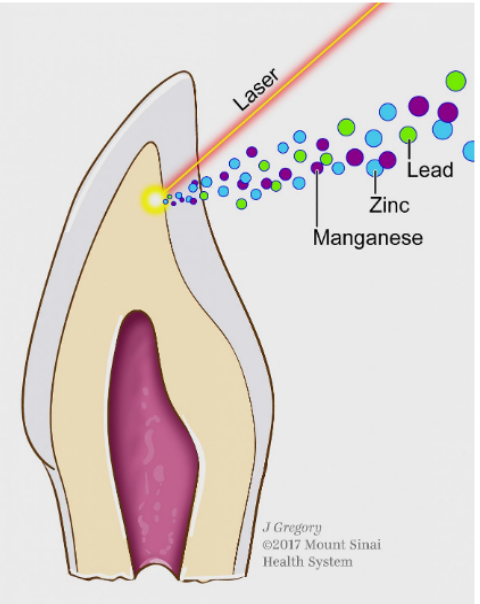
Thursday, June 1, 2017

Baby teeth link autism and heavy metals, NIH study suggests

Baby teeth from children with autism contain more toxic lead and less of the essential nutrients zinc and manganese, compared to teeth from children without autism, according to an innovative study funded by the National Institute of Environmental Health Sciences (NIEHS), part of the National Institutes of Health. The researchers studied twins to control genetic influences and focus on possible environmental contributors to the disease. The findings, published June 1 in the journal Nature Communications, suggest that differences in early-life exposure to metals, or more importantly how a child's body processes them, may affect the risk of autism.

The differences in metal uptake between children with and without autism were especially notable during the months just before and after the children were born. The scientists determined this by using lasers to map the growth rings in baby teeth generated during different developmental periods.

The researchers observed higher levels of lead in children with autism throughout development, with the greatest disparity observed during the period following birth. They also observed lower uptake of manganese in children with autism, both before and after birth. The pattern was more complex for zinc. Children with autism had lower zinc levels earlier in the womb, but these levels then increased after birth, compared to children without autism.



Cross-section of tooth showing laser removal of the dentine layer, in tan, for analysis of metal content. *Mount Sinai Health System*

https://guides.uflib.ufl.edu/precisionpublichealth/topic

Environmental Health Perspectives
https://ehp.niehs.nih.gov/doi/10.1289/ehp.9716

The Association between Blood Lead Levels and Osteoporosis Adults—Results from the Third National Health and Nutrition Examination Survey (NHANES III)

James R. Campbell and Peggy Auinger

Published: 1 July 2007 | https://doi.org/10.1289/ehp.9716 | Cited by: 35

78% of the U.S. population (1970s) had blood lead levels ≥ 10 µg/dL. Bone is a repository for 90–95% of the total body burden of lead and harbors it for years after initial exposure. Thus, a high proportion of adult Americans may currently have elevated bone lead levels. With this many who were exposed to lead when younger, and the morbidity associated with osteoporosis, it is important to investigate whether an association exists between lead exposure and osteoporosis in humans. Our objective was to conduct a secondary analysis to explore an association between lead exposure and osteoporosis in U.S. adults.

Pb from water accumulates in teeth and bones. When Pb leaches out of bones, it may contribute to osteoporosis in adult life. Increased amount of Pb in blood may also contribute to dementia, Alzheimer's and neurotoxicity.

www.ncbi.nlm.nih.gov/pmc/articles/PMC3567843/pdf/nihms367232.pdf

NIH Public Access
Author Manuscript
Curr Alzheimer Res. Author manuscript; available in PMC 2013 February 08.

Published in final edited form as:
Curr Alzheimer Res. 2012 June ; 9(5): 563–573.

Alzheimer's Disease and Environmental Exposure to Lead: The Epidemiologic Evidence and Potential Role of Epigenetics

Kelly M. Bakulski[1], **Laura S. Rozek**[1,2], **Dana C. Dolinoy**[1], **Henry L. Paulson**[3], and **Howard Hu**[1,4,5,*]

[1]University of Michigan, School of Public Health, Department of Environmental Health Sciences

[2]University of Michigan, Medical School, Department of Otolaryngology

[3]University of Michigan, Department of Neurology

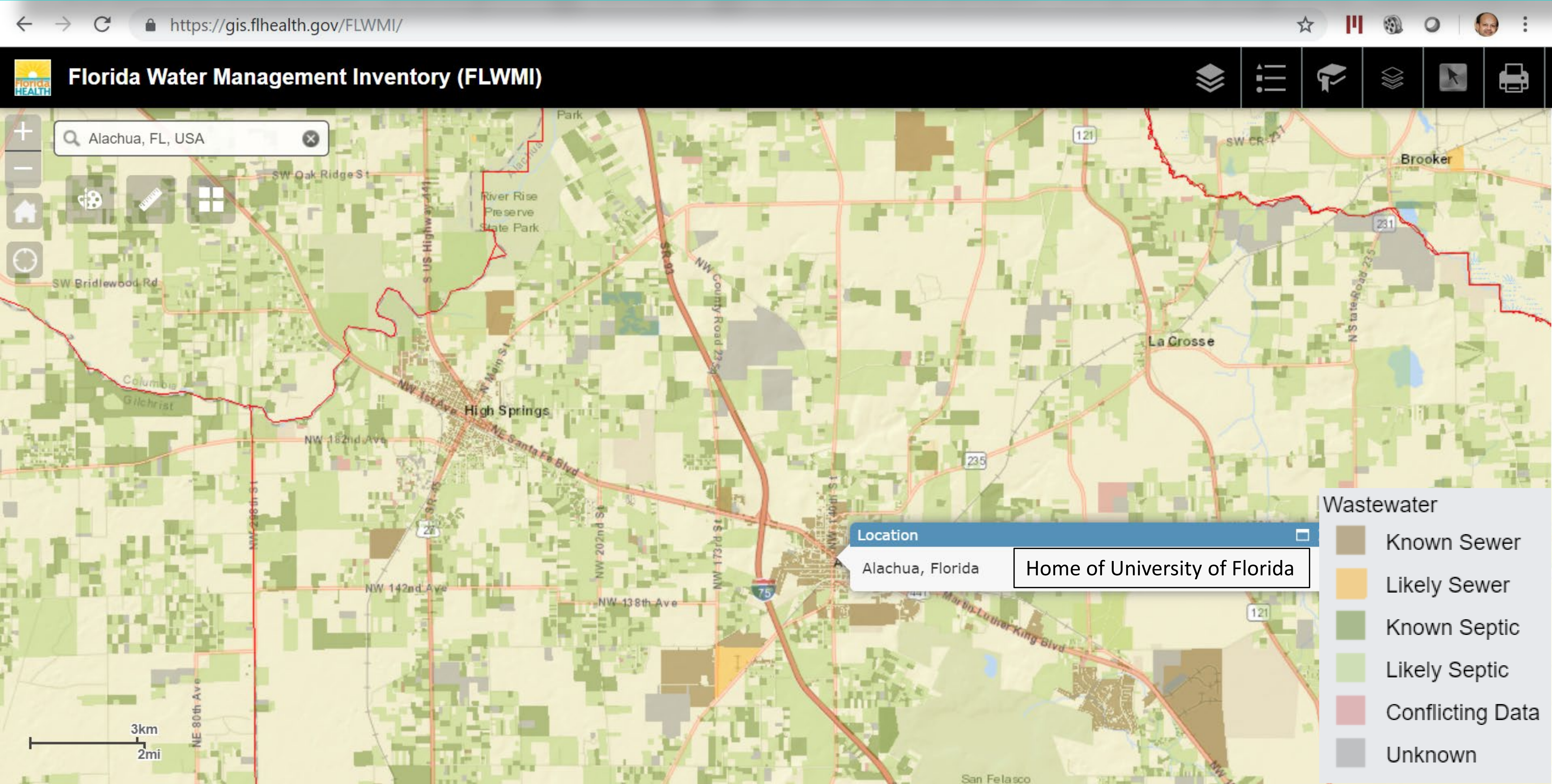
[4]University of Michigan, Department of Epidemiology

[5]University of Michigan, Medical School, Department of Internal Medicine

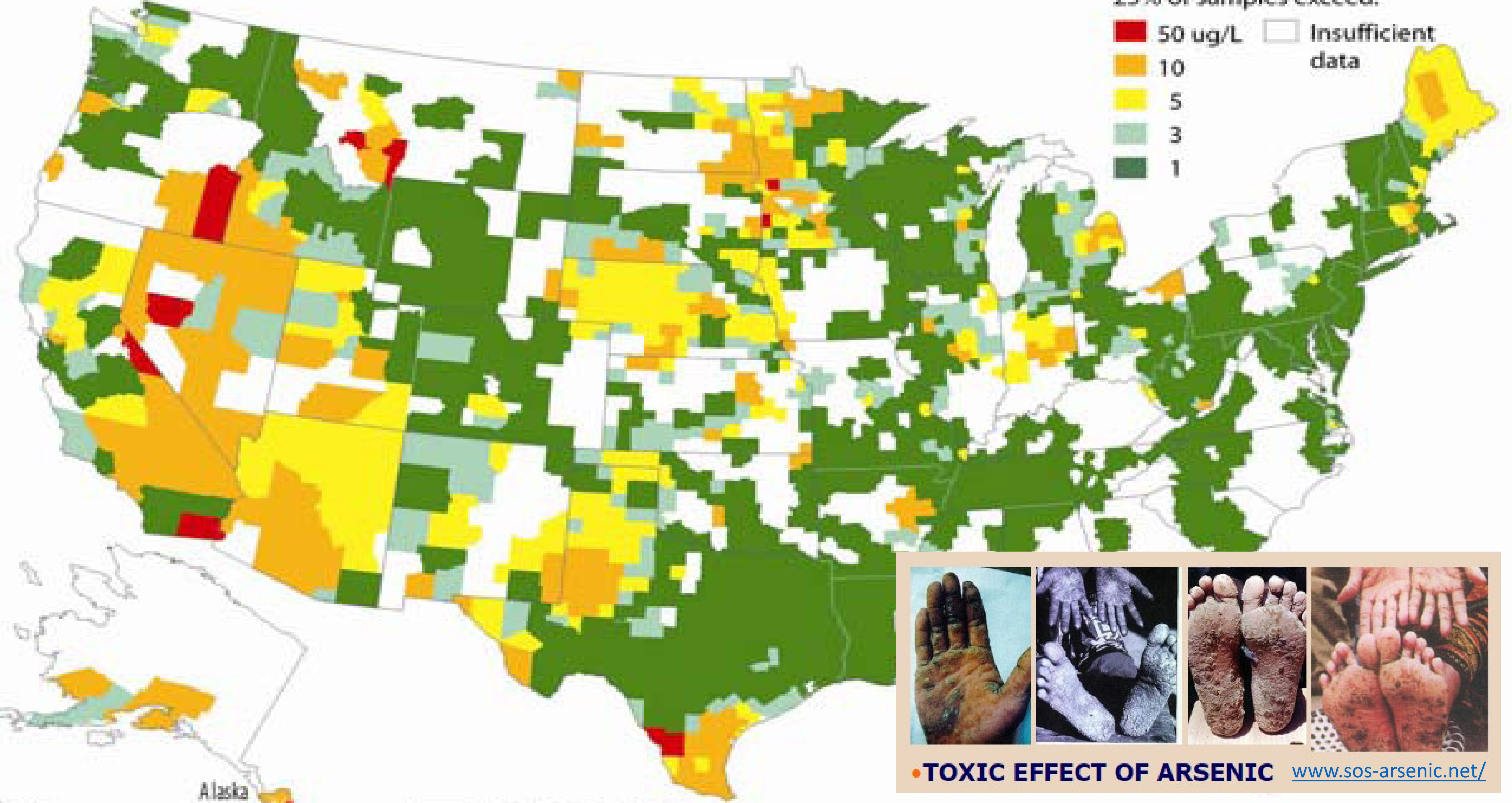
Abstract

Several lines of evidence indicate that the etiology of late-onset Alzheimer's disease (LOAD) is complex, with significant contributions from both genes and environmental factors. Recent research suggests the importance of epigenetic mechanisms in defining the relationship between environmental exposures and LOAD. In epidemiologic studies of adults, cumulative lifetime lead (Pb) exposure has been associated with accelerated declines in cognition. In addition, research in animal models suggests a causal association between Pb exposure during early life, epigenetics, and LOAD. There are multiple challenges to human epidemiologic research evaluating the relationship between epigenetics, LOAD, and Pb exposure. Epidemiologic studies are not well-suited to accommodate the long latency period between exposures during early life and onset of Alzheimer's disease. There is also a lack of validated circulating epigenetics biomarkers and retrospective biomarkers of Pb exposure. Members of our research group have shown bone Pb is an accurate measurement of historical Pb exposure in adults, offering an avenue for future epidemiologic studies. However, this would not address the risk of LOAD attributable to early-life Pb exposures. Future studies that use a cohort design to measure both Pb exposure and validated epigenetic biomarkers of LOAD will be useful to clarify this important relationship.

Wastewater seeping into groundwater from septic tanks?

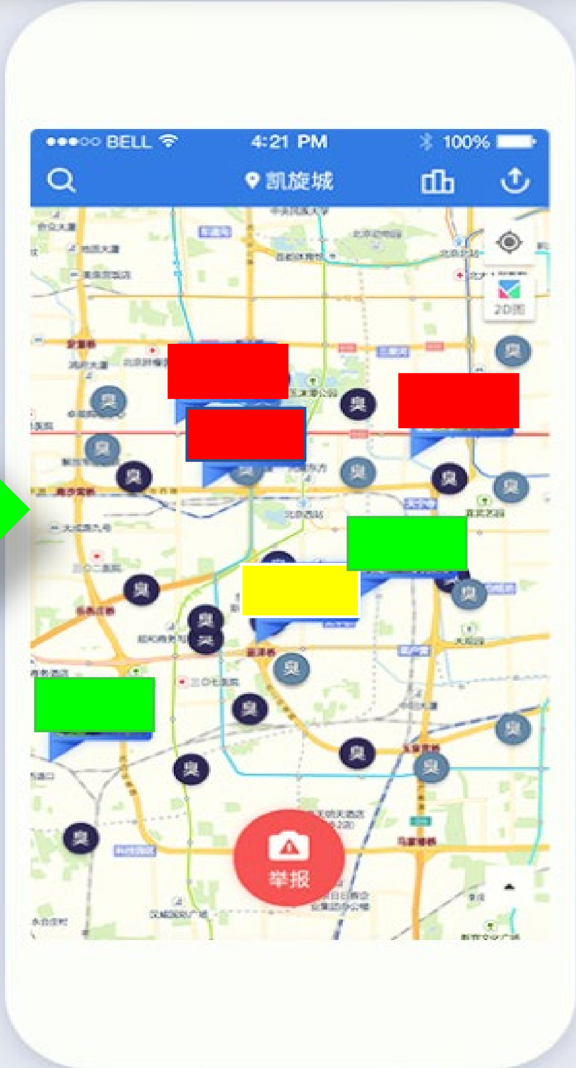


https://www.atsdr.cdc.gov/csem/arsenic/docs/arsenic.pdf



Water ART – IoT Data Analytics of Value to End-User

VALUE



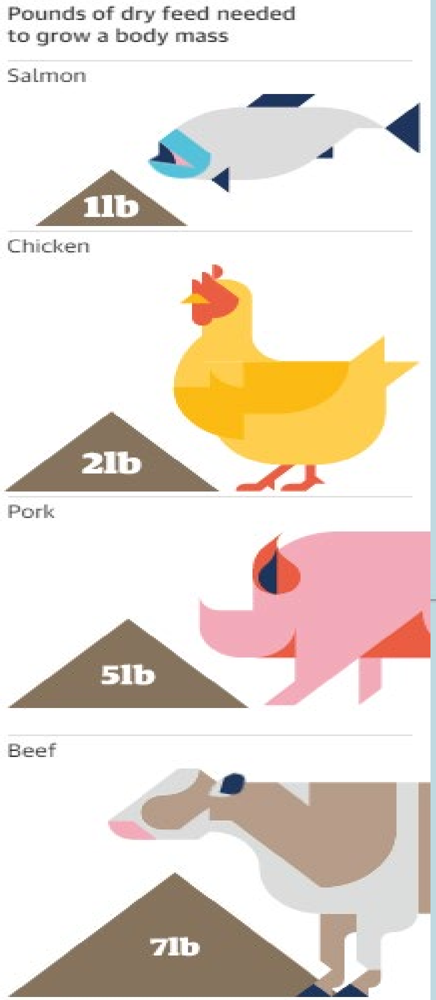
SENSEE inside

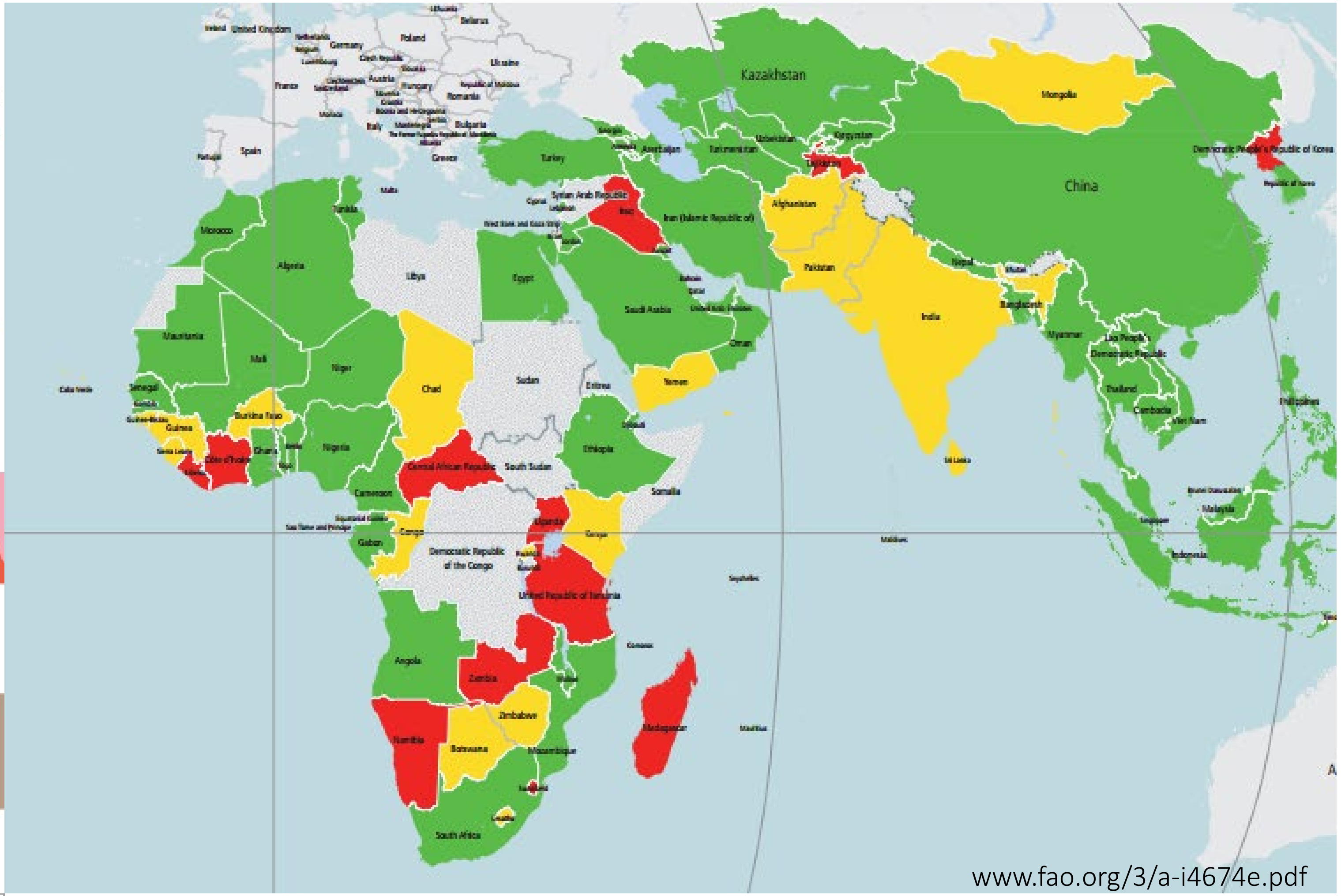
Professor Eric McLamore, University of Florida - www.ncbi.nlm.nih.gov/pubmed/29629449

FOOD v PEOPLE

Prevent Food Waste

Estimated 11 billion people to feed at the dawn of the 22nd Century





www.fao.org/3/a-i4674e.pdf

The End of Plenty
by Joel K Bourne

www.un.org/sustainabledevelopment/blog/2015/07/what-progress-has-been-made-in-ending-global-poverty/

http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Do we really ***waste*** about 68% of the food in the US ?

More than two-thirds of total food wasted – which is ~ 63 million tons

Value wasted ~ $150 Billion (total food wasted value US$218 Billion)

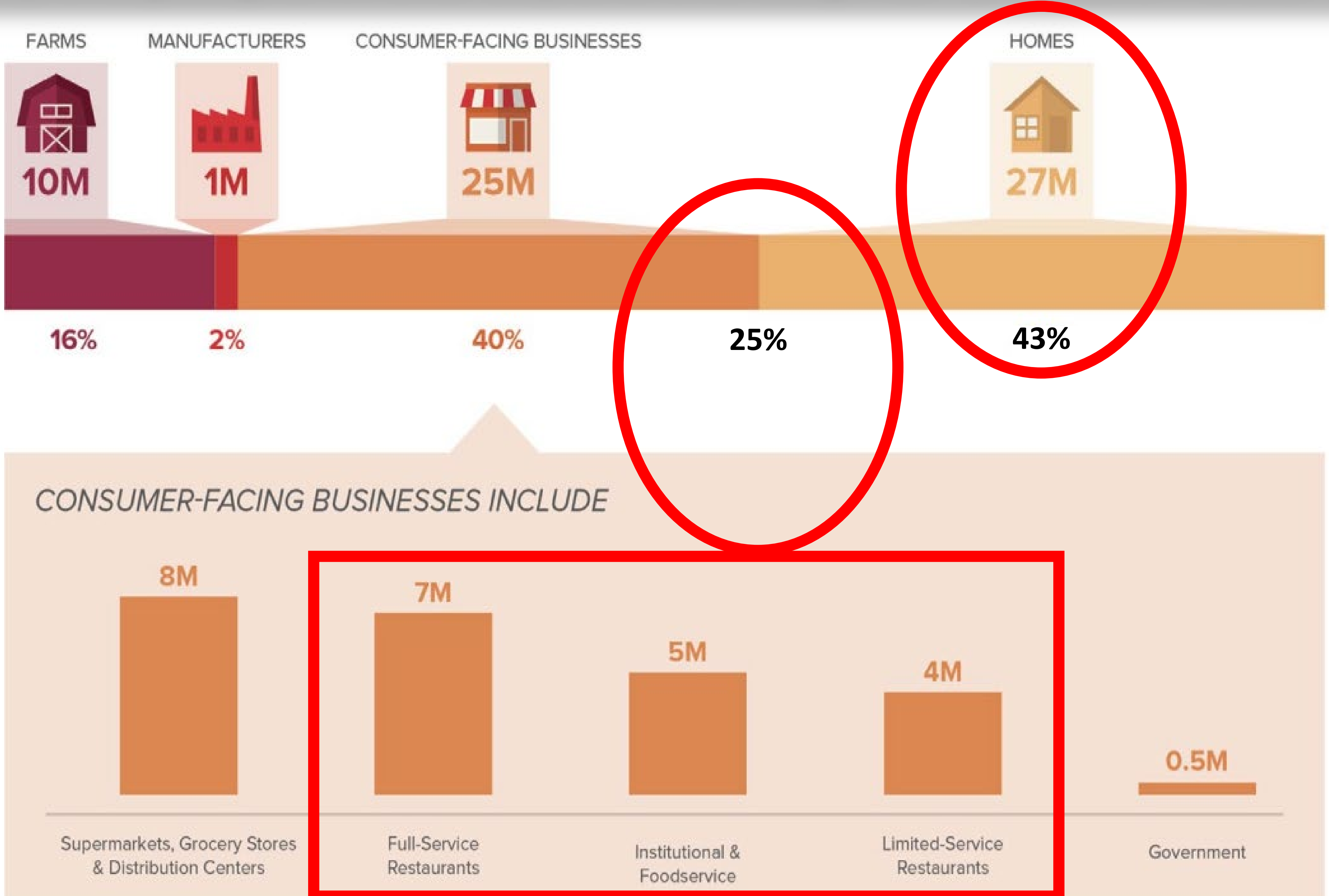
43

million tons

$150 Billion

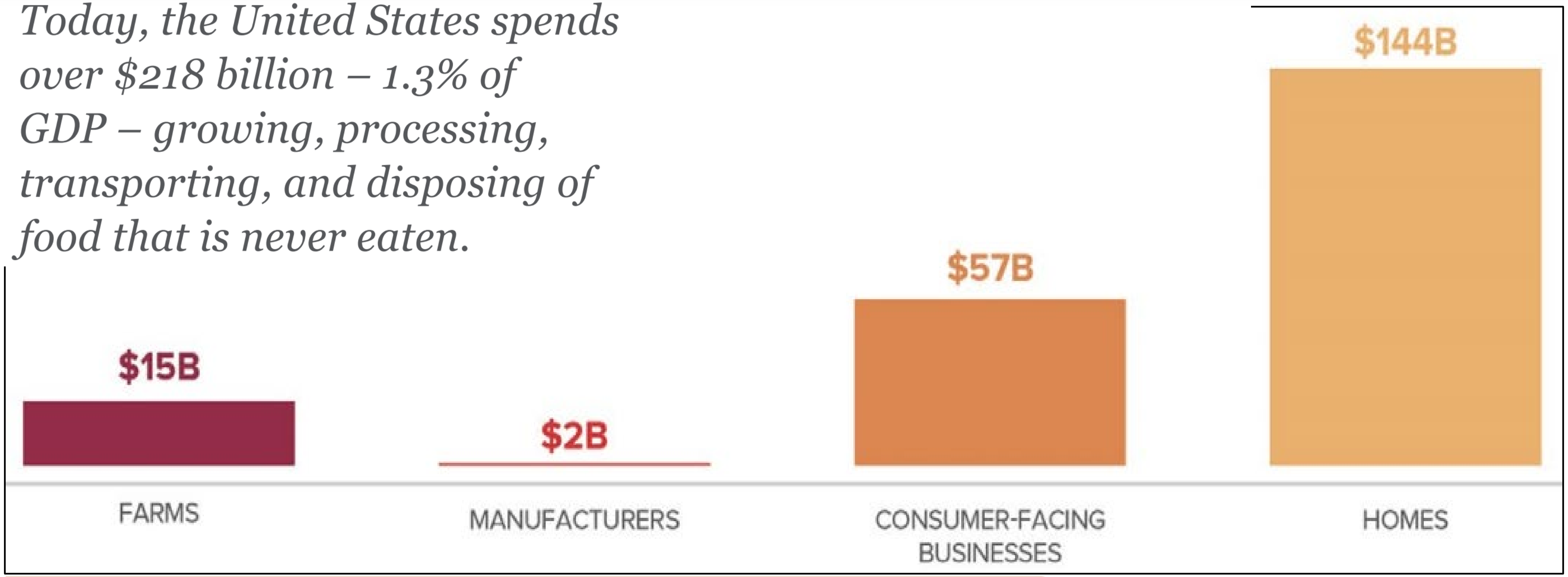
US Report ● http://www.refed.com/downloads/ReFED_Report_2016.pdf
EU Report ● http://data.consilium.europa.eu/doc/document/ST-10730-2016-INIT/en/pdf from www.eu-fusions.org/

