

Managing the Data Deluge: Critical Issues in the Integration and Analysis Of Massive Data in Global Public Health

Dr. George Poste

Chief Scientist, Complex Adaptive Systems Initiative
and Regents Professor in Health Innovation

Arizona State University

george.poste@asu.edu

www.casi.asu.edu

**National Academies Workshop on Big Data and Analysis
for Infectious Disease Research, Operations and Policy**

Pan American Health Organization

Washington, DC

11 May 2016

Data:

The Critical Currency for Improved Global Public Health Capabilities and Preparedness for Epidemic/Pandemic Threats

- **global monitoring of infectious disease dynamics**
- **faster awareness of instabilities and emergent threats**
- **one health: holistic, systems-based analysis of human, animal and ecosystem inter-dependencies**
- **molecular epidemiology, pathogen biology and new technologies for diagnostics/drugs/vaccines**
- **rapid proliferation of diverse data classes**
- **integration, analysis and curation of massive (big) datasets**
- **instructive lessons from other sectors**

The Changing Data Landscape for Biomedical Research, Healthcare Delivery and Public Health

- massive data
- heterogeneous data
- reliable data
- integrated data
- actionable data

Biosurveillance and Situational Awareness in Global Public Health: Mapping The Relentless Changing Dynamics of Infectious Diseases

**old foes:
resurgent drug – resistance**



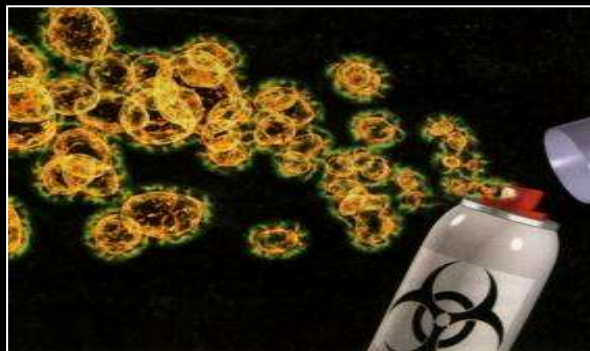
**omnipresent
pandemic threats**



**new foes:
emerging infectious diseases**



**global connectivities
and faster spread**



**bioterrorism and
dual-use technologies**



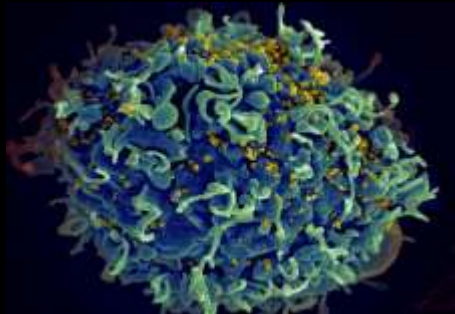
**new technologies: genomics
and synthetic organisms**

One Health: Recognition of the Importance of Zoonotic Diseases as Human Health Threats

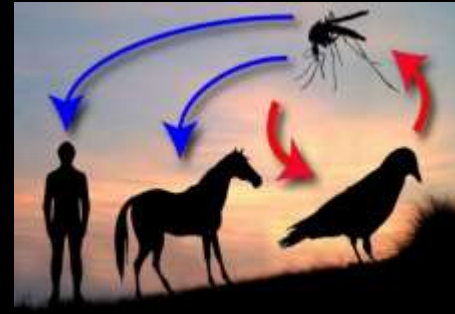
**pandemic (avian)
influenza**



HIV



**West Nile
virus**



MERS



**Ebola
virus**



**bush meat
food chain**



**Zika
virus**



**what's
out there?**

Notice the Resemblance? Hygiene and Quarantine as the Only Controls Absent Drugs or Vaccines

**Bubonic Plague
Physician 15th Century**



**Ebola, Liberia
21st Century**



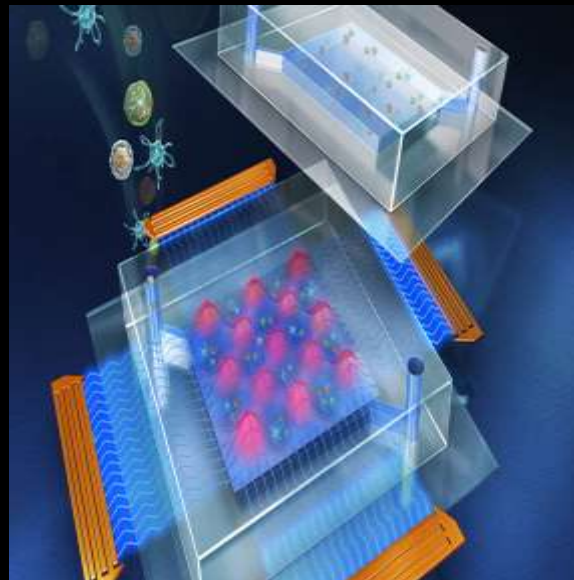
Biosurveillance - Faster Diagnosis Saves Lives: The Primacy of Early Detection and Preparedness Mobilization

Profile



**Comprehensive Genetic
Signature Databanks and
Biorepository for
Infectious Agents**

Detect



**Rapid, Automated
Point-of-Need
Diagnostic Tests for
Far-forward and Low
Resource Settings**

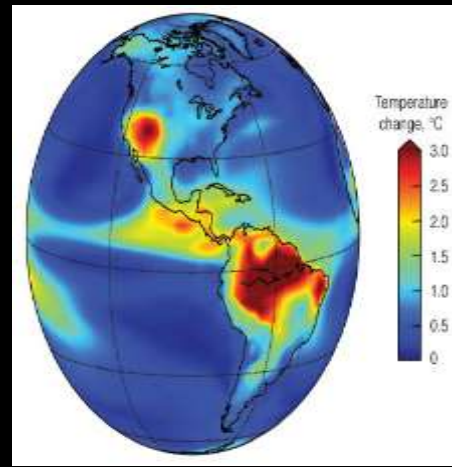
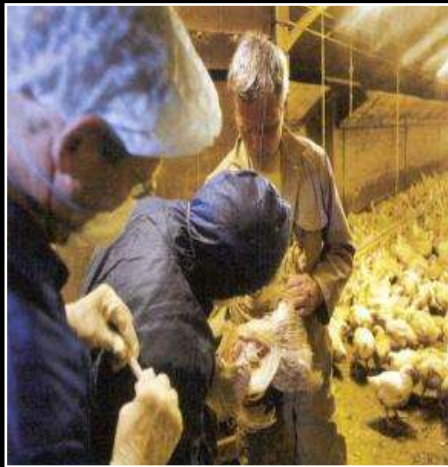
Act



**Real-time Situational
Awareness and
Decision Support**

Geo-Demographic and Geo-Spatial Information Systems: Ground Zero Data and Predictive Modeling

Real-time Intelligence and Faster Preparedness



Real Time Field Reporting of Anomalous Events and Accelerated Detection of Emerging Bioincidents