Yes. We, the people in the US, are the culprits → ~68% FOOD WASTED



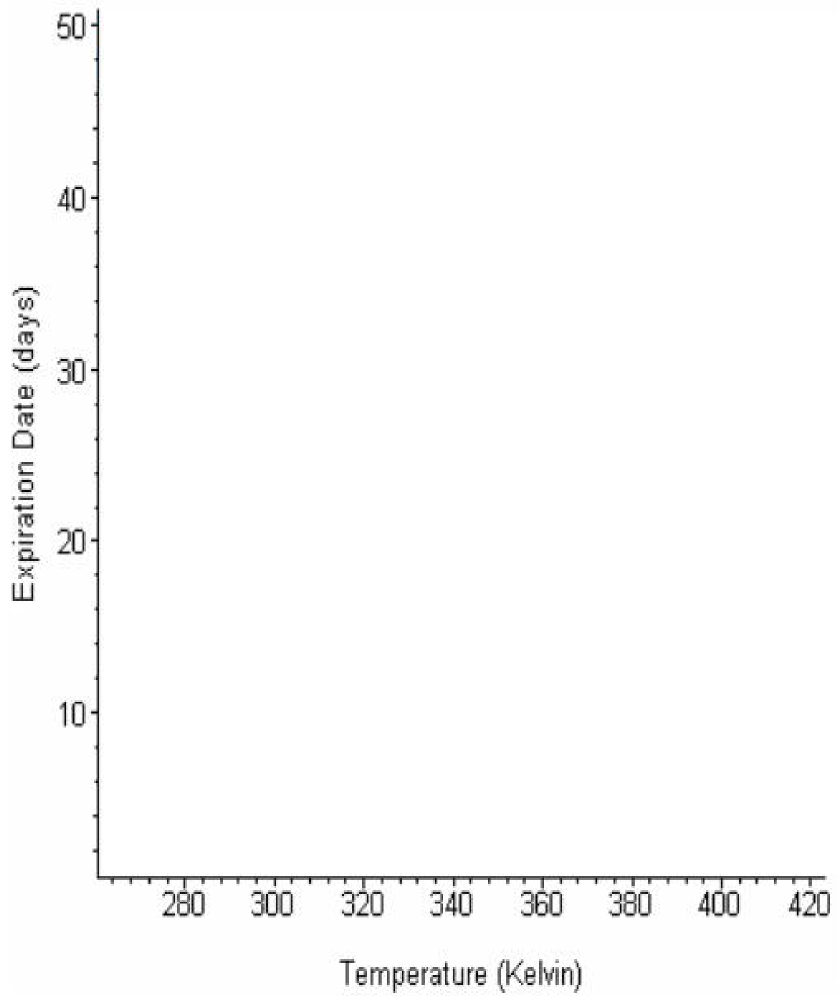
FOOD WASTE in the US: approx. 63 million tons, $218 billion, 1.3% GDP

Today, the United States spends over $218 billion – 1.3% of GDP – growing, processing, transporting, and disposing of food that is never eaten.



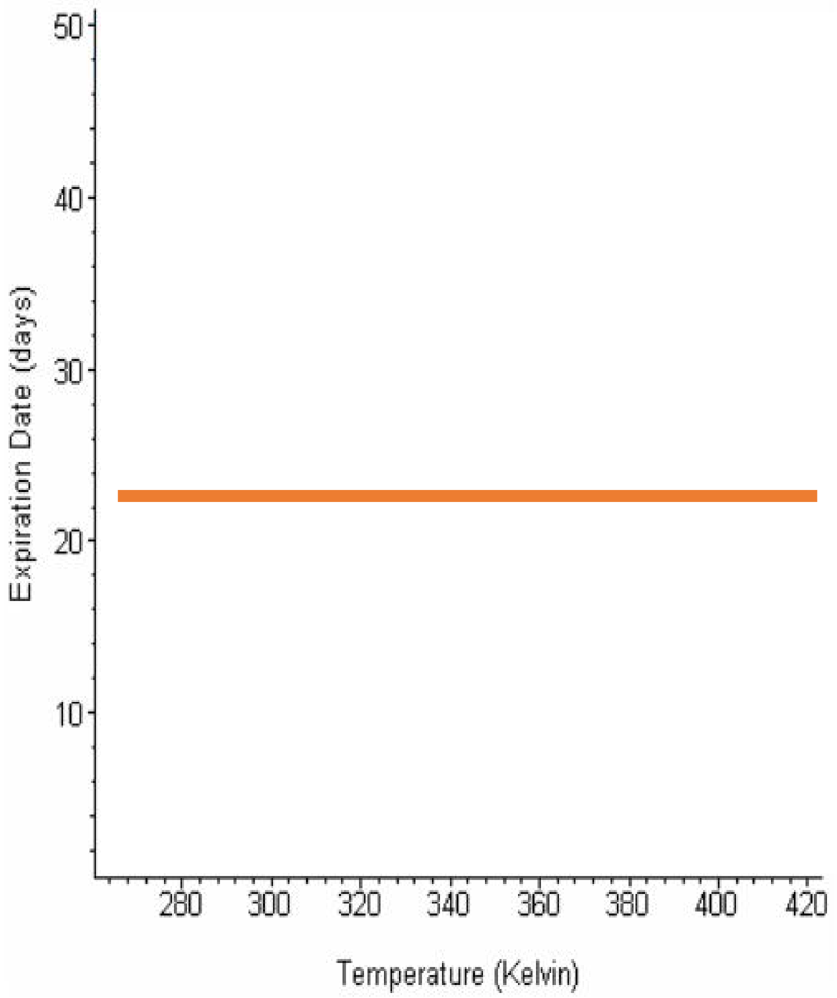
http://www.refed.com/downloads/ReFED_Report_2016.pdf

Storage



2003

Storage



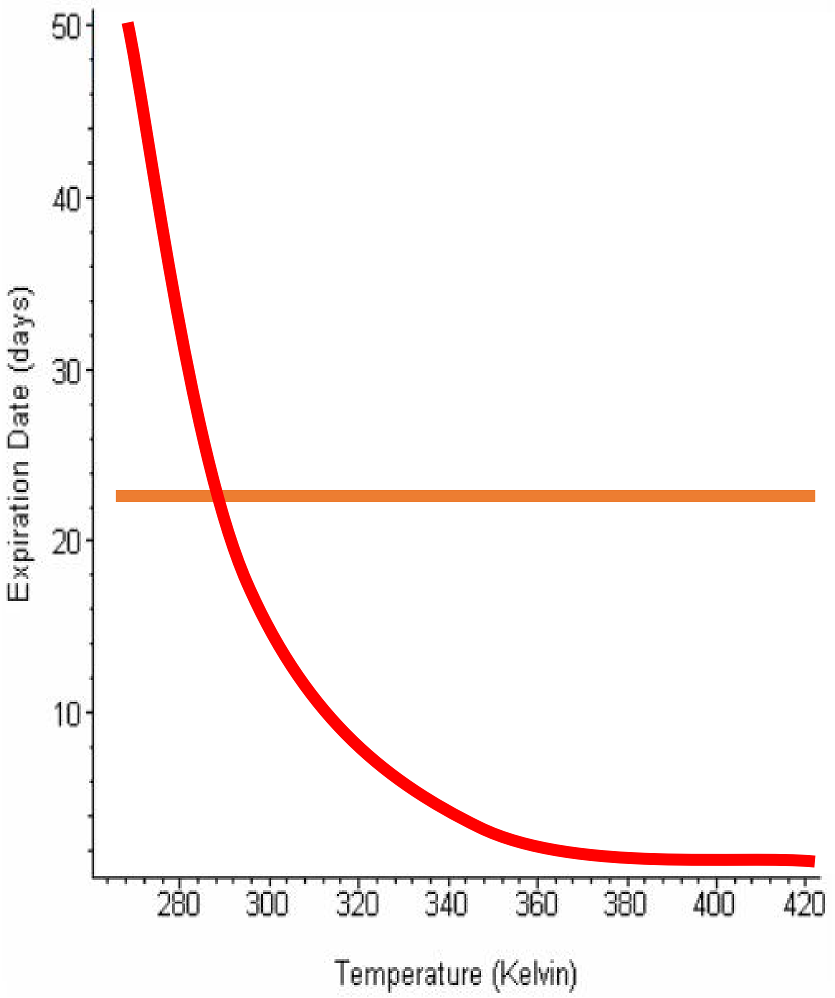
$$\frac{\partial Q}{\partial t} = -k_1 e^{\left[-\frac{E_a}{R_g T(t)}\right]} Q^n$$

Variables

- E_a Activation energy
- k_1 Arrhenius constant
- n Order of the reaction
- T Temperature
- Q Quality
- t Time

2003

Storage



$$\frac{\partial Q}{\partial t} = -k_1 e^{\left[-\frac{E_a}{R_g T(t)}\right]} Q^n$$

Variables

- E_a Activation energy
- k_1 Arrhenius constant
- n Order of the reaction
- T Temperature
- Q Quality
- t Time

2003

Shelf Life

Name: ⊞ Activation Energy

Name: ⊞ Arrhenius Constant

Name: ⊞ Temperature

Name: ⊞ Quality

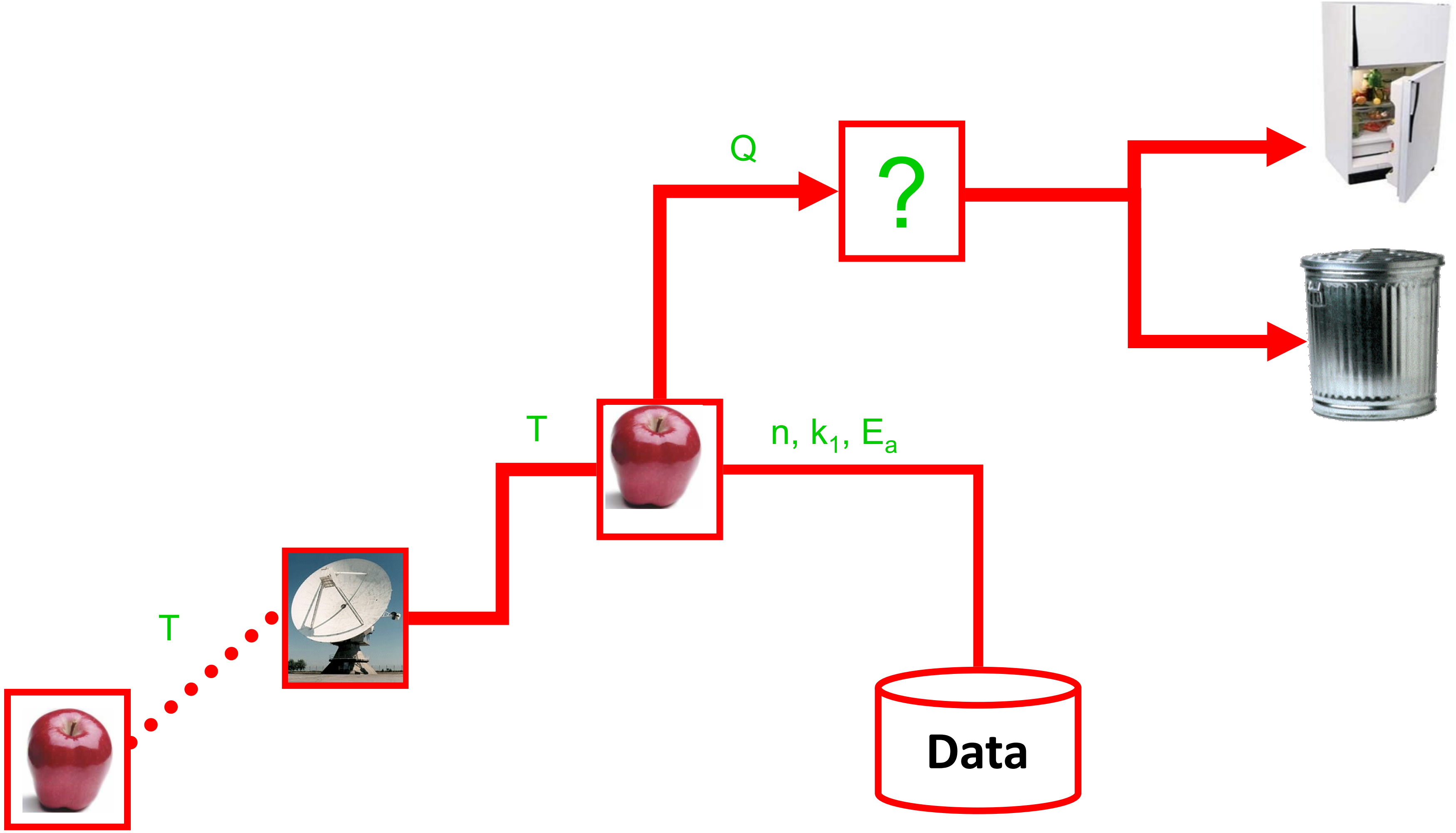
Name: ⊞ Order of Reaction
Description: ⊞Order of Reaction
Symbol: n
Access: Read
ID: EPC: 0102008419100001289731
Class: Scalar
Type: Int
Unit:
Default: 1

Food Quality

Name: ⊞ Food Quality
Description: ⊞Food Quality based Arrhenius
Developer: ⊞Natick Army Laboratories
ID: EPC: 010300908808BF60000000AA
Comp: ⊞$0.25 per month
Type: Analytic
Rate: ⊞1 to 10,000 sec
Algorithm: ⊞

2003

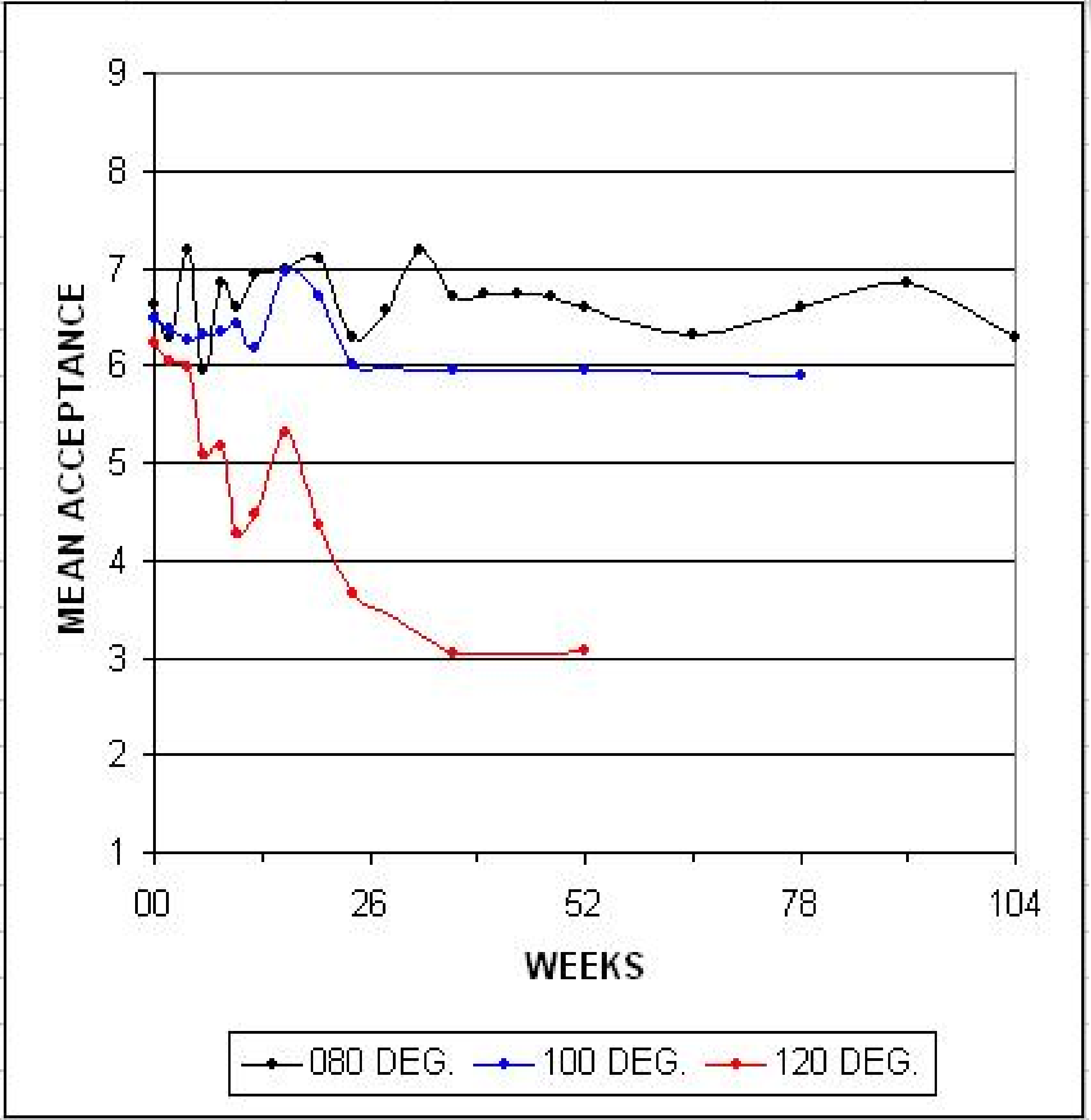
Shelf Life ➔ Answers (not numbers)



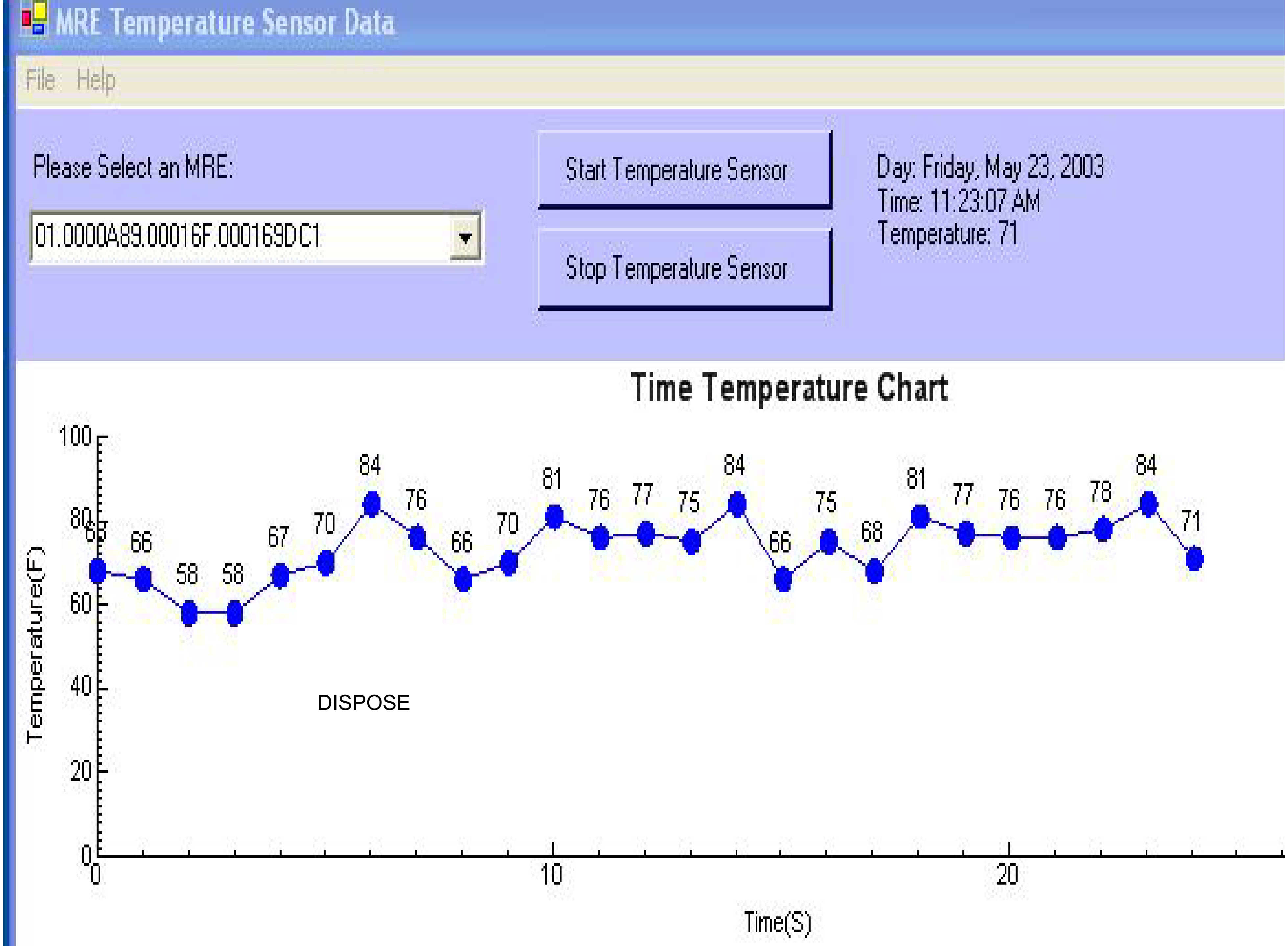
2003

Monitoring Perishables (MRE Simulation)

WKS	080 DEG.	100 DEG.	120 DEG
00	6.622	6.486	6.243
02	6.282	6.359	6.026
04	7.194	6.250	5.972
06	5.949	6.308	5.077
08	6.850	6.350	5.175
10	6.600	6.429	4.286
12	6.944	6.167	4.472
16	7.000	6.947	5.316
20	7.111	6.694	4.361
24	6.300	6.000	3.667
28	6.579		
32	7.189		
36	6.694	5.944	3.028
40	6.730		
44	6.730		
48	6.703		
52	6.583	5.944	3.056
65	6.316		
78	6.583	5.889	
91	6.842		
104	6.300		
130			
156			

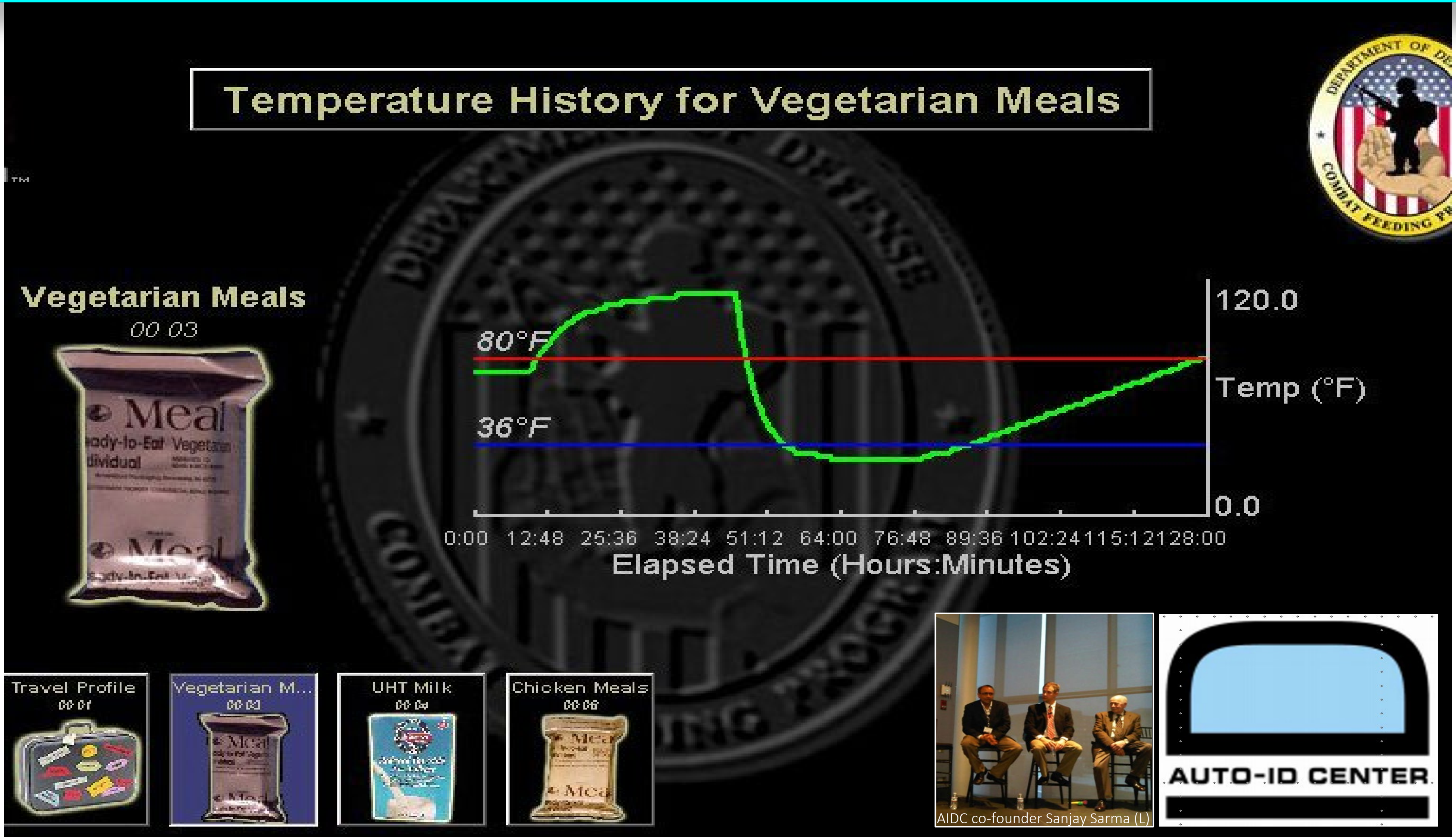


2003


MRE Temperature Sensor Data
File Help
Please Select an MRE:
01.0000A89.00016F.000169DC1
Start Temperature Sensor
Stop Temperature Sensor
Day: Friday, May 23, 2003
Time: 11:23:07 AM
Temperature: 71
Time Temperature Chart
Temperature(F)
Time(S)
DISPOSE
0 20 40 60 80 100
0 10 20
68 66 58 58 67 70 84 76 66 70 81 76 77 75 84 66 75 68 81 77 76 76 78 84 71