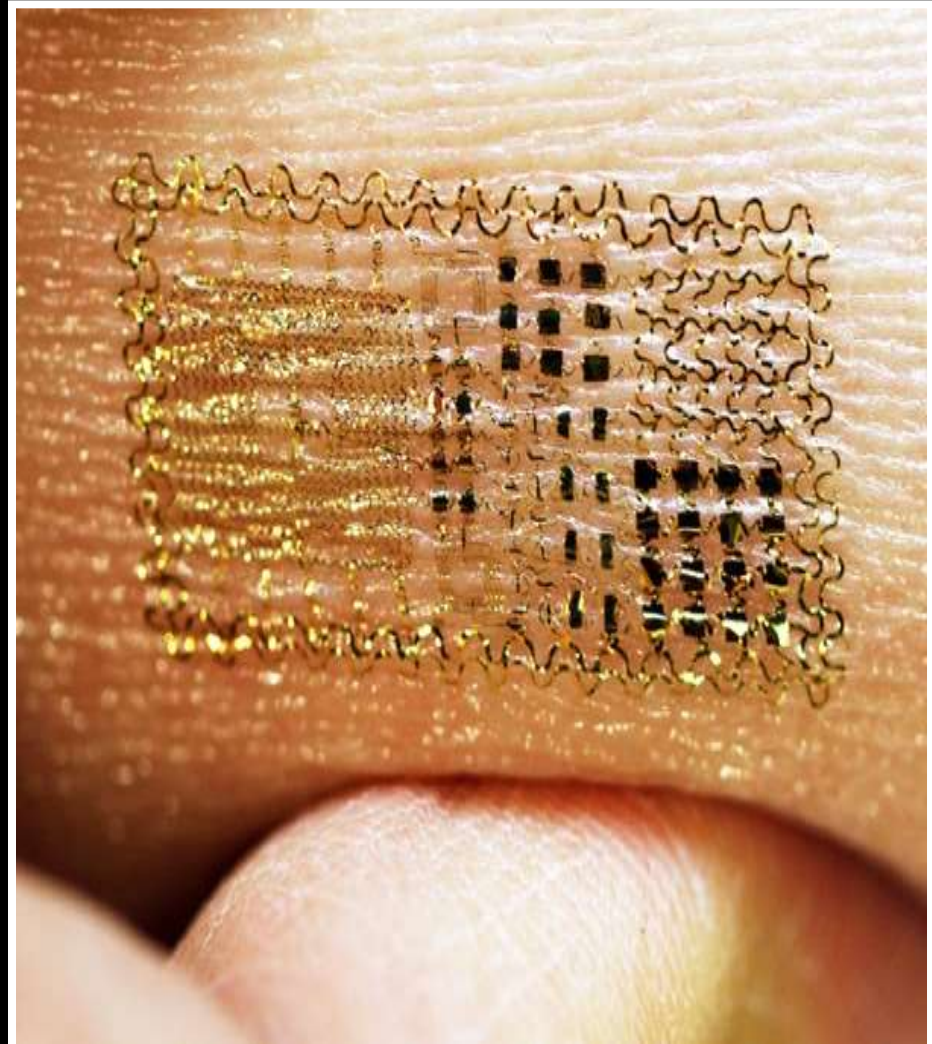


“Lab-In-A-Suitcase”

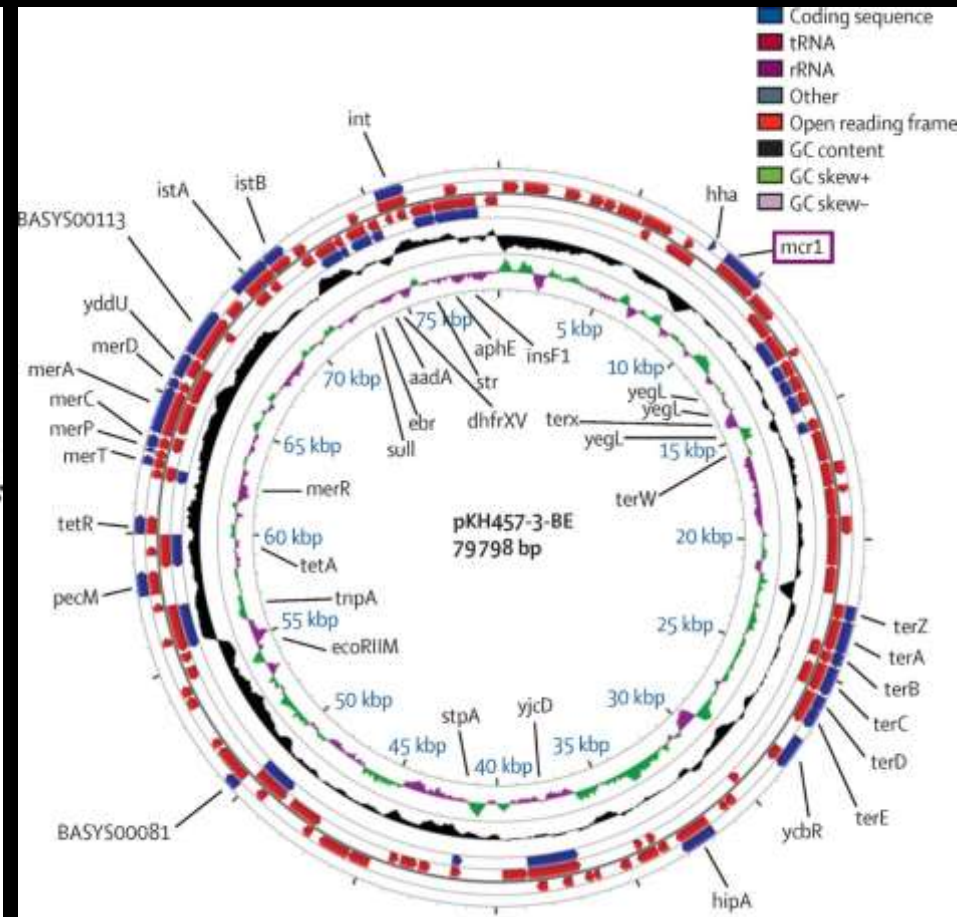
Miniaturized Systems for Point-of-Need Diagnostic Tests