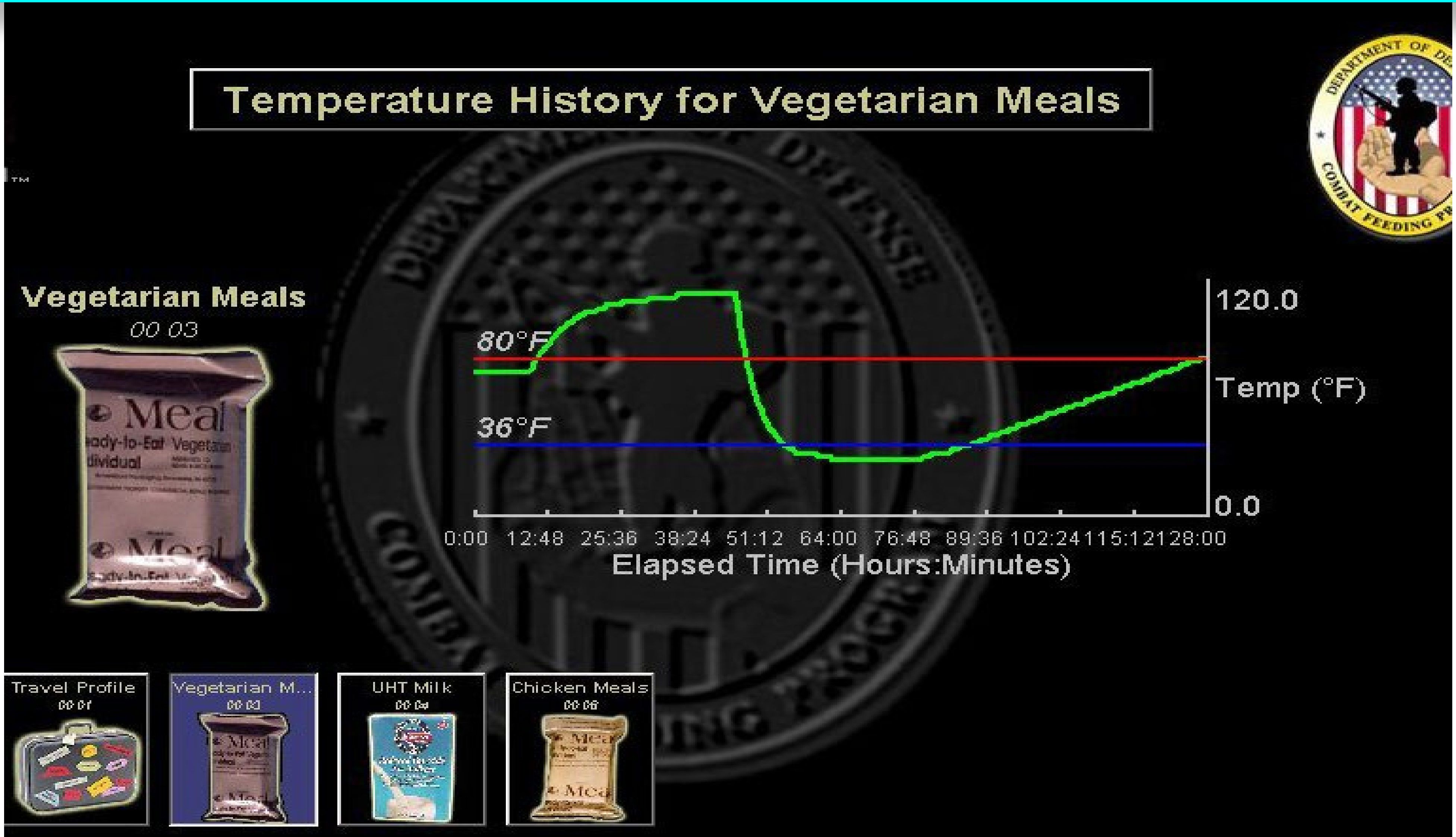
2003

SENSEE can hold RFID + Temperature Sensor Data ● Convergence of Systems



2003

Is this data analytics of value to the end-user on the front lines of a war zone?



2003

NOT GOOD

X

Knowledge Tools	Statistical Tools
• Agent-based model	• Stochastic
• Data-informed models	• Bayesian & Adaptive
• Data visualization	• Dynamic models
• Network models	• Hierarchical models
• Data & database modeling	• Factor analysis
• Document Modeling	• Monte Carlo models
• Metamodeling	• Population Modeling
• Ontological Modeling	• Dynamical system
• Business Process Modeling	• Stochastic system eqn
• Natural language processing	• Social network analysis
• Machine Learning, ANN, DL	• Topic modeling

2003

SIMPLICITY – THE TRAFFIC LIGHT OUTCOME – ANSWERS, not numbers

GOOD

2003

SIMPLICITY – THE TRAFFIC LIGHT OUTCOME – ANSWERS, not numbers

GOOD

INSPECT

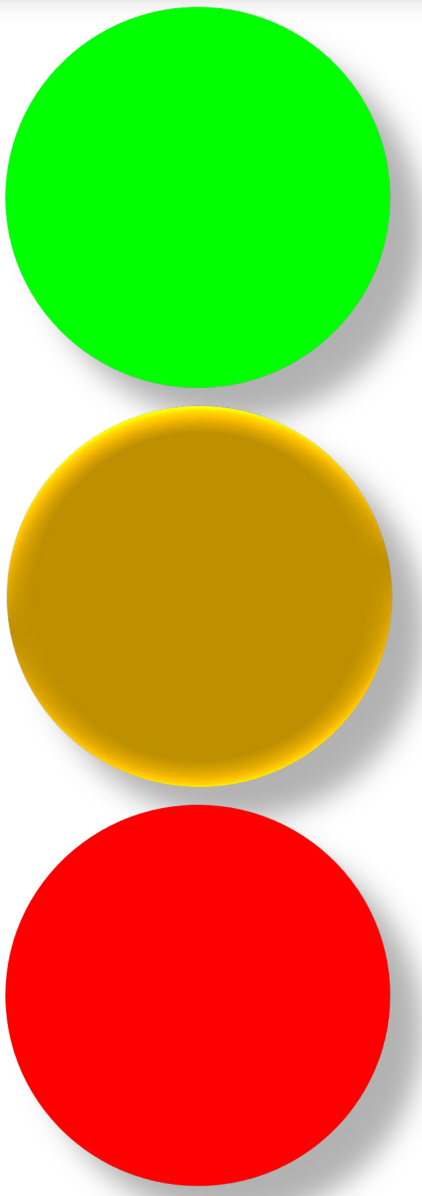
2003

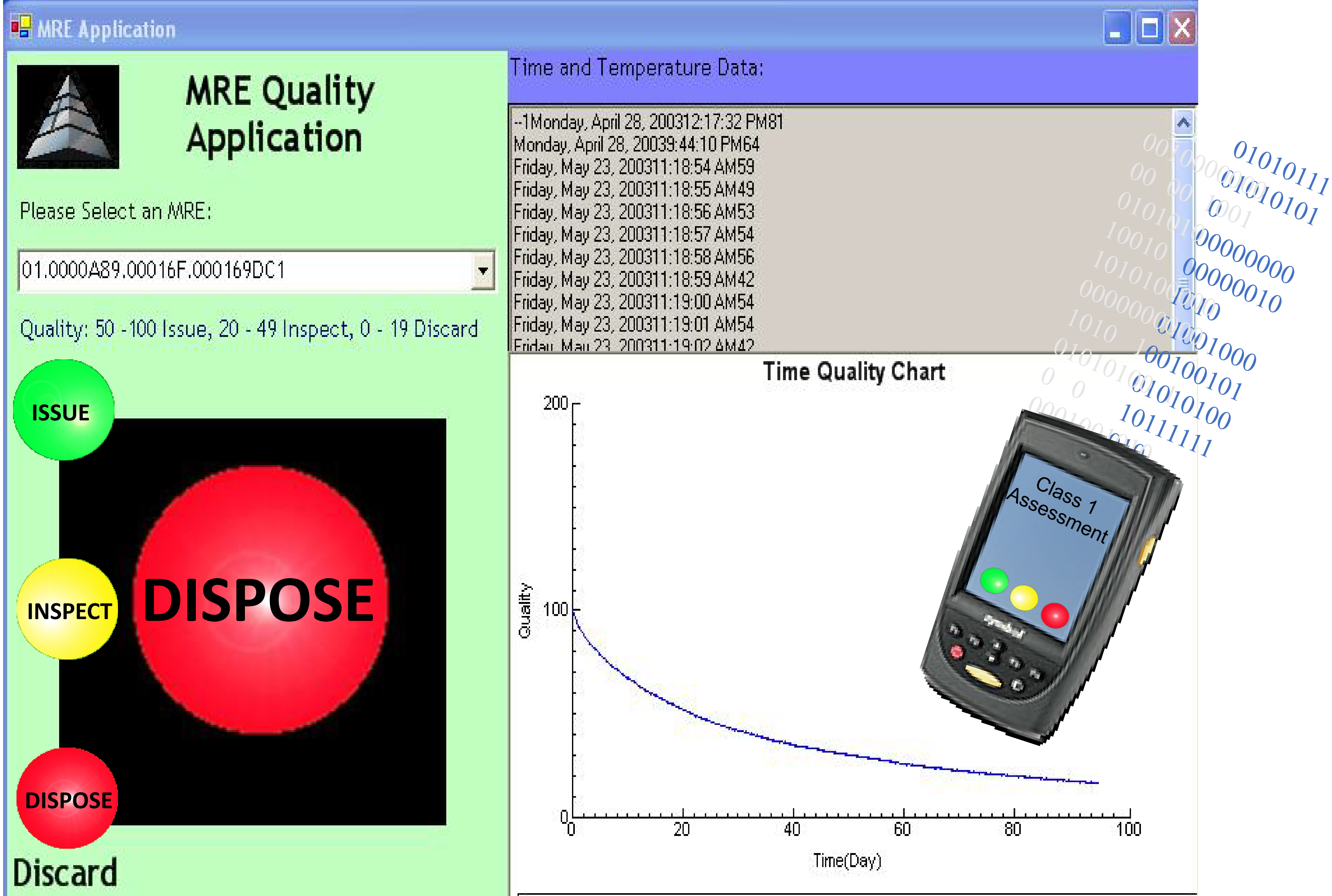
SIMPLICITY – THE TRAFFIC LIGHT OUTCOME – ANSWERS, not numbers

GOOD

INSPECT

REJECT




MRE Application
MRE Quality Application
Please Select an MRE:
01.0000A89.00016F.000169DC1
Quality: 50 - 100 Issue, 20 - 49 Inspect, 0 - 19 Discard
ISSUE
INSPECT
DISPOSE
DISPOSE
Discard
Time and Temperature Data:
--1Monday, April 28, 200312:17:32 PM81
Monday, April 28, 20039:44:10 PM64
Friday, May 23, 200311:18:54 AM59
Friday, May 23, 200311:18:55 AM49
Friday, May 23, 200311:18:56 AM53
Friday, May 23, 200311:18:57 AM54
Friday, May 23, 200311:18:58 AM56
Friday, May 23, 200311:18:59 AM42
Friday, May 23, 200311:19:00 AM54
Friday, May 23, 200311:19:01 AM54
Time Quality Chart
Quality
Time(Day)
200
100
0
0
20
40
60
80
100
Class 1 Assessment

2003

Grocery Store Perishability

Is the spinach fresh? Is the fish smelling fishy? Is the chicken safe to eat?

YOU WANT TO KNOW IF THE CHICKEN IS STILL GOOD TO EAT. YOU DON'T TRUST THE "SELL BY" DATE ON THE LABEL

WHAT IF THE PACK OF "CHICKEN" CAN TALK TO YOU AND OFFER YOU A REAL-TIME UPDATE ABOUT ITS QUALITY AND FOOD SAFETY?

YOU WANT TO KNOW IF THE CHICKEN IS STILL GOOD TO EAT. YOU DON'T TRUST THE "SELL BY" DATE ON THE LABEL

WHAT IF THE PACK OF "CHICKEN" CAN TALK TO YOU AND OFFER YOU A REAL-TIME UPDATE ABOUT ITS QUALITY AND FOOD SAFETY?

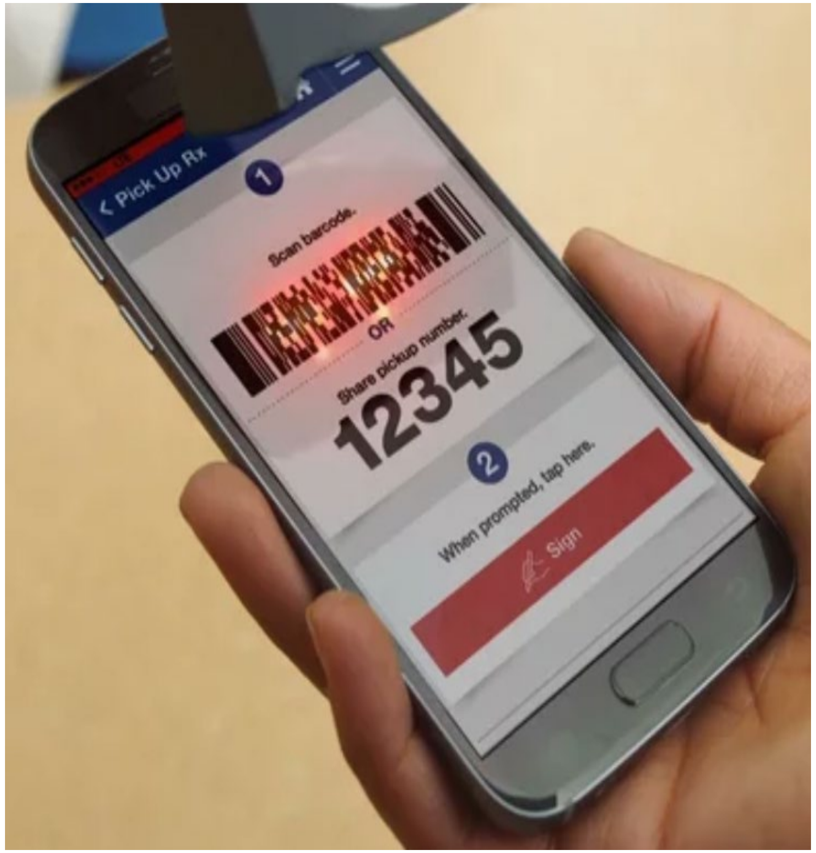

THIS IS A NON-TOXIC
BIOGENIC AMINE SENSOR
TO MONITOR THE SAFETY
OF YOUR FRESH FOOD
HARVESTLAND
PERDUE
fresh:
Chicken Thigh Filets
All Natural† • No Hormones or Steroids Added‡
No Antibiotics Ever
Fed All Vegetarian Diet
No Animal By-Products
Raised Cage Free
PROCESS USDA VERIFIED
USDA PROCESS VERIFIED
USDA A GRADE
Boneless and Skinless • Lean

Yes We Can!

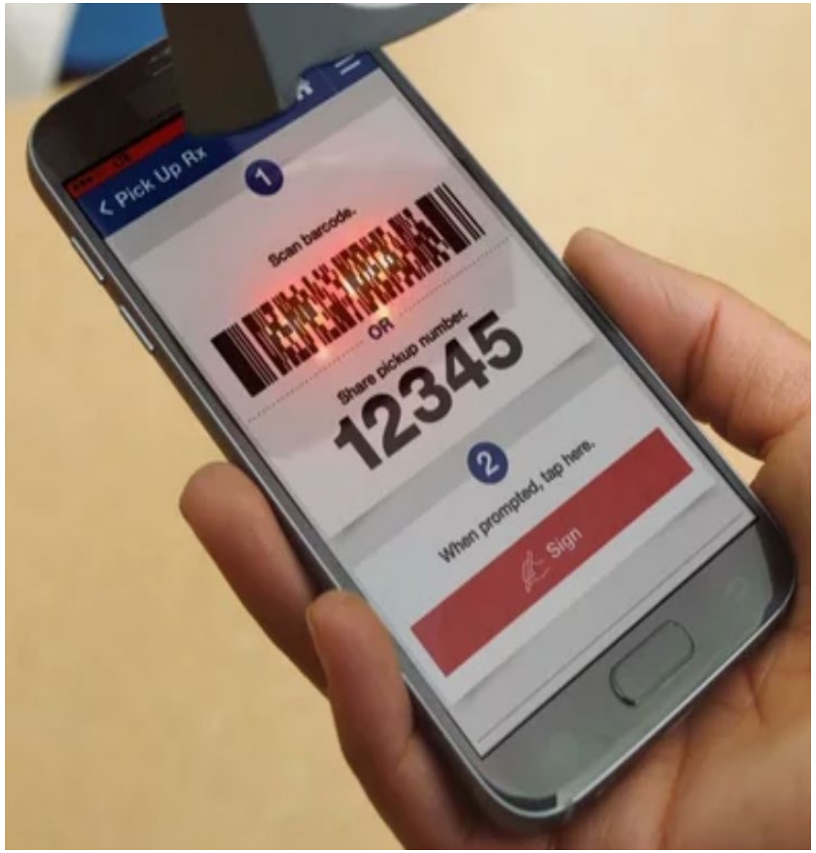

THIS IS A NON-TOXIC BIOGENIC AMINE SENSOR TO MONITOR THE SAFETY OF YOUR FRESH FOOD
HARVESTLAND
PERDUE
fresh:
Chicken Thigh Filets
All Natural • No Hormones or Steroids Added
No Antibiotics Ever
Fed All Vegetarian Diet
No Animal By-Products
Raised Cage Free
PROCESS USDA VERIFIED
Boneless and Skinless • Lean
USDA PROCESS VERIFIED
USDA A GRADE


FOOD SAFETY APP
SKU 12345
PROD CHK
QUALITY

THE SKU (PRODUCT NUMBER) SHOWS ON THE MOBILE APP

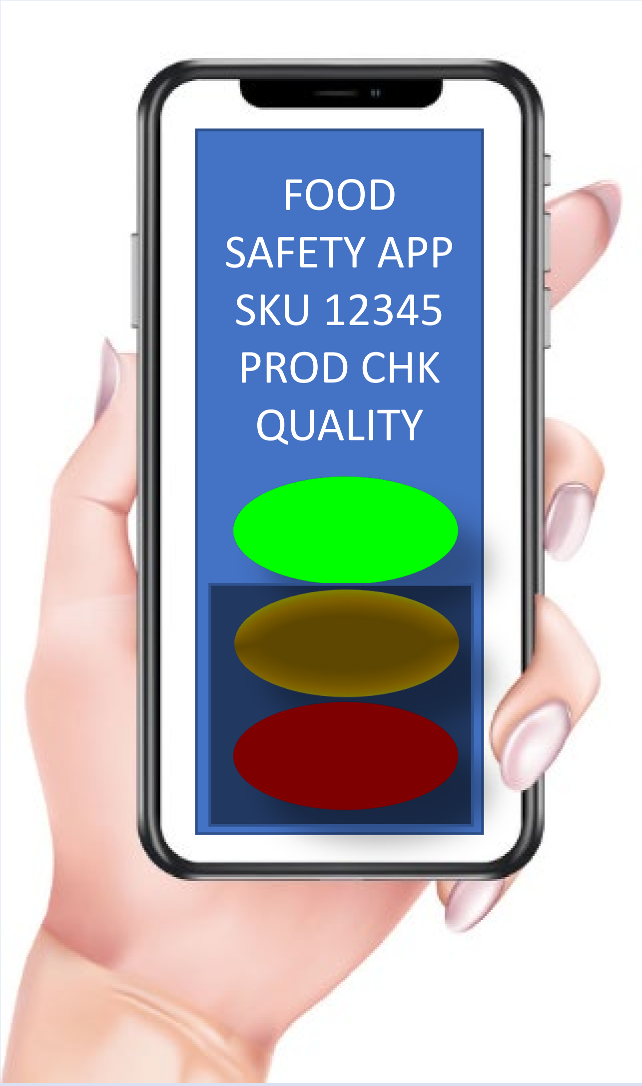


THE SKU (PRODUCT NUMBER) SHOWS ON THE MOBILE APP

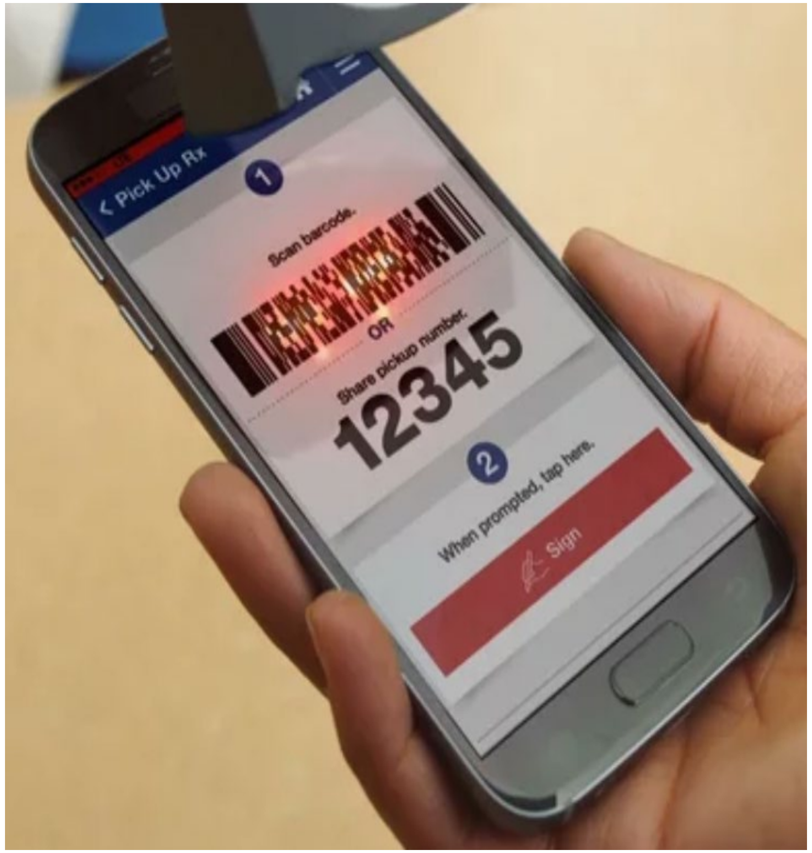


WOULD YOU PAY 1 CENT TO USE THIS FOOD APP HEALTH SAFETY SERVICE ?

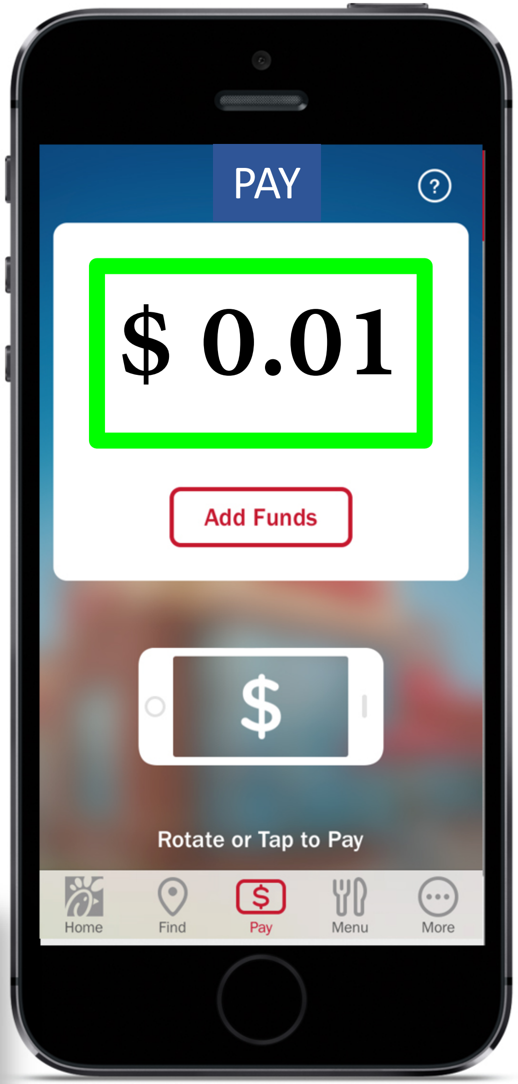
IT IS YOUR HEALTH



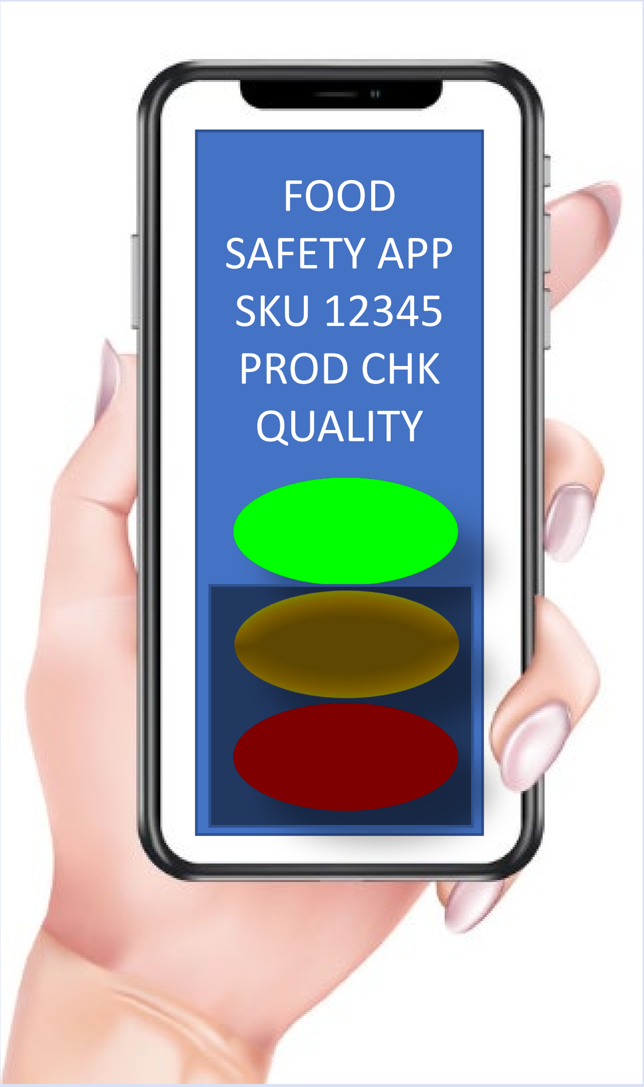
THE SKU (PRODUCT NUMBER) SHOWS ON THE MOBILE APP



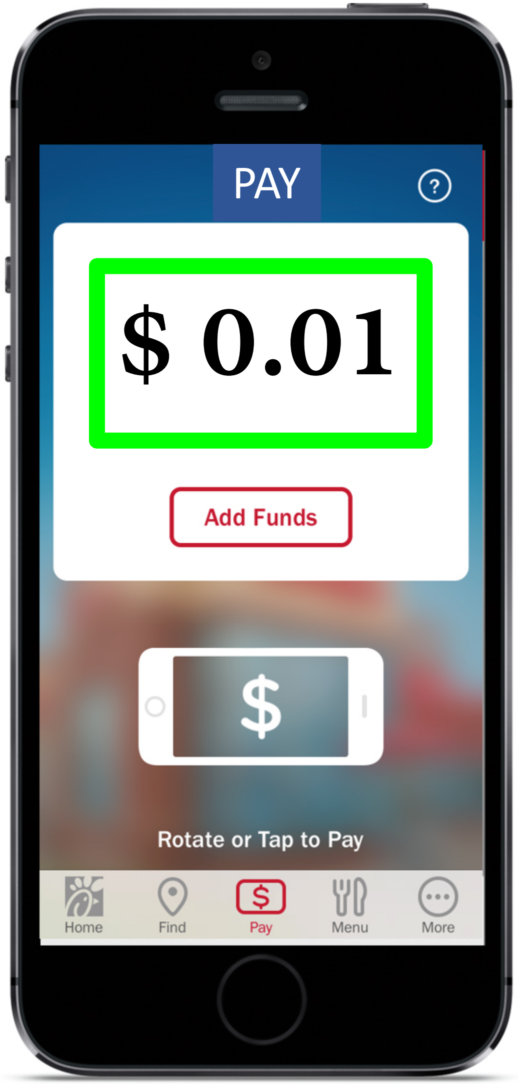
WOULD YOU PAY 1 CENT TO USE THIS FOOD APP HEALTH SAFETY SERVICE ?



IT IS YOUR HEALTH

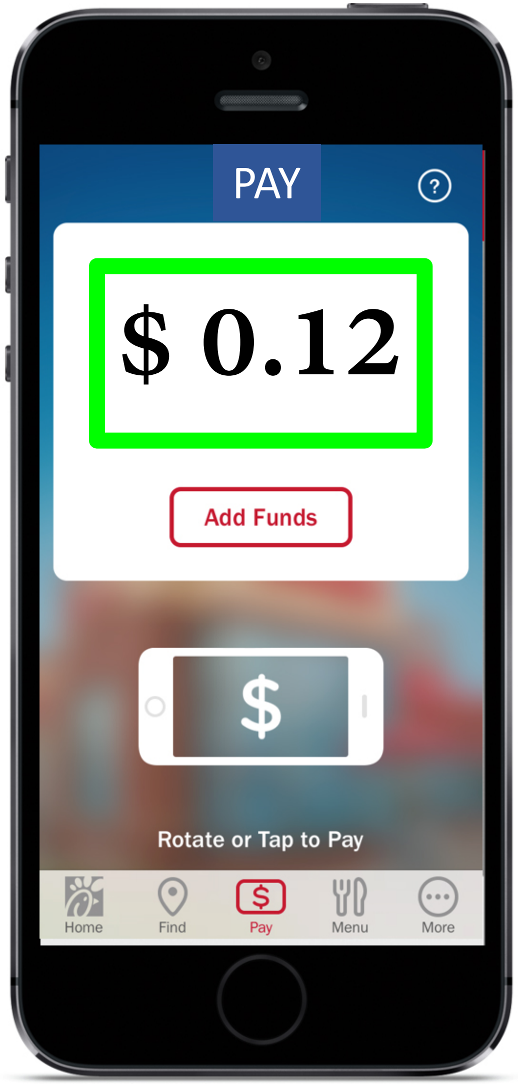


PEAS OF YOUR MIND


PAY
$ 0.01
Add Funds
Rotate or Tap to Pay
Home
Find
Pay
Menu
More

REMEMBER IoT ? HOW NANO-FEES MAY GENERATE MEGA-MILLIONS

0.12 cents

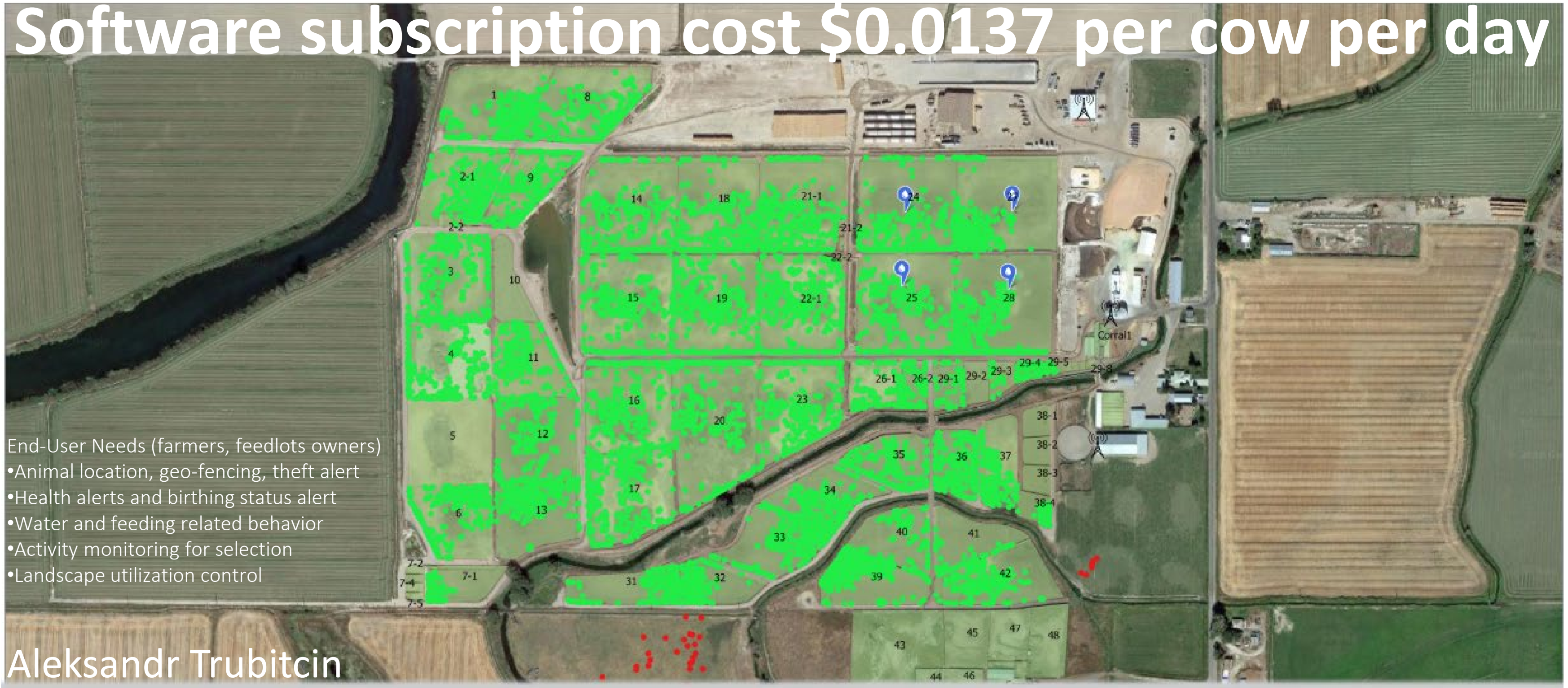


REMEMBER IoT ? HOW NANO-FEES MAY GENERATE MEGA-MILLIONS

See SIGNALS - https://dspace.mit.edu/handle/1721.1/111021

The user will also need access to the cattle tracking and monitoring web application. For example, the cost of an annual subscription to the Cattle Tags Technologies app will be $5 per animal.

0.12 cents



LoRaWAN ear tag from Cattle Tags Technologies starts from $39. Tags have embedded GPS receiver, accelerometer, temperature sensor and replaceable battery. Operator reads RFID-tag with Bluetooth reader (ID sent to ERP system). Installation of activated LoRaWAN ear tag follows. alex.trubitcin@gmail.com

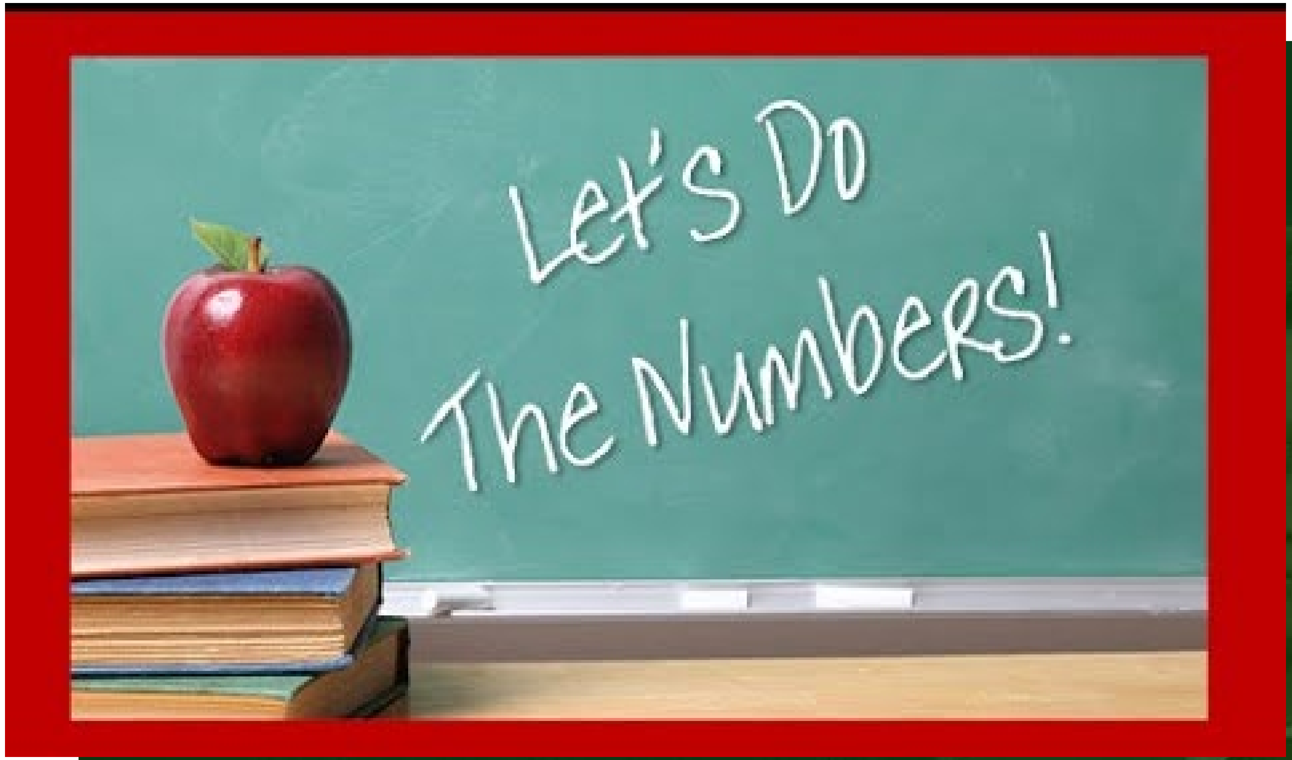
Digital Transformation (Connectivity+App+Tag) Cost $0.12 per cow per day

Proposal generates following remark from a VC (Mr Vinod Vaticinator, Venture Capitalist)

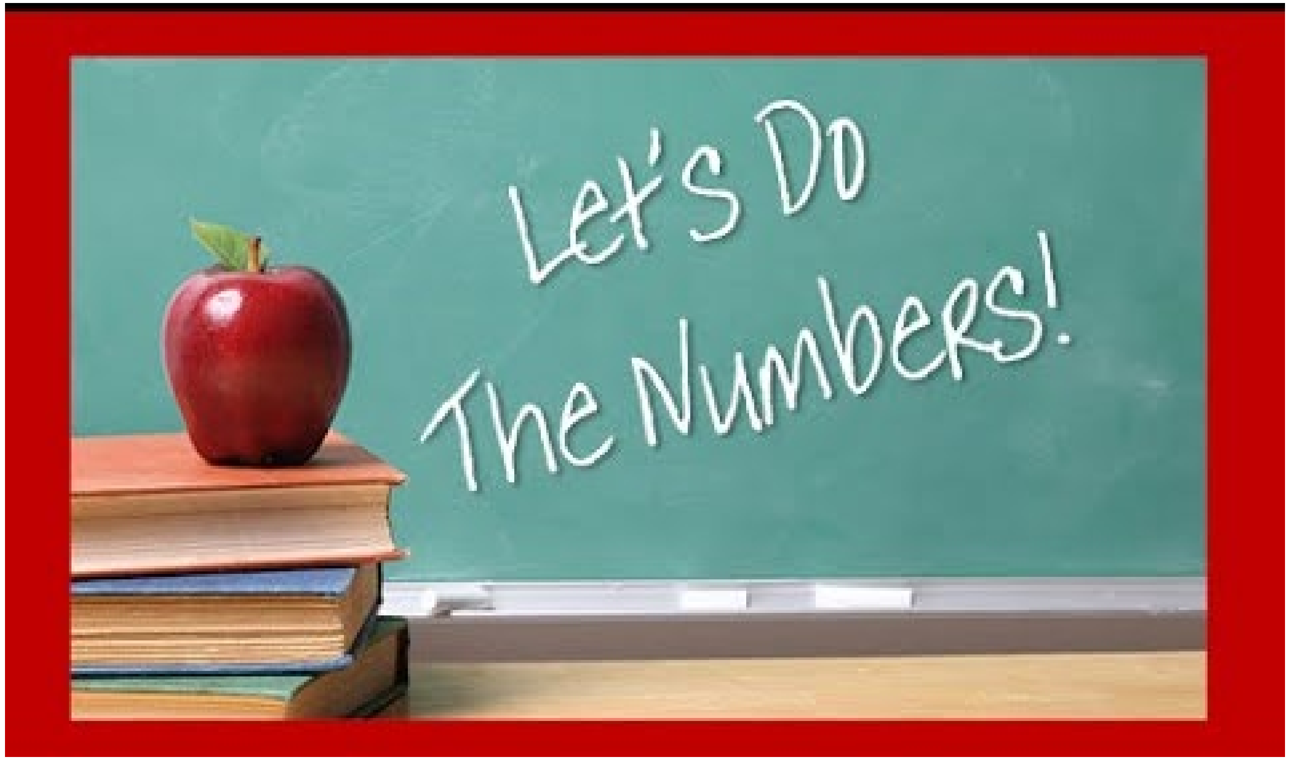
A joke? 1 cent ?

Bad business !!

You are fired !



https://www.marketplace.org/shows/marketplace/07092018/



Number of Supermarkets - 2018

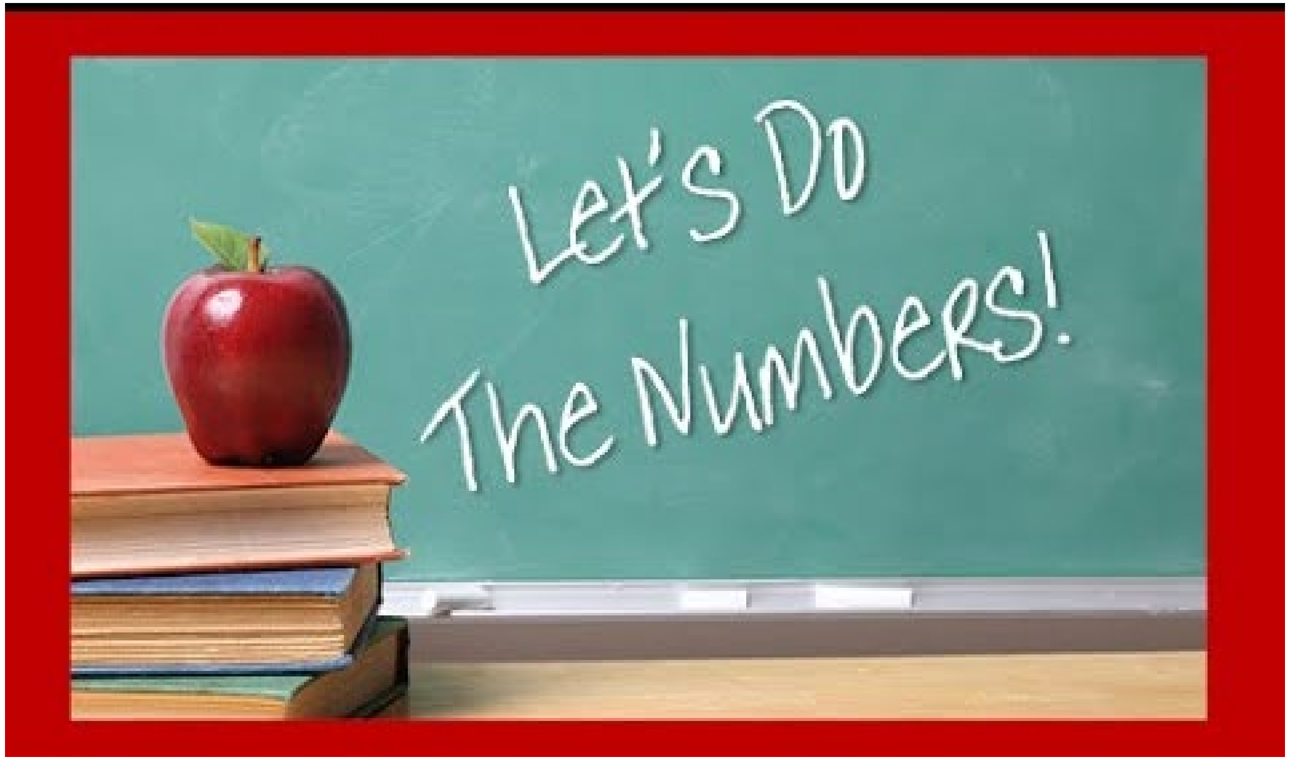
*2 million or more in annual sales

Source: Progressive Grocer Magazine

38,307

Transaction count is the metric you seek. Also known as customer count. It is the number of transactions per period. The most accessible of these is transactions per day. Even a smaller successful grocery store will approach 2000 transactions per a day. You can assume the actual foot traffic is higher as people shop together, but this number is not usually kept track of.

www.fmi.org/our-research/supermarket-facts
www.quora.com/What-is-the-average-foot-traffic-at-a-local-grocery-store



28 BILLION

ANNUAL GROCERY TRANSACTIONS IN US

Number of Supermarkets - 2018

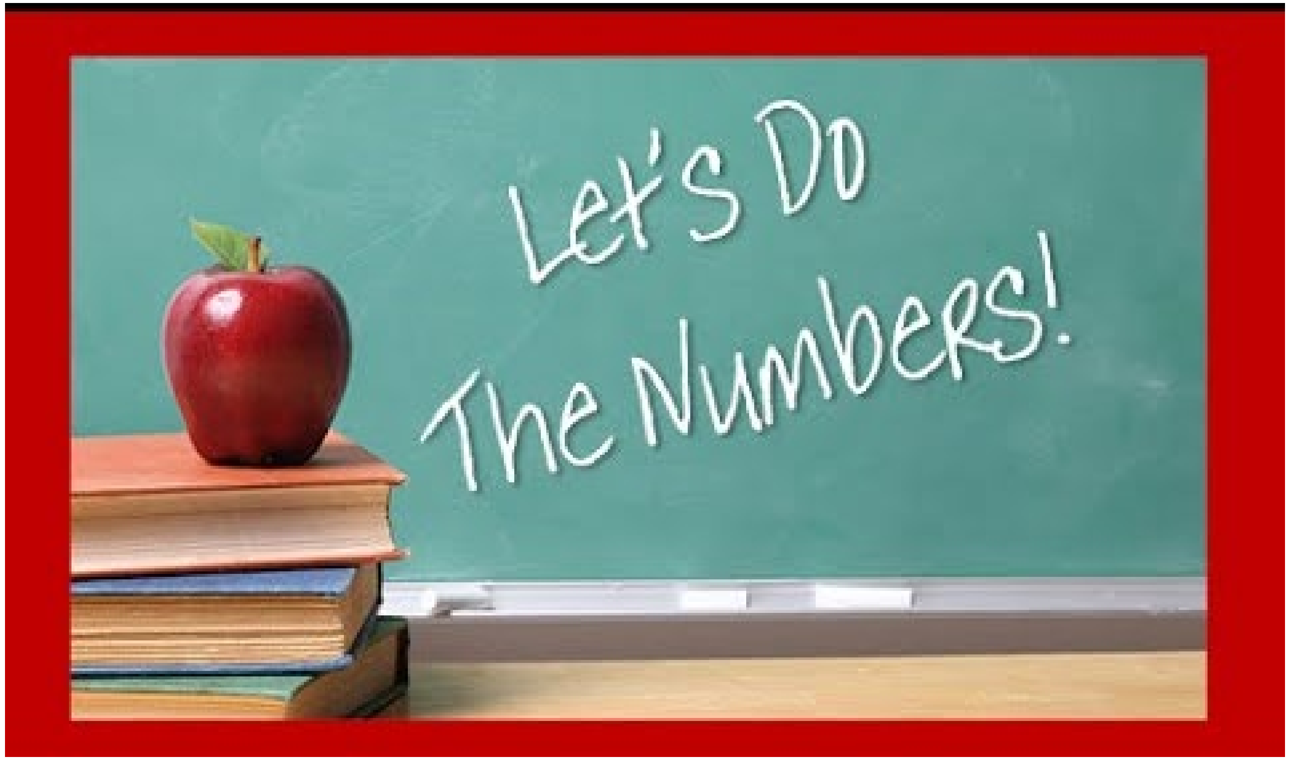
*2 million or more in annual sales

Source: Progressive Grocer Magazine

38,307

Transaction count is the metric you seek. Also known as customer count. It is the number of transactions per period. The most accessible of these is transactions per day. Even a smaller successful grocery store will approach 2000 transactions per a day. You can assume the actual foot traffic is higher as people shop together, but this number is not usually kept track of.

www.fmi.org/our-research/supermarket-facts
www.quora.com/What-is-the-average-foot-traffic-at-a-local-grocery-store