On Body: In-Body Sensor Systems for Real-Time Remote Monitoring and Evaluation of Health Status



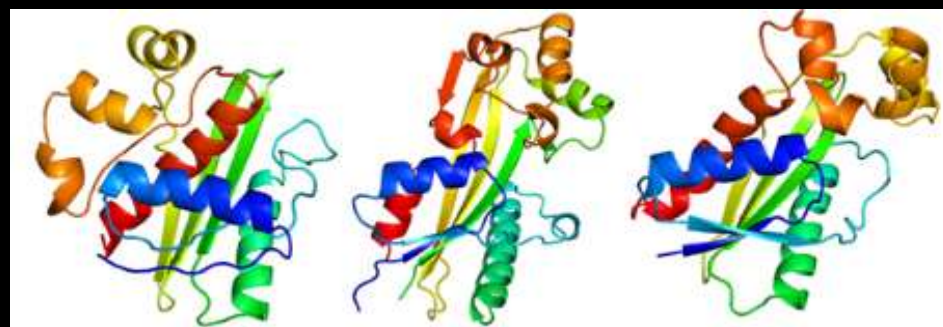
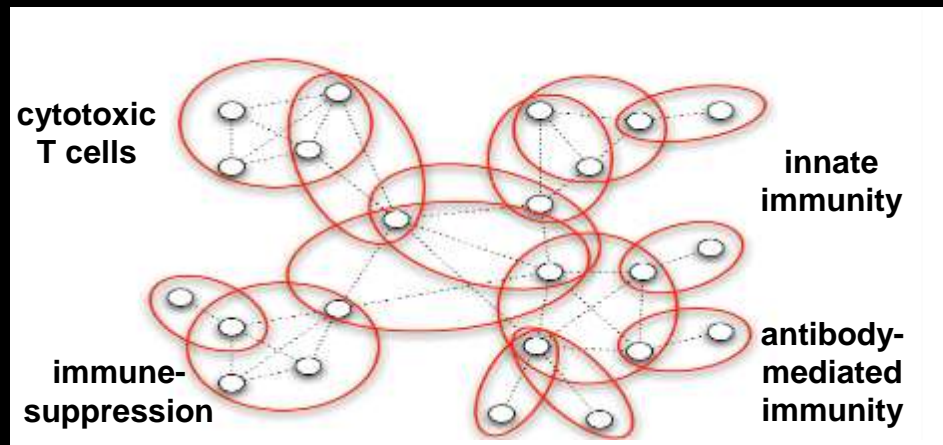
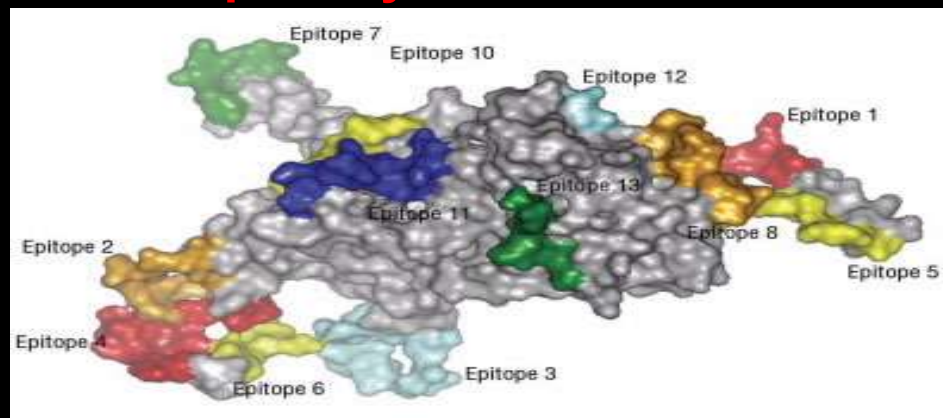
MRSA