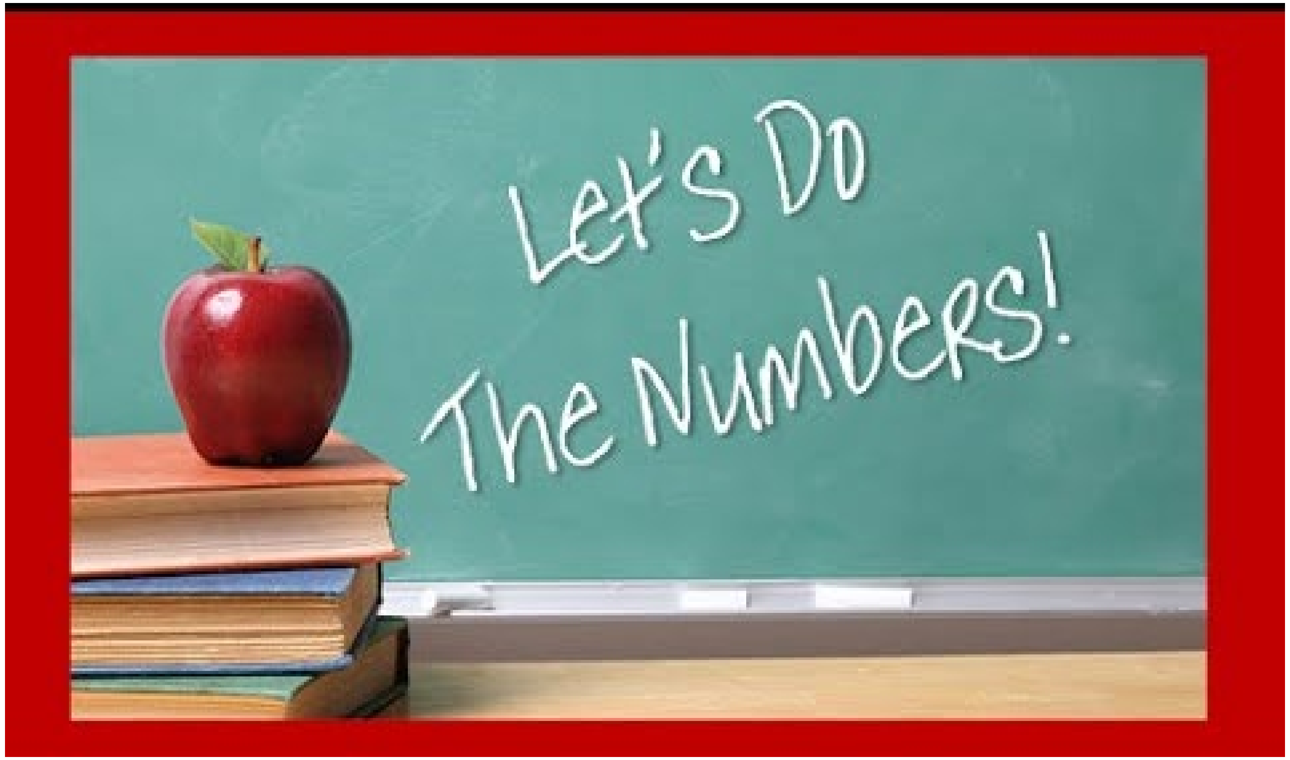
0.28 BILLION

ONLY 1% TRANSACTIONS BOUGHT CHKN AND CUSTOMER WILLING TO PAY 1 CENT

Number of Supermarkets - 2018	38,307
*2 million or more in annual sales	
Source: Progressive Grocer Magazine	

Transaction count is the metric you seek. Also known as customer count. It is the number of transactions per period. The most accessible of these is transactions per day. Even a smaller successful grocery store will approach 2000 transactions per a day. You can assume the actual foot traffic is higher as people shop together, but this number is not usually kept track of.

www.fmi.org/our-research/supermarket-facts
www.quora.com/What-is-the-average-foot-traffic-at-a-local-grocery-store



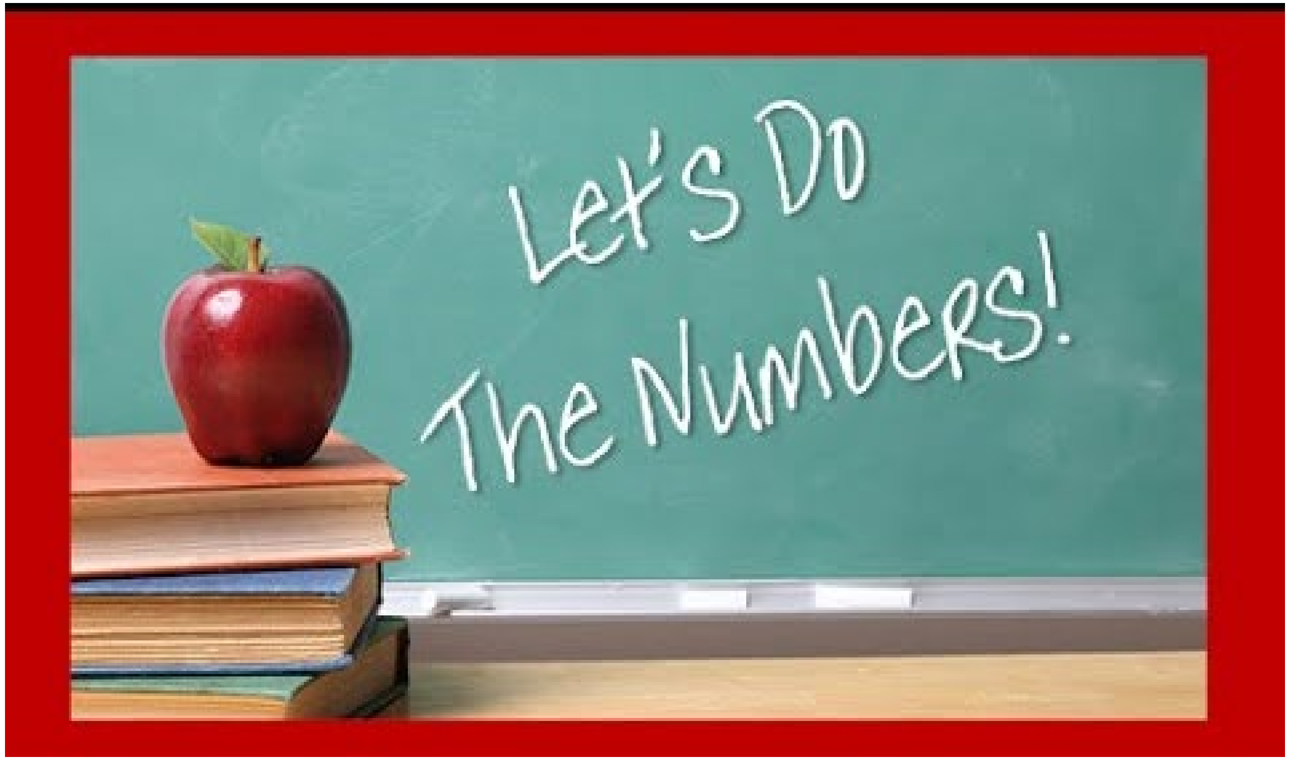
0.28 BILLION

ONLY 1% TRANSACTIONS BOUGHT CHKN
AND CUSTOMER WILLING TO PAY 1 CENT

FROM 1 USE - PAY A PENNY PER USE (PAPPU) - FOR ONLY 1 ITEM

$2.8 million

ANNUAL REVENUE ONLY IN USA



Average number items carried in a supermarket in 2017 **30,098**

Source: Food Marketing Institute

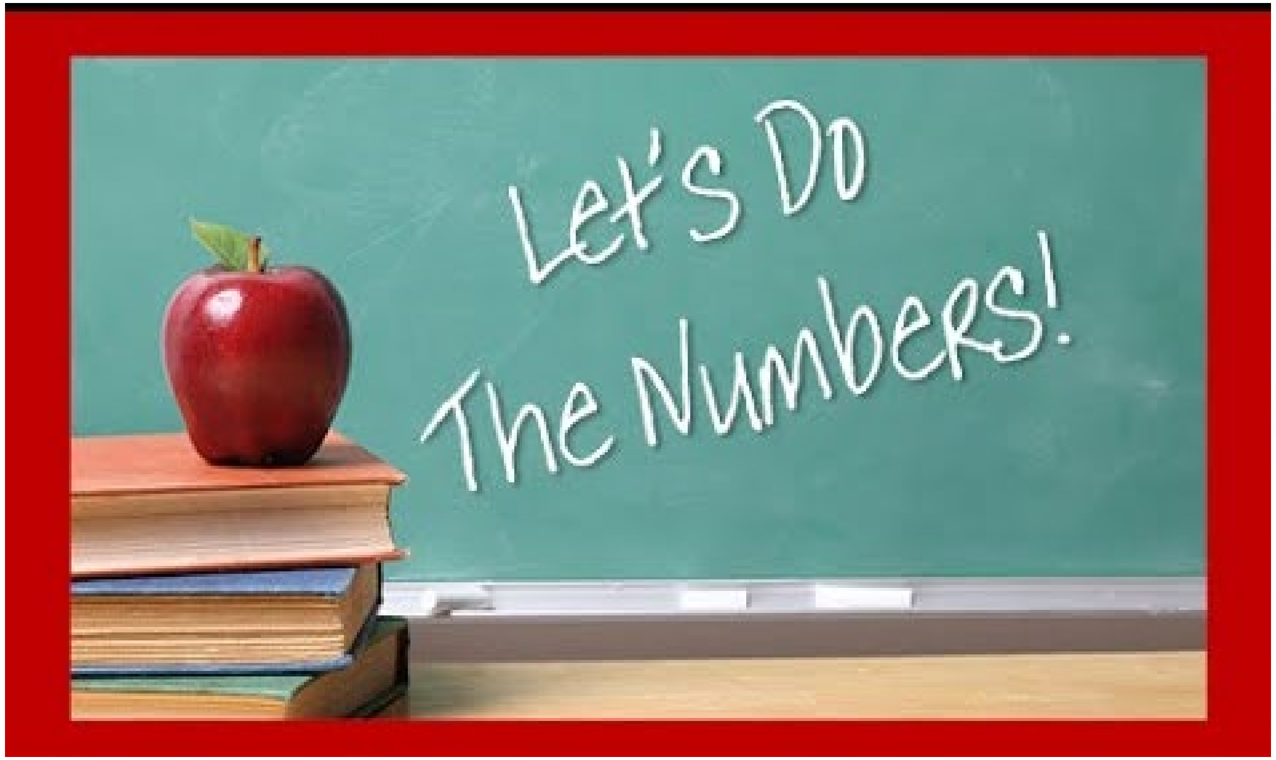
1% OF ITEMS USE FOOD SAFETY SENSOR

CUSTOMERS PAY 10 CENT/TRANSACTION

FOOD SAFETY - PAY A PENNY PER USE (PAPPU) - 10% TRANSACTIONS

$280 million

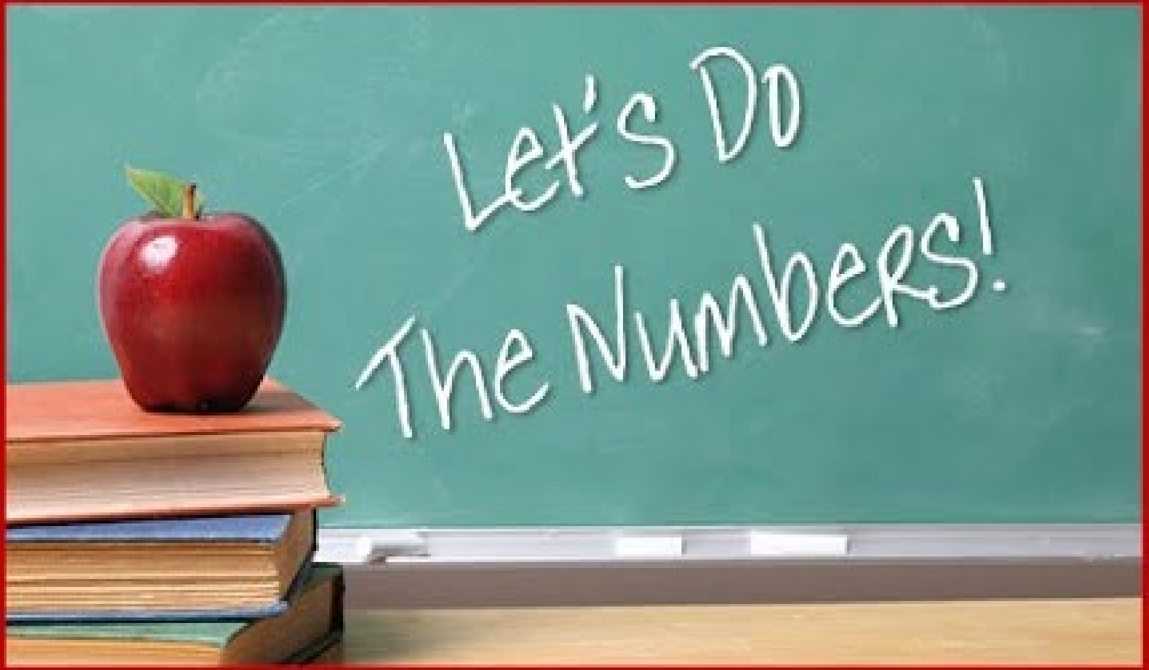
10% of 28 billion transactions pays $0.10 per transaction = $280 million revenue p.a.



FOOD SAFETY – PAY A PENNY PER USE (PAPPU) – GLOBAL POTENTIAL

Billion Dollar Industry?

ANNUAL GROSS REVENUE

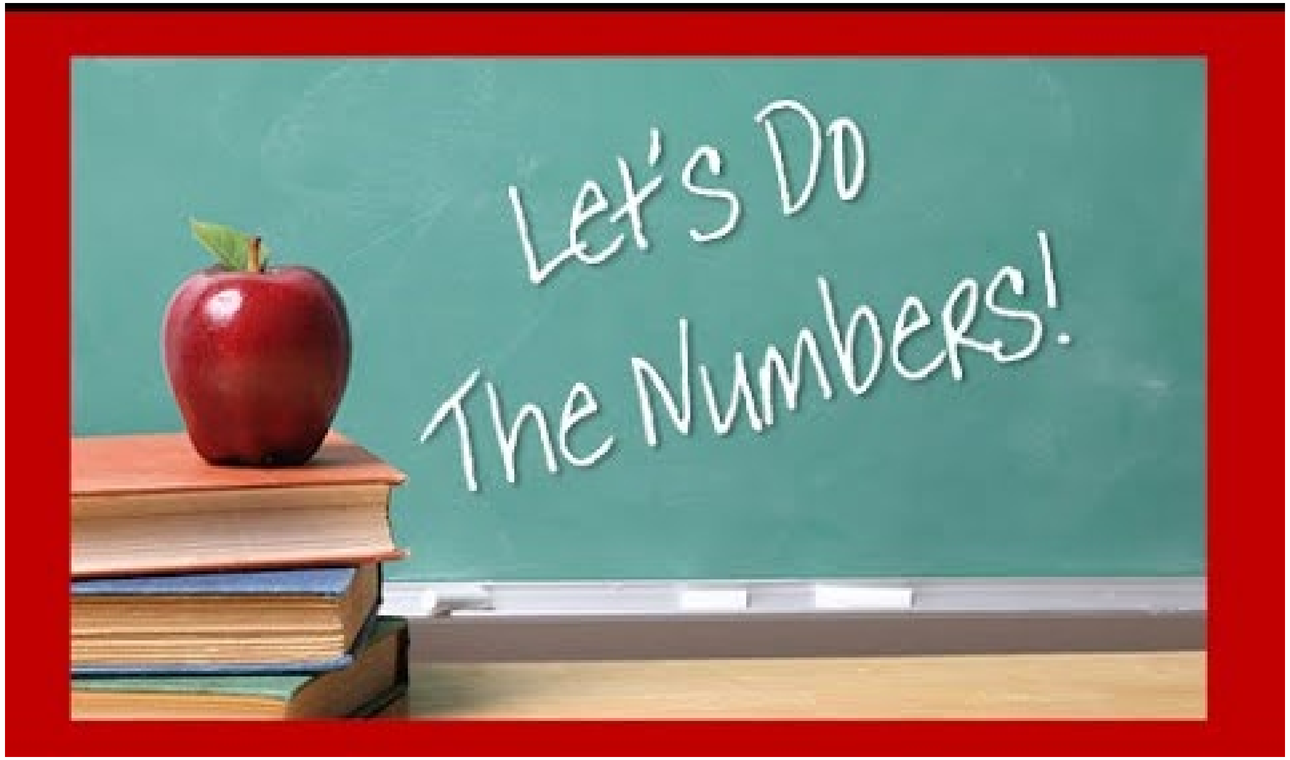


JUST
1 cent
PAPPU

FOOD SAFETY – PAY A PENNY PER USE (PAPPU) – GLOBAL POTENTIAL

Billion Dollar Industry

http://bit.ly/Economics-of-Technology
https://www.law.uchicago.edu/files/file/coase-nature.pdf
http://web.pdx.edu/~nwallace/EHP/TCEProgression.pdf
https://pdfs.semanticscholar.org/e4e8/a0486808360d056dbe212f7424273558538c.pdf
http://www.economics-ejournal.org/economics/discussionpapers/2007-3/at_download/file



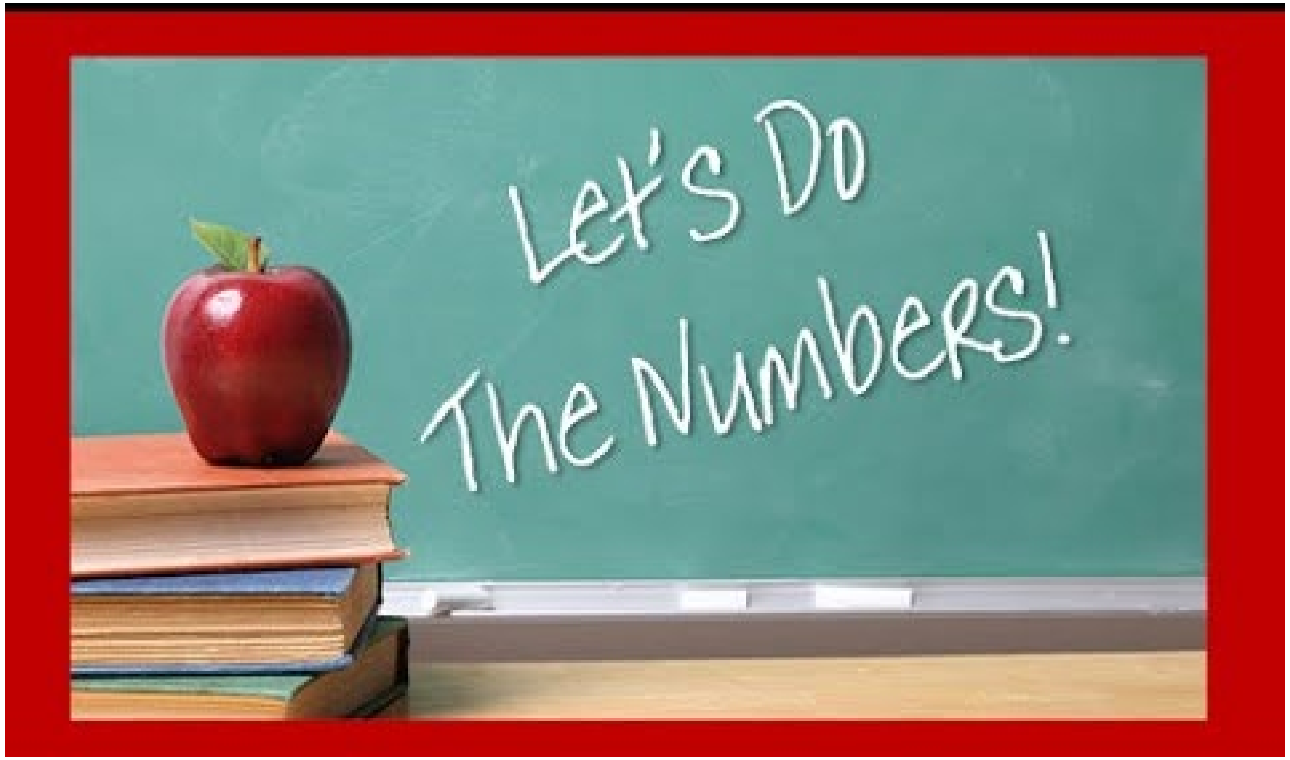
REALITY CHECK

Economies of scale may take several years to reach market penetration to generate mega revenues from nano payments, for example, the PAPPU model.

FOOD SAFETY – PAY A PENNY PER USE (PAPPU) – GLOBAL POTENTIAL

Billion Dollar Industry

THINK FOOD SAFETY AS PAY PER USE PERSONALIZED SERVICE



PAY A PENNY PER USE GLOBAL SERVICES

HPE CEO Pledges to Sell 'Everything as a Service' by 2022

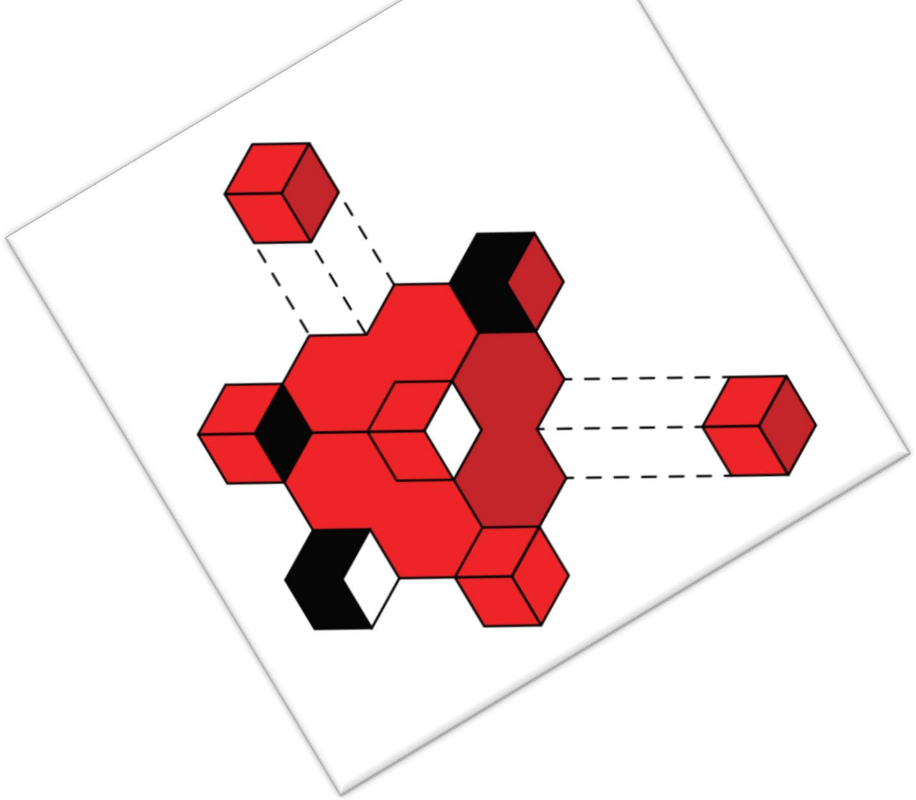
In its boldest move yet to make on-prem IT more like public cloud, the company says GreenLake is its future.

Yevgeniy Sverdlik | Jun 19, 2019

Three years from now, every product Hewlett Packard Enterprise sells will be available as a service. That's the pledge CEO Antonio Neri made from stage Tuesday afternoon during his keynote at the company's Discover conference in Las Vegas. The pledge covers both hardware and software in the enterprise tech giant's sprawling portfolio.

THINK FOOD SAFETY AS PAY PER USE PERSONALIZED SERVICE

https://www.datacenterknowledge.com/hewlett-packard-enterprise/hpe-ceo-pledges-sell-everything-service-2022



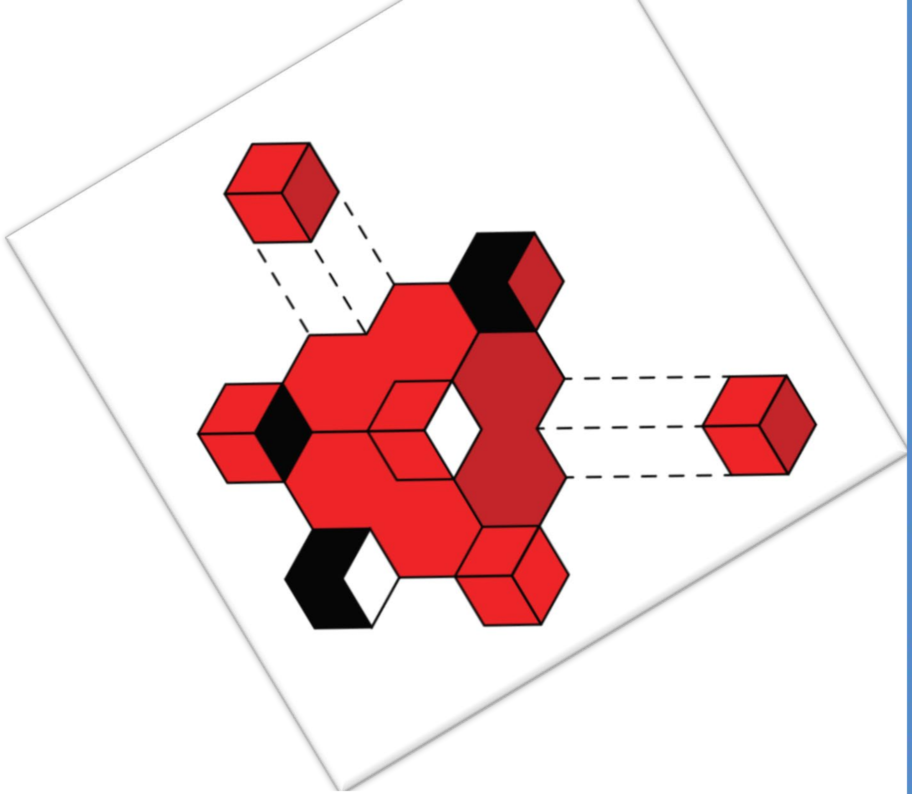
WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Insight Report

Top 10 Emerging Technologies 2019

http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf



WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Insight Report

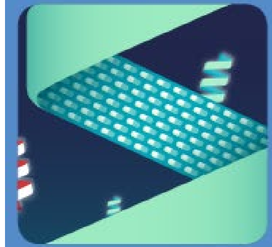
Top 10 Emerging Technologies 2019

http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf

1. Bioplastics for a Circular Economy

2. Social Robots

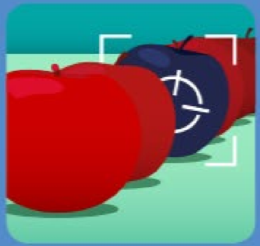
3. Tiny Lenses for Miniature Devices



4. Disordered Proteins as Drug Targets

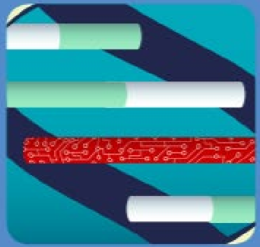
5. Smarter Fertilizers Can Reduce Environmental Contamination

6. Collaborative Telepresence



7. Advanced Food Tracking and Packaging

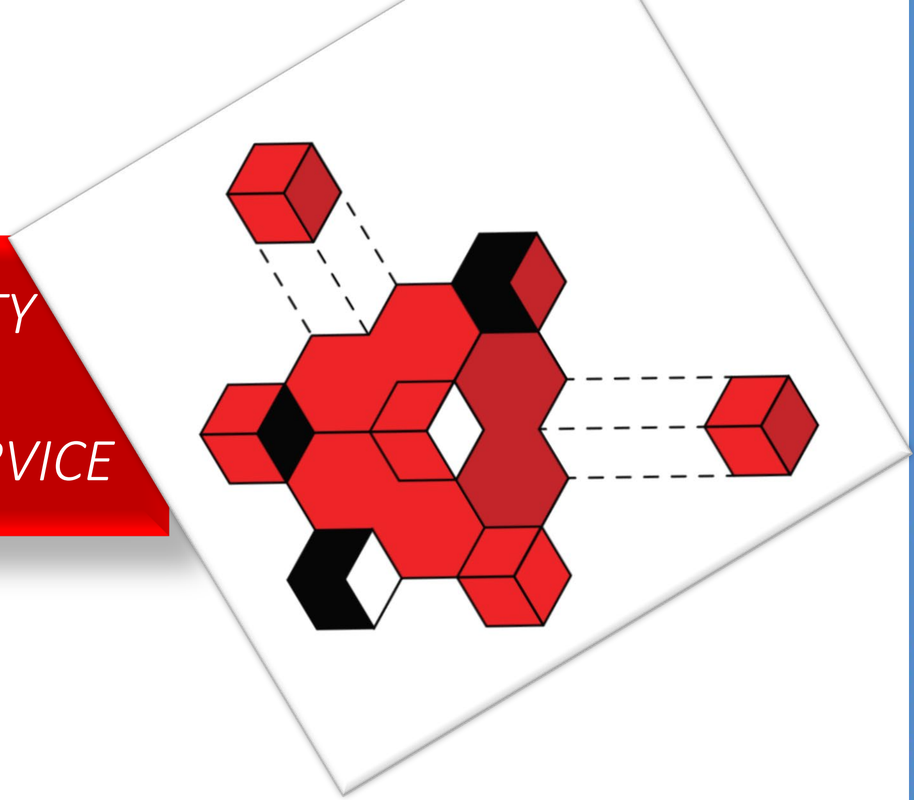
8. Safer Nuclear Reactors



9. DNA Data Storage

10. Utility-Scale Storage of Renewable Energy

THINK FOOD SAFETY AS A PAY PER USE PERSONALIZED SERVICE



WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Insight Report

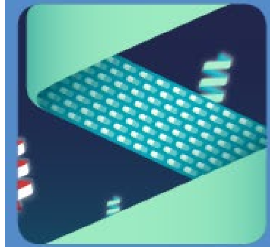
Top 10 Emerging Technologies 2019

http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf

1. Bioplastics for a Circular Economy

2. Social Robots

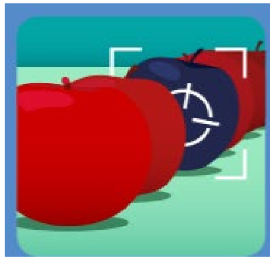
3. Tiny Lenses for Miniature Devices



4. Disordered Proteins as Drug Targets

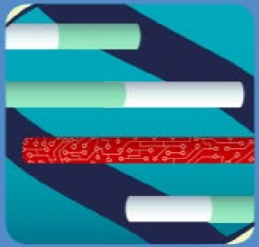
5. Smarter Fertilizers Can Reduce Environmental Contamination

6. Collaborative Telepresence



7. Advanced Food Tracking and Packaging

8. Safer Nuclear Reactors



9. DNA Data Storage

10. Utility-Scale Storage of Renewable Energy

HPE CEO Pledges to Sell 'Everything as a Service' by 2022

In its boldest move yet to make on-prem IT more like public cloud, the company says GreenLake is its future.

Yevgeniy Sverdlik | Jun 19, 2019

Three years from now, every product Hewlett Packard Enterprise sells will be available as a service. That's the pledge CEO Antonio Neri made from stage Tuesday afternoon during his keynote at the company's Discover conference in Las Vegas. The pledge covers both hardware and software in the enterprise tech giant's sprawling portfolio.

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

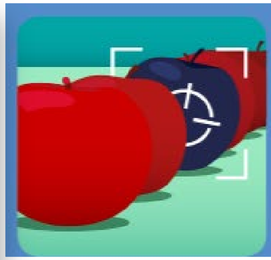
Insight Report

Top 10 Emerging Technologies 2019

THINK FOOD SAFETY AS A PAY PER USE PERSONALIZED SERVICE

1. Bioplastics for a Circular Economy

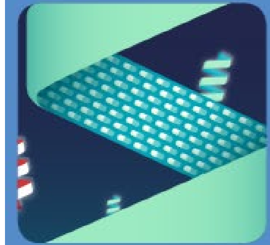
6. Collaborative Telepresence



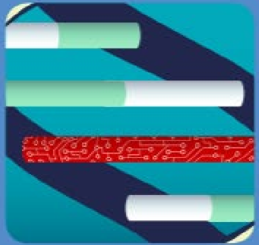
7. Advanced Food Tracking and Packaging

3. Tiny Lenses for Miniature Devices

8. Safer Nuclear Reactors



4. Disordered Proteins as Drug Targets



9. DNA Data Storage

5. Smarter Fertilizers Can Reduce Environmental Contamination

10. Utility-Scale Storage of Renewable Energy

FOOD ART ?

In a Grocery Store Near You

FRESH SENSE™

CHECK WITH YOUR FOOD SAFETY APP

FRESH SENSE – A TRADEMARKED SERVICE FROM SHOP-RITE TO GUARANTEE YOUR FRESHNESS IN REAL TIME ON YOUR PHONE

FRESH SENSE™

CHECK WITH YOUR FOOD SAFETY APP

SENSEE

SMALL PART BUT USEFUL

FRESH SENSE – A TRADEMARKED SERVICE FROM SHOP-RITE TO GUARANTEE YOUR FRESHNESS IN REAL TIME ON YOUR PHONE

FRESH SENSE™

FRESH SENSE – A TRADEMARKED SERVICE FROM SHOP-RITE TO GUARANTEE YOUR FRESHNESS IN REAL TIME ON YOUR PHONE

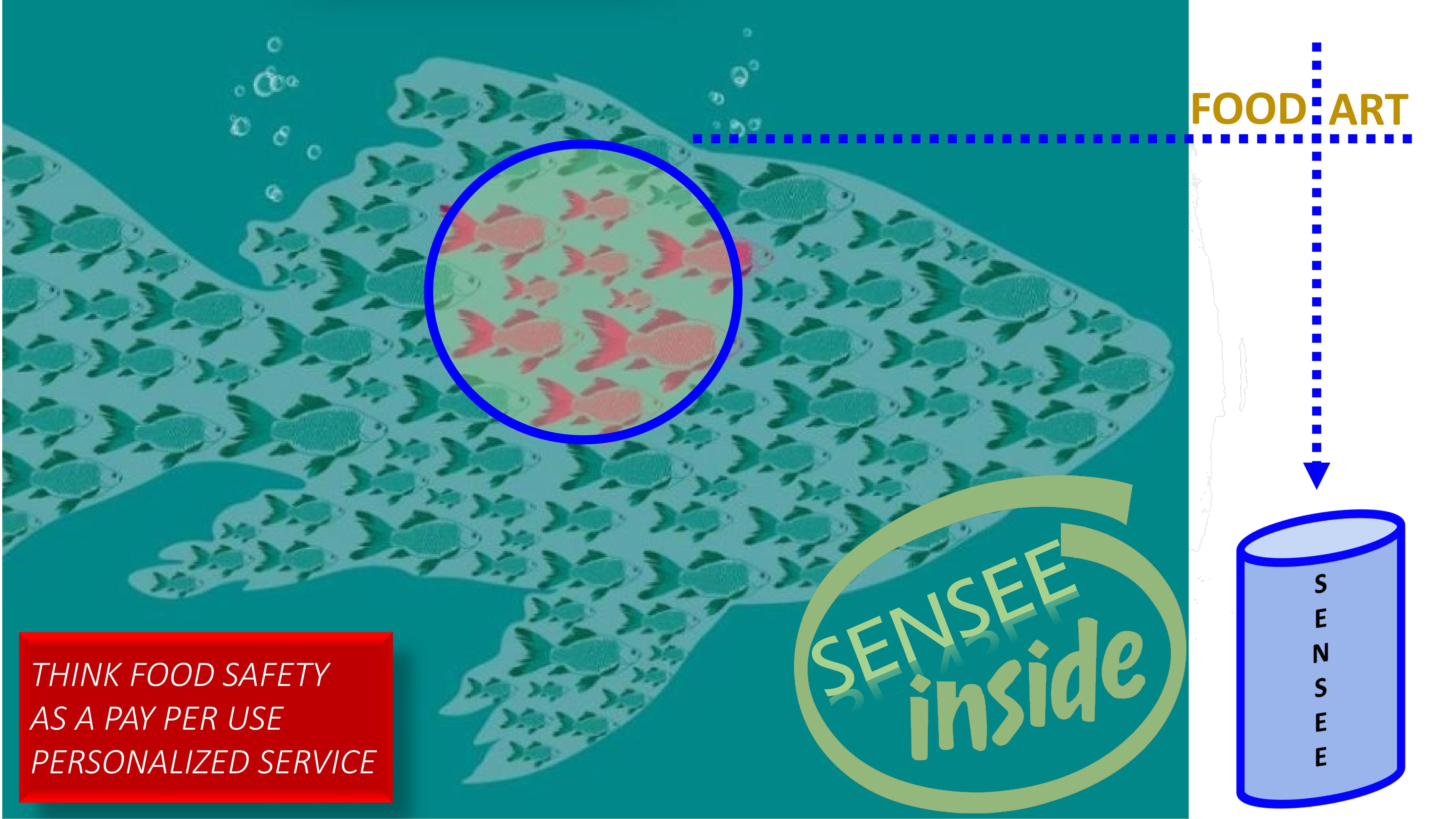
FRESH SENSE™

KIDS

FRESH SENSE – A TRADEMARKED SERVICE FROM SHOP-RITE TO GUARANTEE YOUR FRESHNESS IN REAL TIME ON YOUR PHONE

FRESH SENSE – A TRADEMARKED SERVICE FROM SHOP-RITE TO GUARANTEE YOUR FRESHNESS IN REAL TIME ON YOUR PHONE


FRESH SENSE™
CHECK WITH YOUR FOOD SAFETY APP
FRUIT & VEG
ALDI
Dole fresh
BABY SPINACH
FOOD SAFETY APP
SKU 56789
PROD SPNH
QUALITY
SENSE
FOOD SAFETY APP
SKU 98765
PROD FISH
QUALITY


FOOD ART
SENSEE
SENSEE inside
THINK FOOD SAFETY
AS A PAY PER USE
PERSONALIZED SERVICE

Summary – A Sense of the Future

SENSEE, SNAPS, ART, DIDA'S, KIDS

Are these ingredients for a Google of Ag and/or fuel for future scientific research ?

PEAS

Platform for the Agro-Ecosystem

A few lofty goals, perhaps best attempted in stages, from data to data-informed, with knowledge-informed as a future performance index (KPI). Granularity of data from sensors feed decisions with information and knowledge. In the short term, offer logic tools (ART) for users, reduce food waste and contribute to food safety.

Being KIDS – Is "knowledge-informed" a challenging goal?

Knowledge combination/integration beyond (heterogenous) rules and ontologies are not only difficult, but calls for **new thinking**. The semantics of knowledge bases other than rules (for example, descriptions of temporal processes like workflows which could logically decide when the irrigation system must turn on/off the water pumps, or protocols in spatio-temporal logic) must be integrated. We may need some form of logic framework in which knowledge modules, having different native semantics, can be overlaid with meaningful semantics, preferably agnostic of linguistic bias, ideally as a "plug and play" operation, graph-friendly for non-expert end-users to decompose and/or re-compose the choice of logic and logic tools, based on experience from expert humans in the loop. Chaperoning convergence between distributed domain(s) knowledge, operational rules, data, information, and systems science, is a daunting and challenging goal.

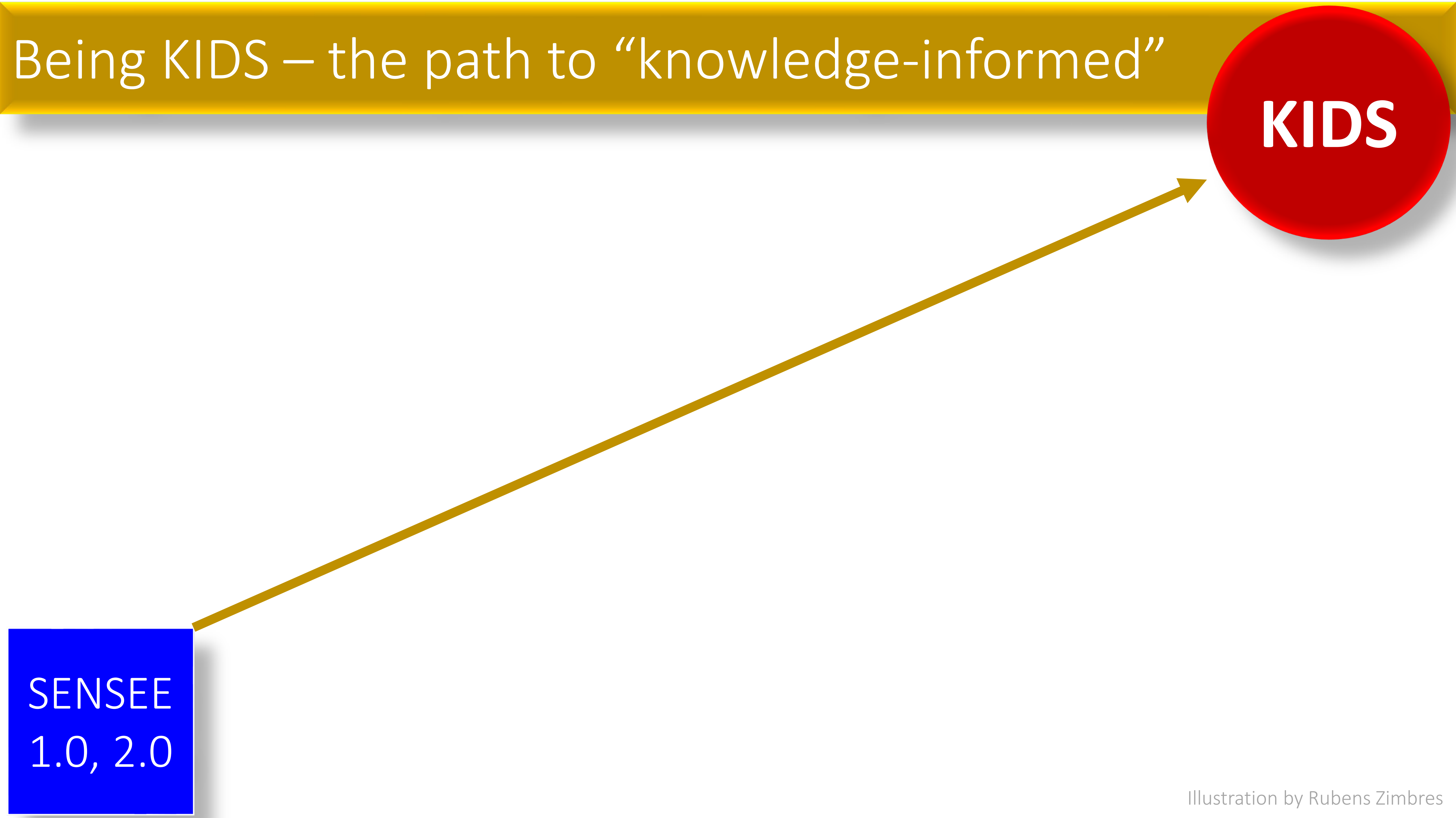
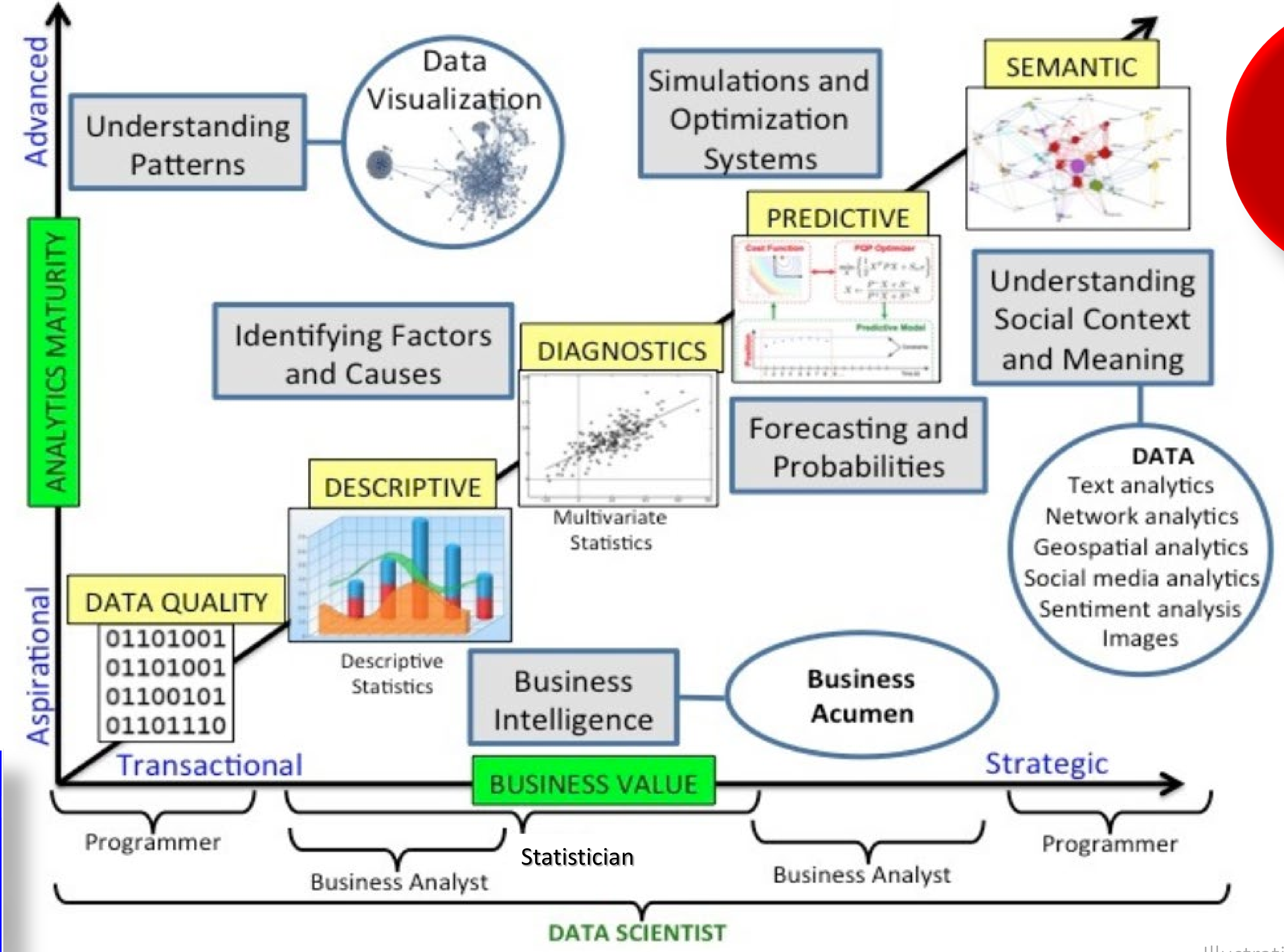


Illustration by Rubens Zimbres

KIDS

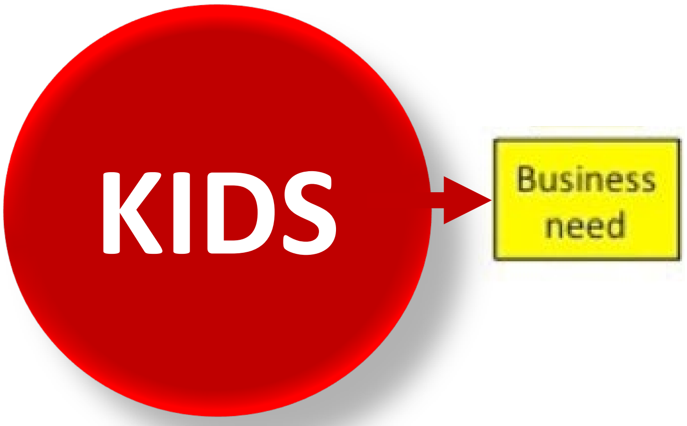


SENSEE 1.0, 2.0

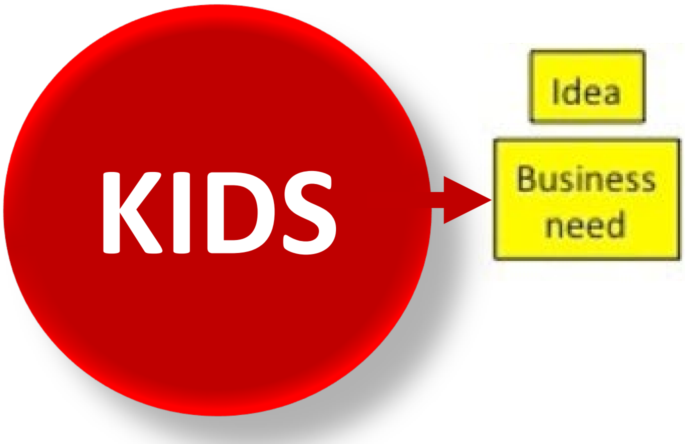
Illustration by Rubens Zimbres

KIDS → **USER**

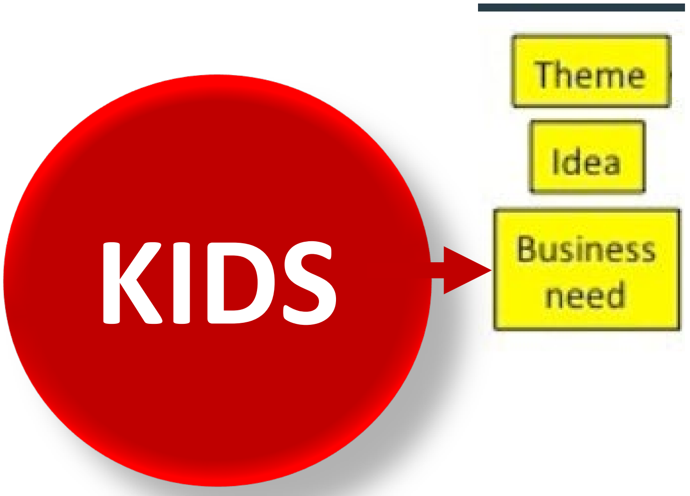
In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.



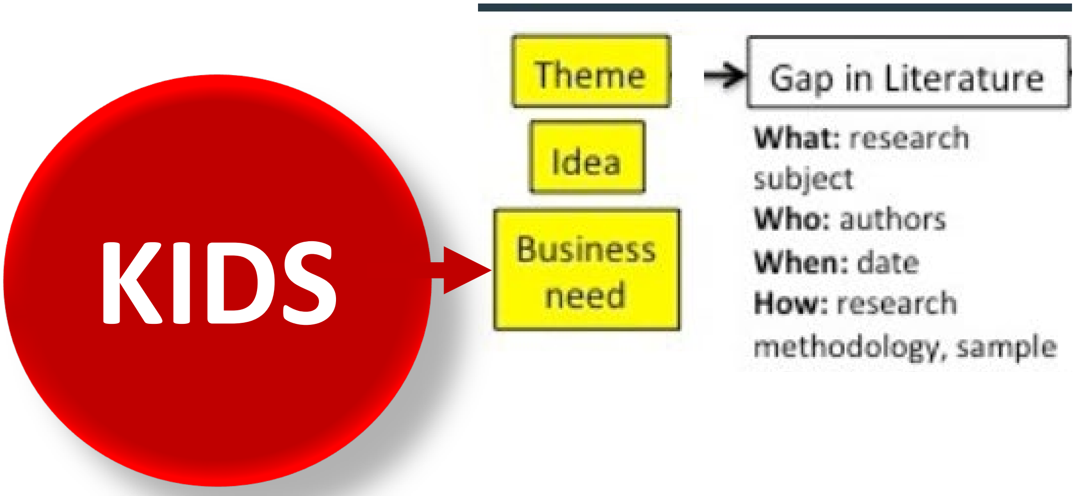
In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.



In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.

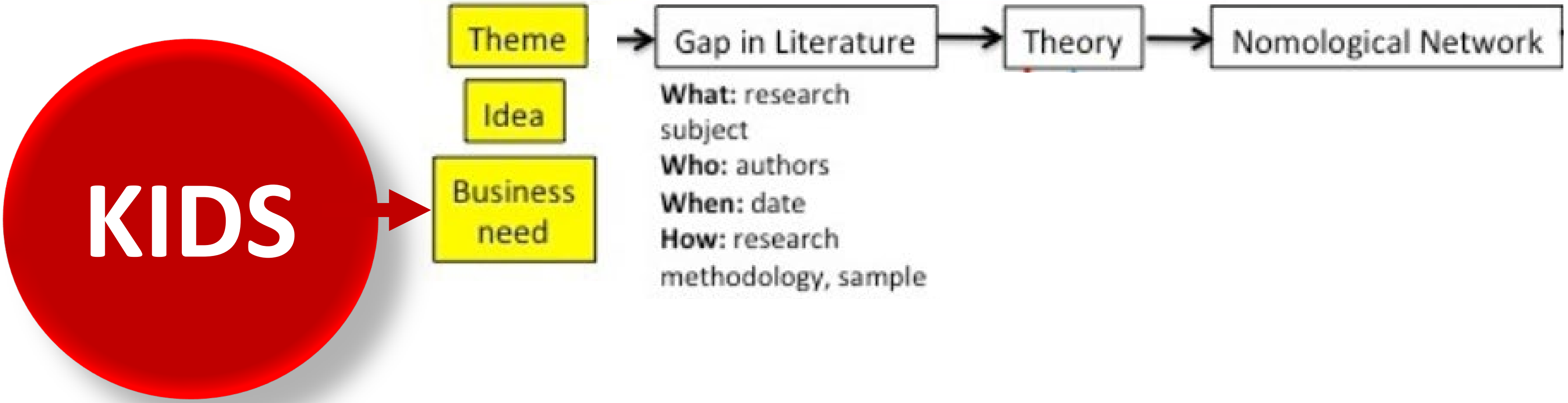


In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.



In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.

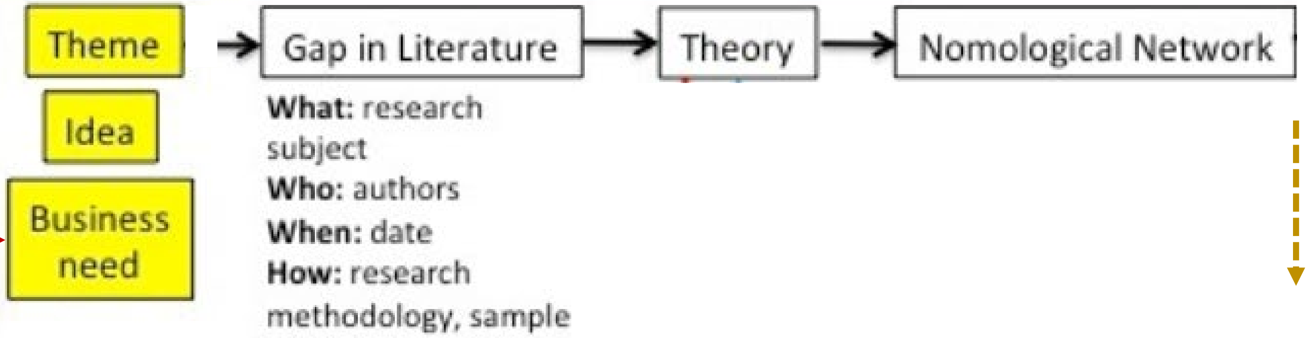
Illustration by Rubens Zimbres



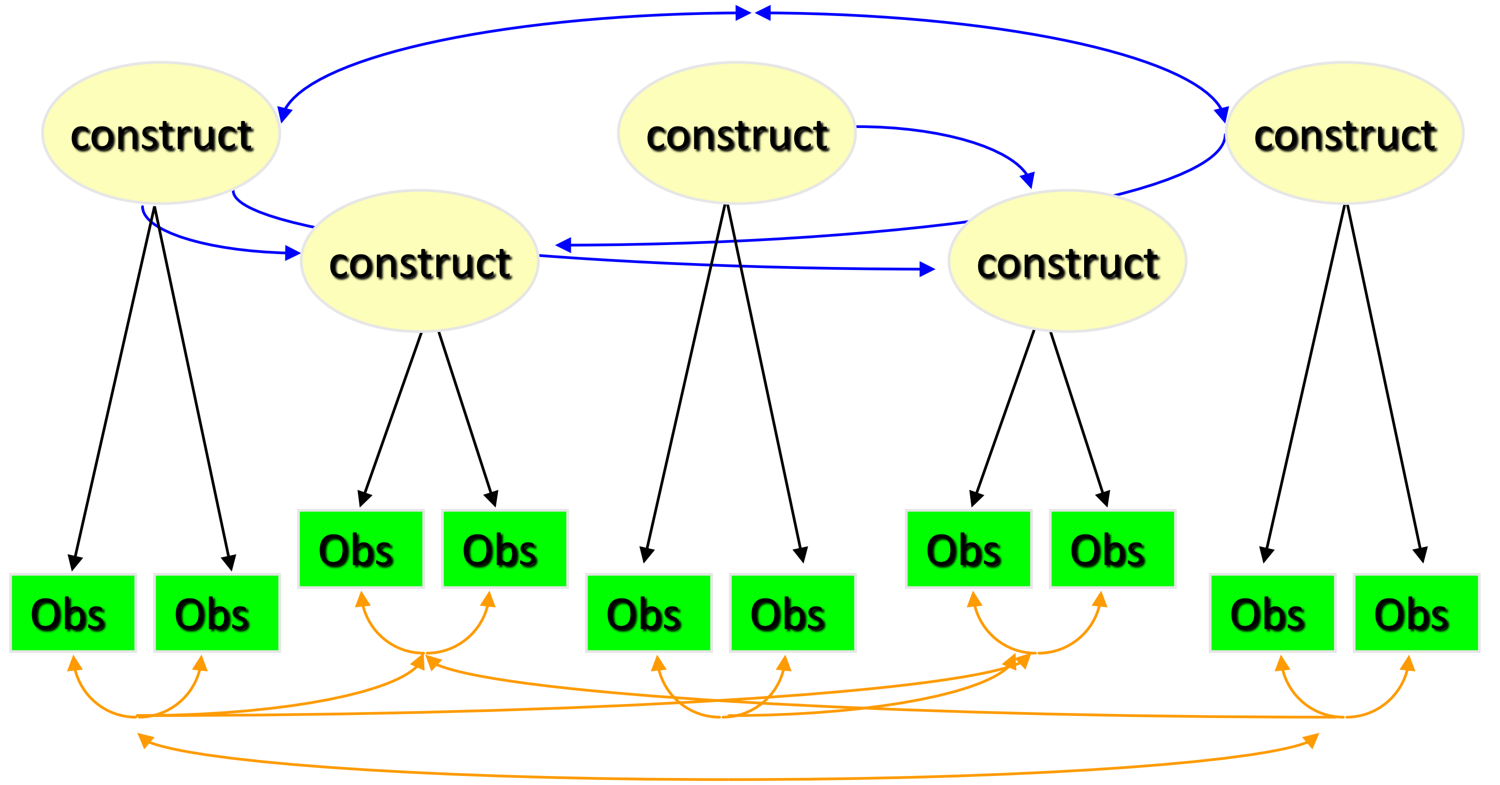
In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.

Illustration by Rubens Zimbres

KIDS

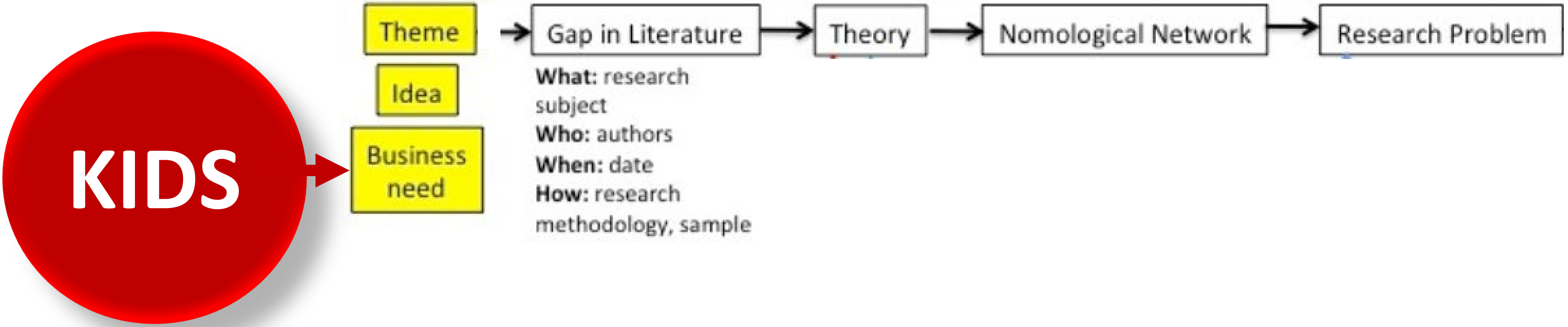


In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.



An idea circa 1955, the nomological net represents concepts (constructs) of interest, their observable manifestations and interrelationships. It appears to share common grounds with system dynamics and knowledge graphs.

https://psychclassics.yorku.ca/Cronbach/construct.htm ■ http://dx.doi.org/10.1037/h0040957 ■ https://conservancy.umn.edu/bitstream/handle/11299/184279/1_07_Cronbach.pdf?sequence=1



In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.

Illustration by Rubens Zimbres

KIDS

In pursuing the knowledge-informed paradigm using the PEAS platform, users will be confronted with questions that they cannot answer. Unmet needs may fuel research. Hence, KIDS is a catalyst for new ideas, innovation, research in science and engineering.

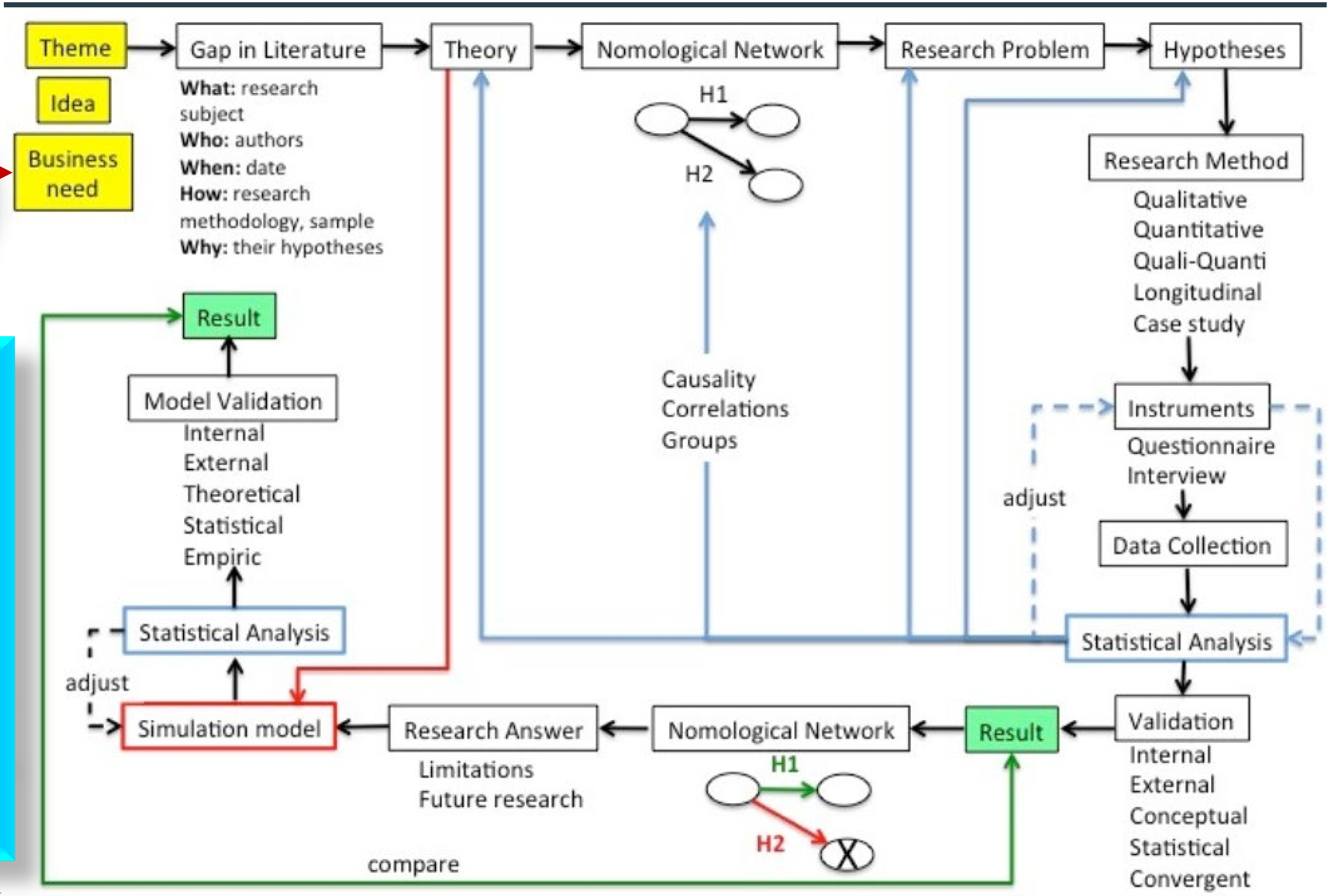
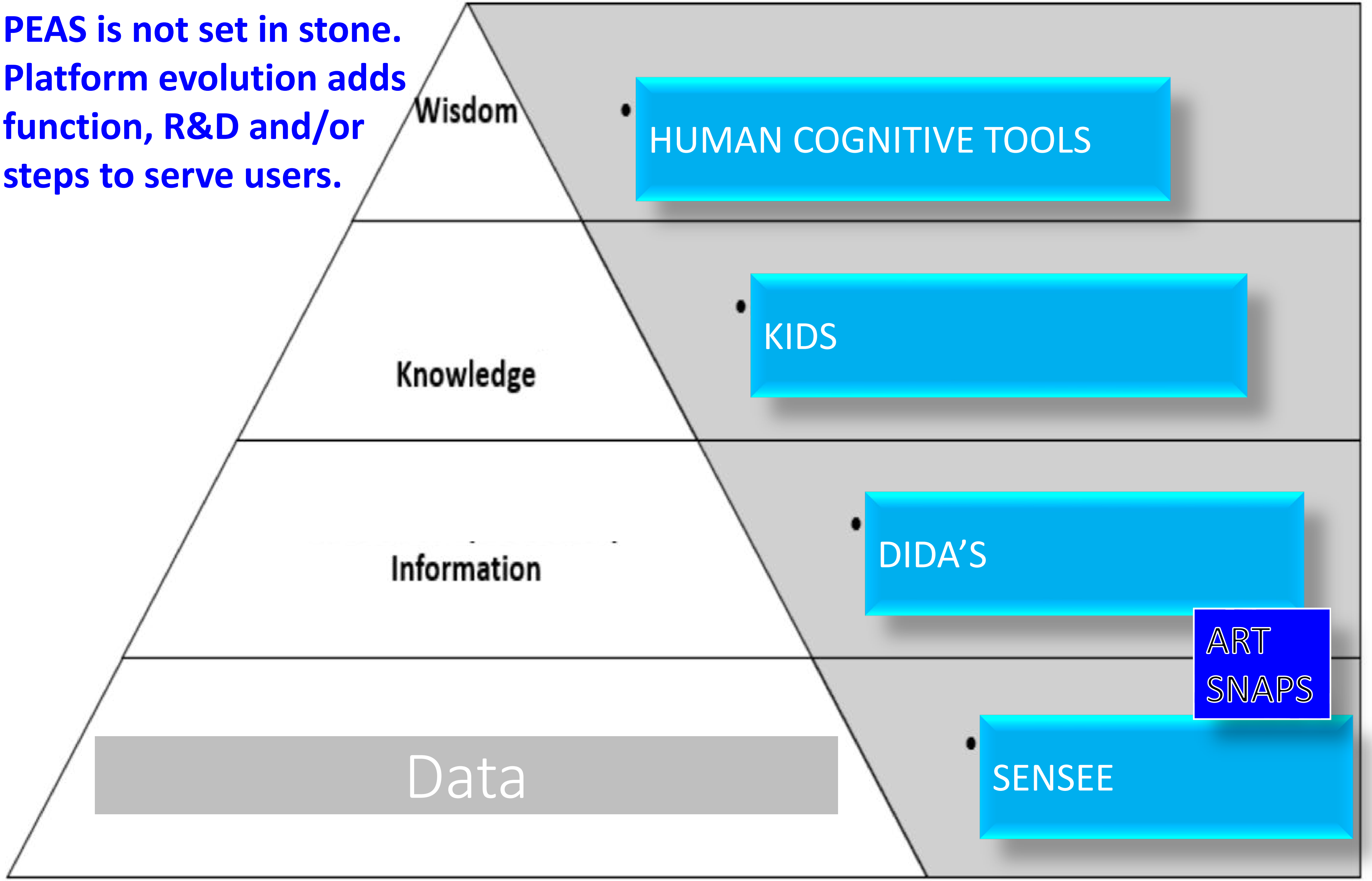
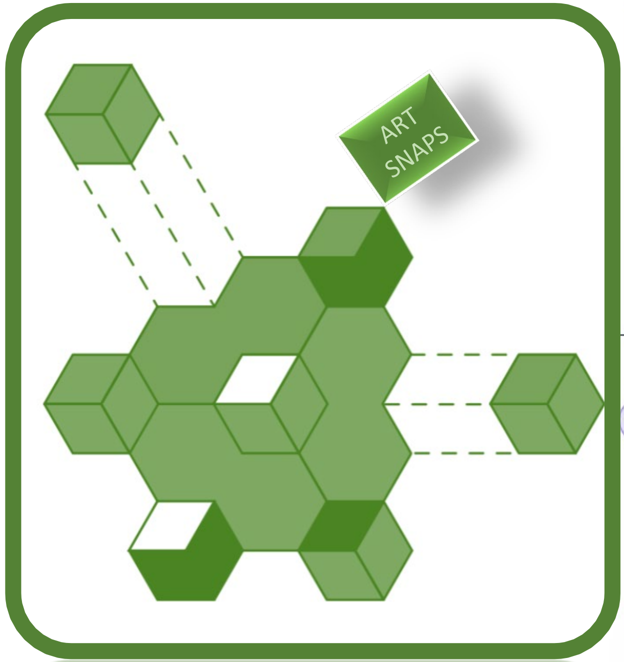


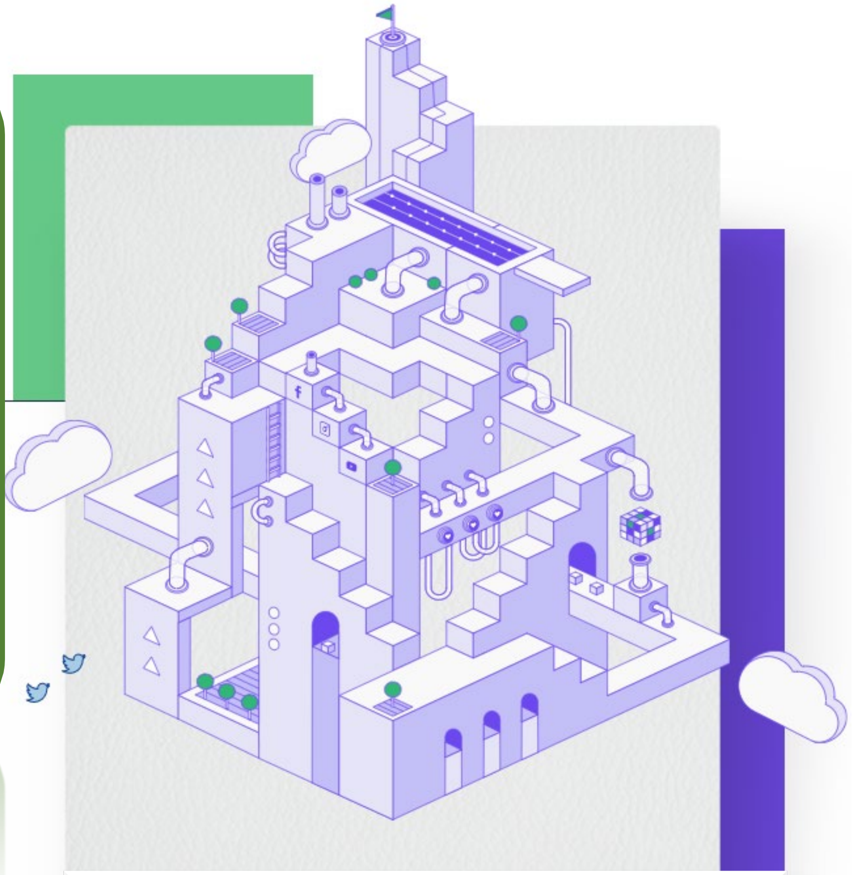
Illustration by Rubens Zimbres

PEAS PLATFORM

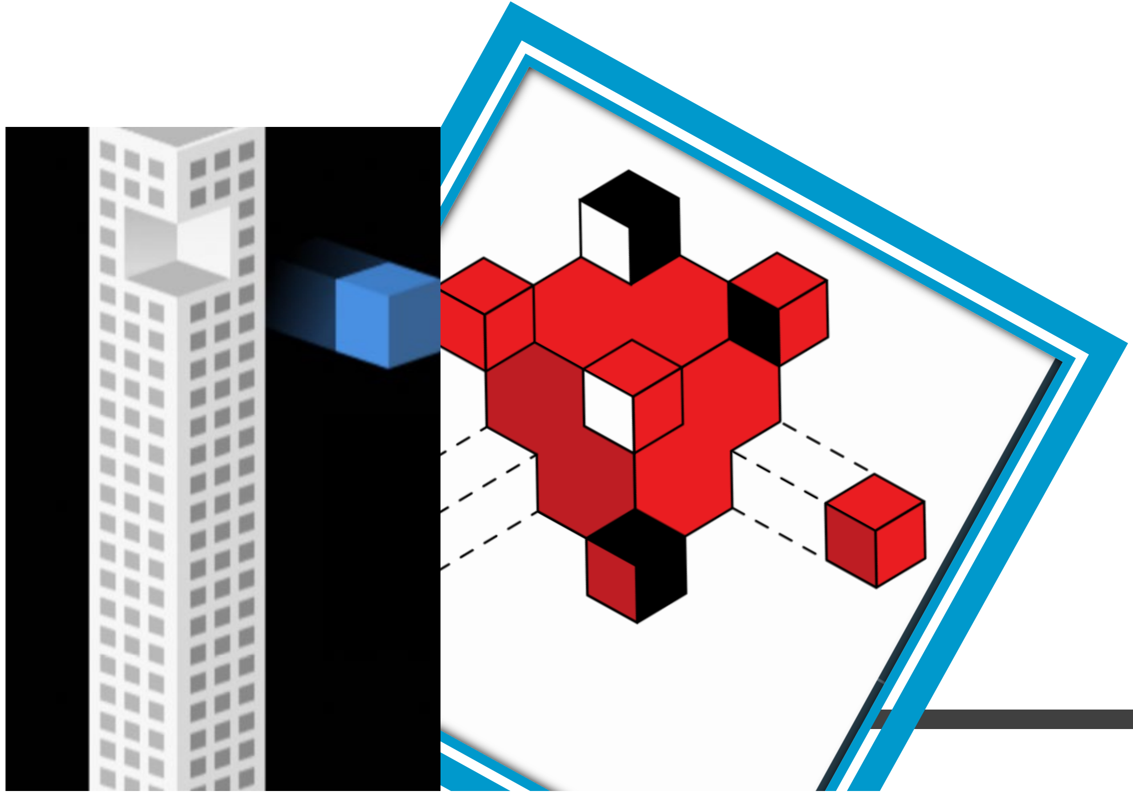
PEAS is not set in stone. Platform evolution adds function, R&D and/or steps to serve users.







KNOWLEDGE SYSTEMS SOLUTIONS
ARE OFTEN MULTI-LAYERED IDEAS



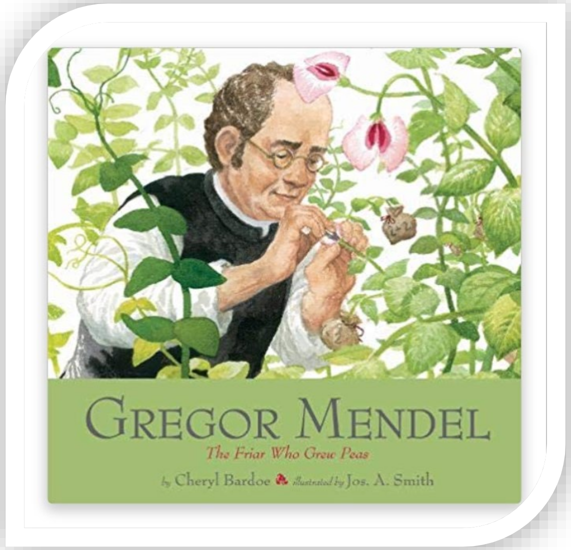
WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Insight Report

Top 10 Emerging Technologies 2019

http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf



7. Advanced Food Tracking and Packaging

5. Smarter Fertilizers Can Reduce Environmental Contamination

SUMMARY

1) Data ≠ Information ≠ Knowledge
2) Develop portfolio of ART (pareto solutions - logic tools - for the next billion users)
3) Context determines the perishability phase of data > information > decision > knowledge
4) Relationships must be relative to context before connecting relevant contextual data (R2C2)

Suggested topics for in-depth exploration:

[0] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-441-information-theory-spring-2010/

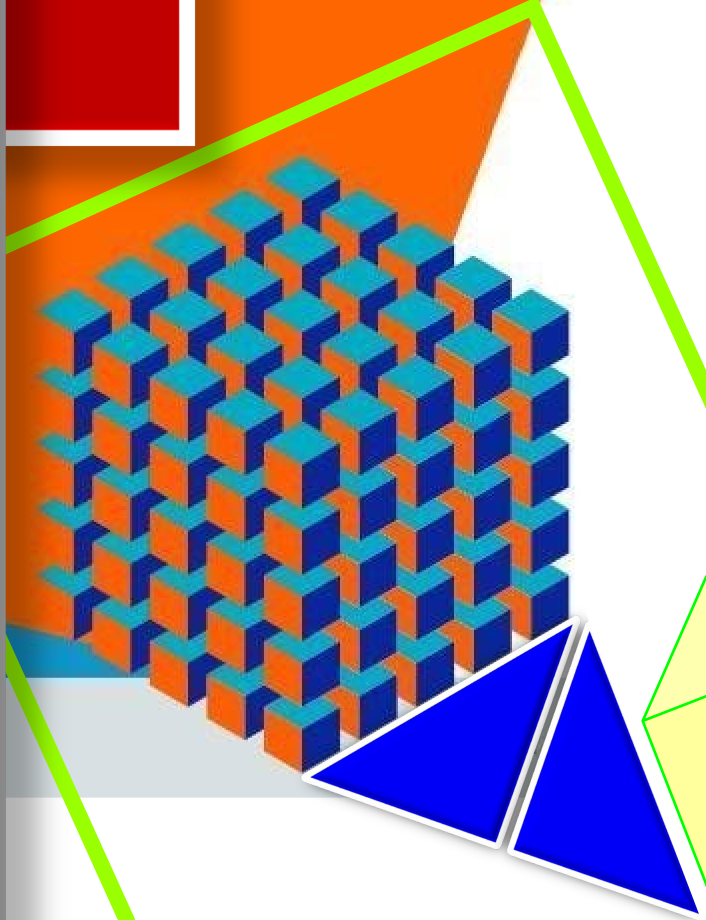
[1] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/lecture-videos/index.htm

[2] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-spring-2005/

[3] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/lecture-videos/

[4] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-825-techniques-in-artificial-intelligence-sma-5504-fall-2002/

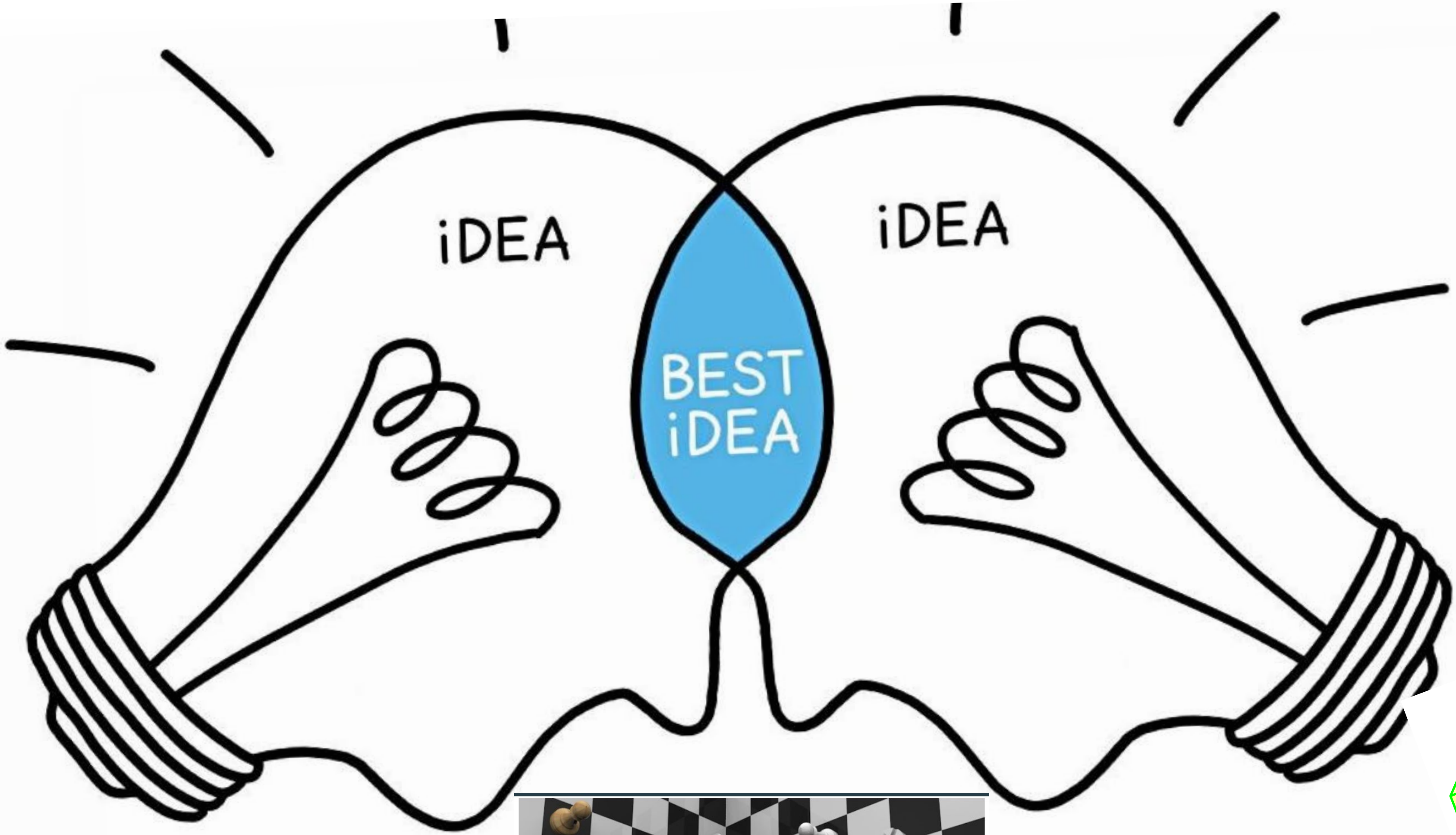
[5] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-868j-the-society-of-mind-fall-2011/video-lectures/

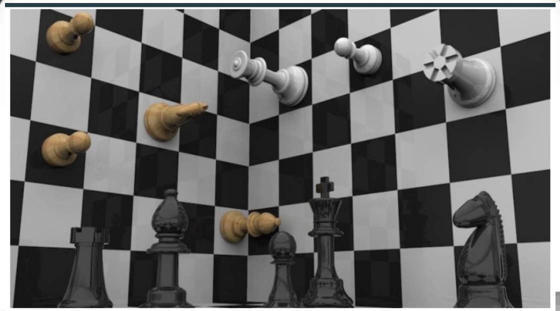


Professor Eric S McLamore
ABE, University of Florida
USDA Center of Excellence

Dr Shoumen Palit Austin Datta
http://bit.ly/SIGNALS-SIGNALS

"If your actions inspire others to dream more, learn more, do more and become more, then you are an enabler." JQA



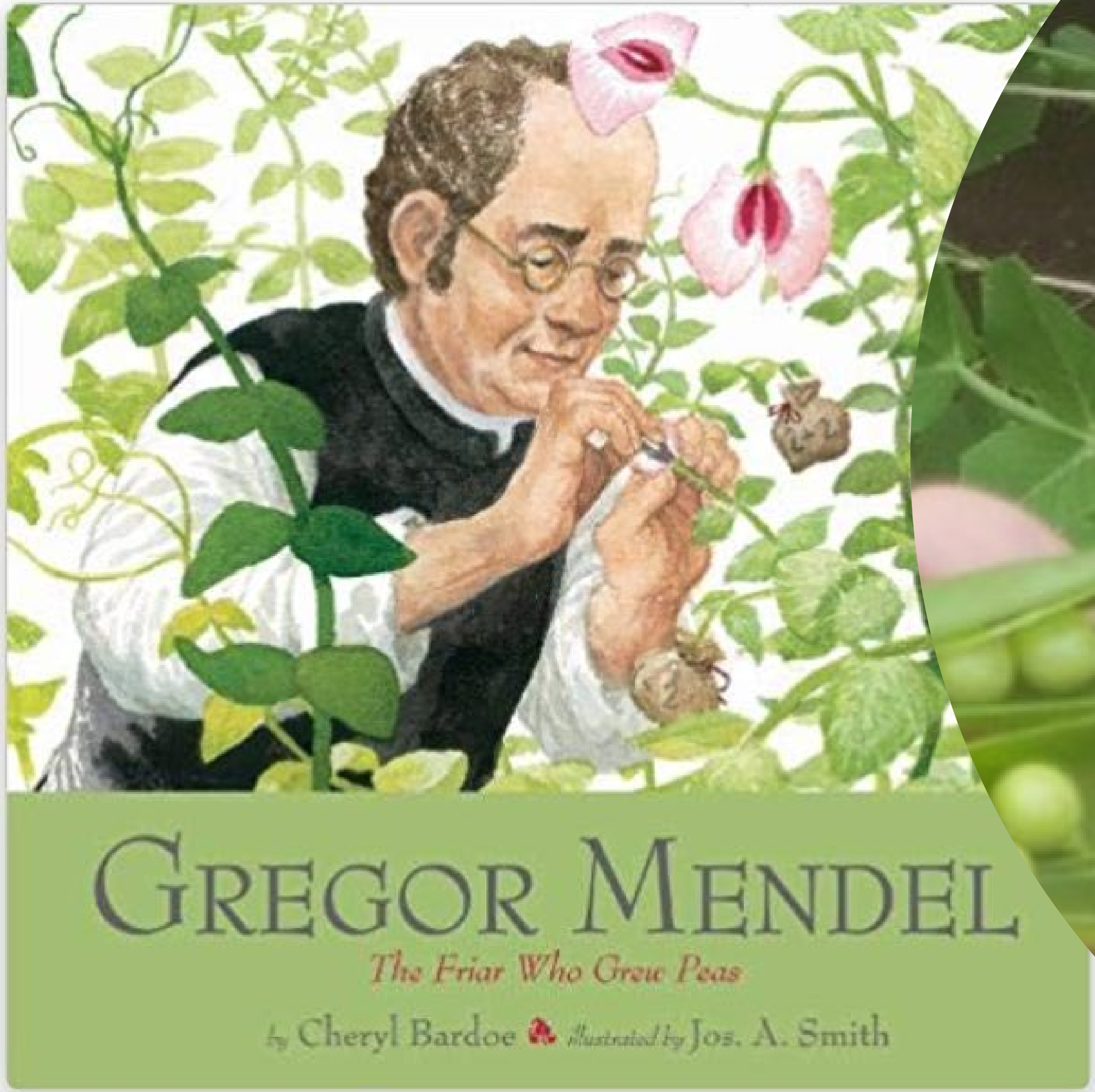


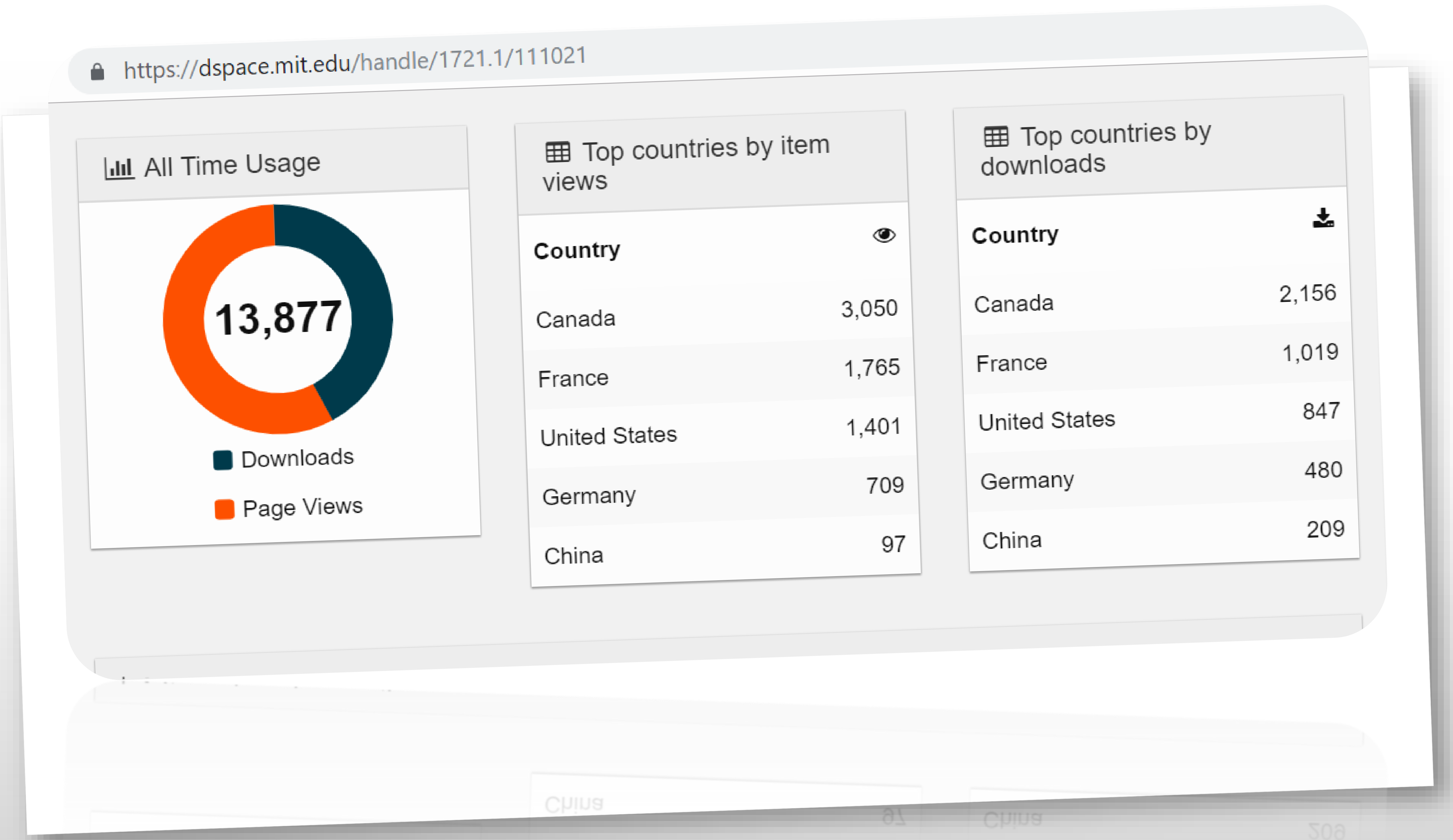
Professor Eric S McLamore
ABE, University of Florida
USDA Center of Excellence

Dr Shoumen Palit Austin Datta
http://bit.ly/SIGNALS-SIGNALS

"If your actions inspire others to dream more, learn more, do more and become more, then you are an enabler." John Quincy Adams

PEAS


GREGOR MENDEL
The Friar Who Grew Peas
by Cheryl Bardoe illustrated by Jos. A. Smith



Thank you for visiting https://dspace.mit.edu/handle/1721.1/111021

CAUSATION AND CORRELATION

Logic and reasoning are quintessential ingredients to make sense of data in any attempt to evolve from data-informed tools to knowledge-informed systems. Causation and correlation are at the core of logic and reasoning processes.

Fundamental Principles

Causation & Correlation: Ground Zero for Data-Informed & Knowledge-Informed Systems

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[3] 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 any tool or development 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 municipalities or NGOs 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-informed[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 key for medicine, farming, optimizing milk production and anywhere decision is desired.

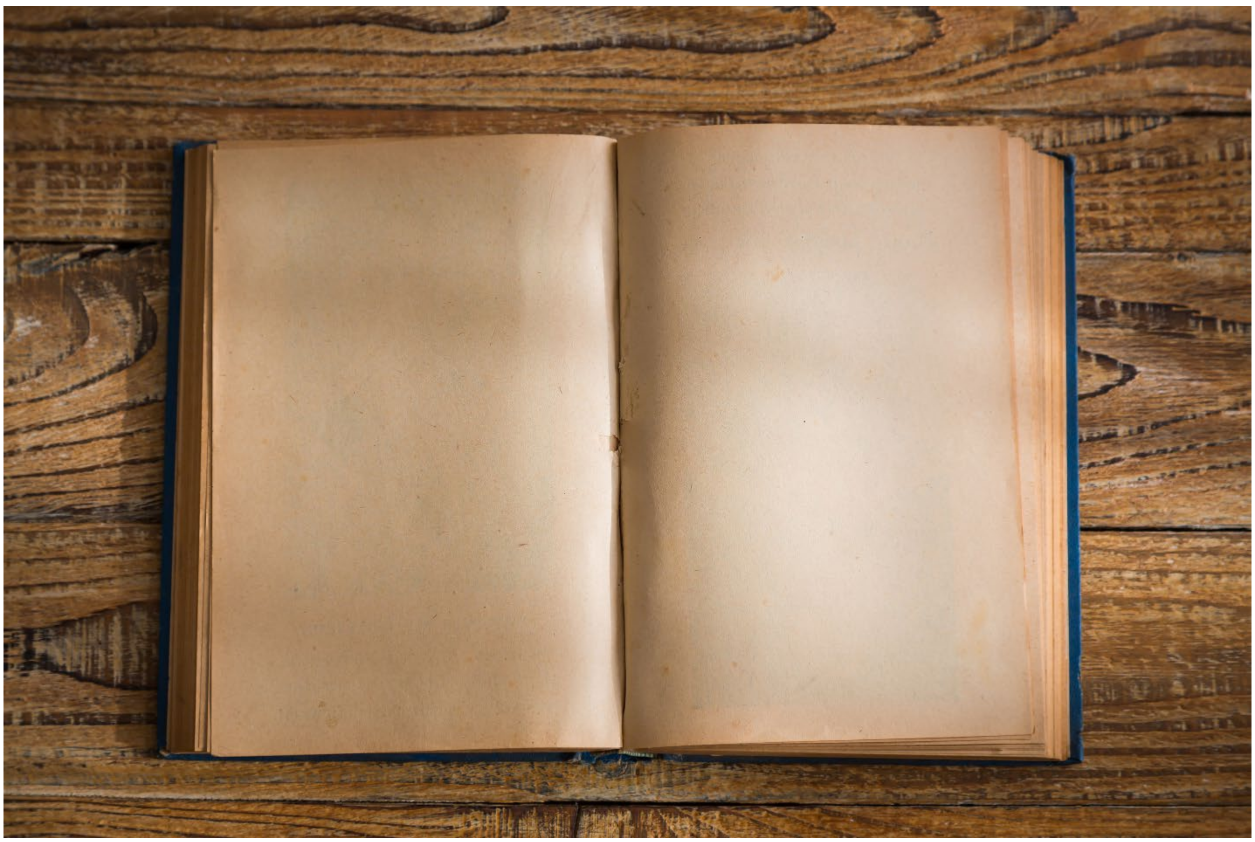
---

[1] https://ncase.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
[4] https://arxiv.org/ftp/arxiv/papers/1108/1108.3235.pdf
[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://cfpub.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://thesystemsthinker.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/scresponse/repository/Rice_SCResp_Article_SCMR.pdf
[18] https://rmas.fad.harvard.edu/pages/change-control
[19] https://dspace.mit.edu/handle/1721.1/41914
[20] https://link.springer.com/content/pdf/10.1007%2Fs13753-017-0154-5.pdf
[21] https://doi.org/10.1007/s13753-018-0190-9
[22] http://science.sciencemag.org/content/359/6373/325
[23] https://www.researchgate.net/publication/267635903
[24] https://issues.org/esty-2/
[25] https://www.brettonwoodsproject.org/2005/08/art-320747/
[26] https://www.ebrd.com/downloads/research/guides/finance.pdf
[27] https://nautilus.org/gps/applied-gps/global-public-goods/what-are-global-public-goods/
[28] http://dx.doi.org/10.1109/JIOT.2017.2755620
[29] http://www.bookmetrix.com/detail/book/7153d069-09a3-4211-a6cf-34588ef367e9#reviews
[30] http://supplychain.mit.edu/supply-chain-games/beer-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
[34] https://preventioncentre.org.au/wp-content/uploads/2018/08/080818_Diabetes_FactSheet.pdf
[35] http://links.jstor.org/sici?sici=0012-9682%28196908%2937%3A3%3C424%3AICRBEM%3E2.0.CO%3B2-L
[36] https://www.amazon.com/Freakonomics-Economist-Explores-Hidden-Everything/dp/0060731338

ESSAY BY ISAAC ASIMOV (1959)

CREATIVITY

How do people get new ideas?



Essay by Isaac Asimov on Creativity (1959)

Isaac Asimov

How do people get new ideas?

Presumably, the process of creativity, whatever it is, is essentially the same in all its branches and varieties, so that the evolution of a new art form, a new gadget, a new scientific principle, all involve common factors. We are most interested in the "creation" of a new scientific principle or a new application of an old one, but we can be general here.

One way of investigating the problem is to consider the great ideas of the past and see just how they were generated. Unfortunately, the method of generation is never clear even to the "generators" themselves.

But what if the same earth-shaking idea occurred to two men, simultaneously and independently? Perhaps, the common factors involved would be illuminating. Consider the theory of evolution by natural selection, independently created by Charles Darwin and Alfred Wallace.

There is a great deal in common there. Both traveled to far places, observing strange species of plants and animals and the manner in which they varied from place to place. Both were keenly interested in finding an explanation for this, and both failed until each happened to read Malthus's "Essay on Population."

Both then saw how the notion of overpopulation and weeding out (which Malthus had applied to human beings) would fit into the doctrine of evolution by natural selection (if applied to species generally).

Obviously, then, what is needed is not only people with a good background in a particular field, but also people capable of making a connection between item 1 and item 2 which might not ordinarily seem connected.

Undoubtedly in the first half of the 19th century, a great many naturalists had studied the manner in which species were differentiated among themselves. A great many people had read Malthus. Perhaps some both studied species and read Malthus. But what you needed was someone who studied species, read Malthus, and had the ability to make a cross-connection.

That is the crucial point that is the rare characteristic that must be found. Once the cross-connection is made, it becomes obvious. Thomas H. Huxley is supposed to have exclaimed after reading *On the Origin of Species*, "How stupid of me not to have thought of this."

But why didn't he think of it? The history of human thought would make it seem that there is difficulty in thinking of an idea even when all the facts are on the table. Making the cross-connection requires a certain daring. It must, for any cross-connection that does not require daring is performed at once by many and develops not as a "new idea," but as a mere "corollary of an old idea."

It is only afterward that a new idea seems reasonable. To begin with, it usually seems unreasonable. It seems the height of unreason to suppose the earth was round instead of flat, or that it moved instead of the sun, or that objects required a force to stop them when in motion, instead of a force to keep them moving, and so on.

A person willing to fly in the face of reason, authority, and common sense must be a person of considerable self-assurance. Since he occurs only rarely, he must seem eccentric (in at least that respect) to the rest of us. A person eccentric in one respect is often eccentric in others.

Consequently, the person who is most likely to get new ideas is a person of good background in the field of interest and one who is unconventional in his habits. (To be a crackpot is not, however, enough in itself.)

Once you have the people you want, the next question is: Do you want to bring them together so that they may discuss the problem mutually, or should you inform each of the problem and allow them to work in isolation?

My feeling is that as far as creativity is concerned, isolation is required. The creative person is, in any case, continually working at it. His mind is shuffling his information at all times, even when he is not conscious of it. (The famous example of Kekule working out the structure of benzene in his sleep is well-known.)

The presence of others can only inhibit this process, since creation is embarrassing. For every new good idea you have, there are a hundred, ten thousand foolish ones, which you naturally do not care to display.

Nevertheless, a meeting of such people may be desirable for reasons other than the act of creation itself. No two people exactly duplicate each other's mental stores of items. One person may know A and not B, another may know B and not A, and either knowing A and B, both may get the idea—though not necessarily at once or even soon.

Furthermore, the information may not only be of individual items A and B, but even of combinations such as A-B, which in themselves are not significant. However, if one person mentions the unusual combination of A-B and another the unusual combination A-C, it may well be that the combination A-B-C, which neither has thought of separately, may yield an answer.

It seems to me then that the purpose of cerebration sessions is not to think up new ideas but to educate the participants in facts and fact-combinations, in theories and vagrant thoughts.

But how to persuade creative people to do so? First and foremost, there must be ease, relaxation, and a general sense of permissiveness. The world in general disapproves of creativity, and to be creative in public is particularly bad. Even to speculate in public is rather worrisome. The individuals must, therefore, have the feeling that the others won't object.

If a single individual present is unsympathetic to the foolishness that would be bound to go on at such a session, the others would freeze. The unsympathetic individual may be a gold mine of information, but the harm he does will more than compensate for that. It seems necessary to me, then, that all people at a session be willing to sound foolish and listen to others sound foolish.

If a single individual present has a much greater reputation than the others, or is more articulate, or has a distinctly more commanding personality, he may well take over the conference and reduce the rest to little more than passive obedience. The individual may himself be extremely useful, but he might as well be put to work solo, for he is neutralizing the rest.

The optimum number of the group would probably not be very high. I should guess that no more than five would be wanted. A larger group might have a larger total supply of information, but there would be the tension of waiting to speak, which can be very frustrating. It would probably be better to have a number of sessions at which the people attending would vary, rather than one session including them all. (This would involve a certain repetition, but even repetition is not in itself undesirable. It is not what people say at these conferences, but what they inspire in each other later on.)

For best purposes, there should be a feeling of informality.

Joviality, the use of first names, joking, relaxed kidding are, I think, of the essence—not in themselves, but because they encourage a willingness to be involved in the folly of creativeness. For this purpose I think a meeting in someone's home or over a dinner table at some restaurant is perhaps more useful than one in a conference room.

Probably more inhibiting than anything else is a feeling of responsibility. The great ideas of the ages have come from people who weren't paid to have great ideas, but were paid to be teachers or patent clerks or petty officials, or were not paid at all. The great ideas came as side issues.

To feel guilty because one has not earned one's salary because one has not had a great idea is the surest way, it seems to me, of making it certain that no great idea will come in the next time either.

Yet your company is conducting this cerebration program on government money. To think of congressmen or the general public hearing about scientists fooling around, boondoggling, telling dirty jokes, perhaps, at government expense, is to break into a cold sweat. In fact, the average scientist has enough public conscience not to want to feel he is doing this even if no one finds out.

I would suggest that members at a cerebration session be given sinecure tasks to do—short reports to write, or summaries of their conclusions, or brief answers to suggested problems—and be paid for that, the payment being the fee that would ordinarily be paid for the cerebration session. The cerebration session would then be officially unpaid-for and that, too, would allow considerable relaxation.

I do not think that cerebration sessions can be left unguided. There must be someone in charge who plays a role equivalent to that of a psychoanalyst. A psychoanalyst, as I understand it, by asking the right questions (and except for that interfering as little as possible), gets the patient himself to discuss his past life in such a way as to elicit new understanding of it in his own eyes.

In the same way, a session-arbiter will have to sit there, stirring up the animals, asking the shrewd question, making the necessary comment, bringing them gently back to the point. Since the arbiter will not know which question is shrewd, which comment necessary, and what the point is, his will not be an easy job.

As for "gadgets" designed to elicit creativity, I think these should arise out of the bull sessions themselves. If thoroughly relaxed, free of responsibility, discussing something of interest, and being by nature unconventional, the participants themselves will create devices to stimulate discussion.

It's okay not to be a "safe" academic.

Sometimes you have to stand up for causes that are bigger than you are, believing that doors that are meant to be opened will be and that nothing can shut them.

Sometimes your character needs to speak louder than your scholarship.