Combating “Agent-X”

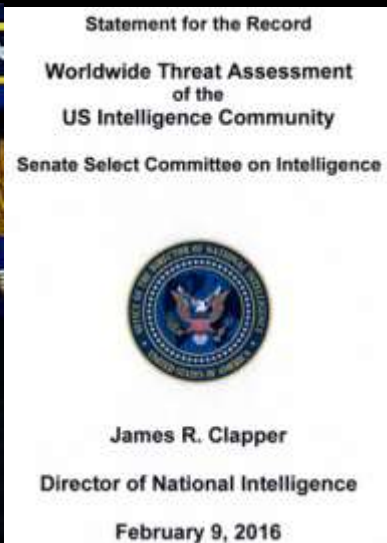
- imperative for faster mobilization to develop vaccines against EID's with major epidemic/pandemic potential
- transition from current protracted ‘Pasteurian’ manufacturing concepts to computational prediction of immunizing epitopes
 - femtosecond laser technology and cyro-electronmicroscopy for protein 3D structure analysis without x-ray crystallography
 - mapping epitope SAR rule sets

Conversion of Vaccine Production from Protracted “Biological” Manufacturing to a Rapid “Synthetic” Chemical Process and New PON Capabilities

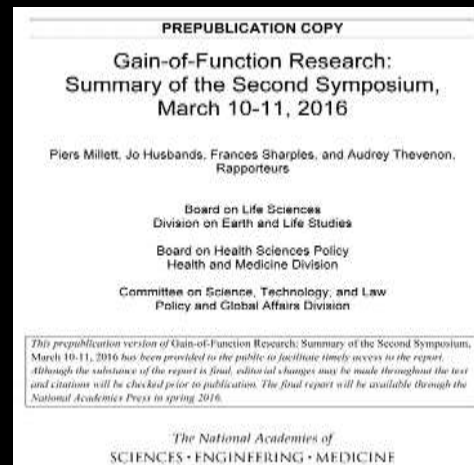


- computational epitope mapping
- analytics to identify commonalities (rule sets) in the composition and structure of proteins that trigger different types of immune responses
- rapid profiling of ‘Agent X’ for ‘epitope rule matching’ and chemical synthesis of epitopes at point-of-need facilities

Synthetic Biology, Genome Editing and National Security: The Ultimate Dual-Use Technology for Modification of Biological Systems?



**Genome Editing Listed
In The Top Six
Existential Threats
To USA**



**Dual-Use Research
of Concern:
Engineered Microbial
Virulence or
Accidental Release**



**Technology Diffusion
and Ubiquity:
Tracking Illicit
Activities**

The Pending Zettabyte Era

1,000,000,000,000,000,000,000,000



Managing Big Data in Biomedicine is Not a Simple Extrapolation from Current IT Practices

Current Public Health, Biomedical Research and Clinical Institutional Structures Are Ill-Prepared For The Data Deluge

Managing the Data Deluge: Annotation, Integration, Analysis and Curation of Massive, Complex Heterogeneous Data

Big Data: the V6 – D3 challenge

V6: volume, variety, velocity, veracity, visualization, value
D3: dynamics, dimensionality, decisions

Global Disease Surveillance



EMERGENCY ID NET



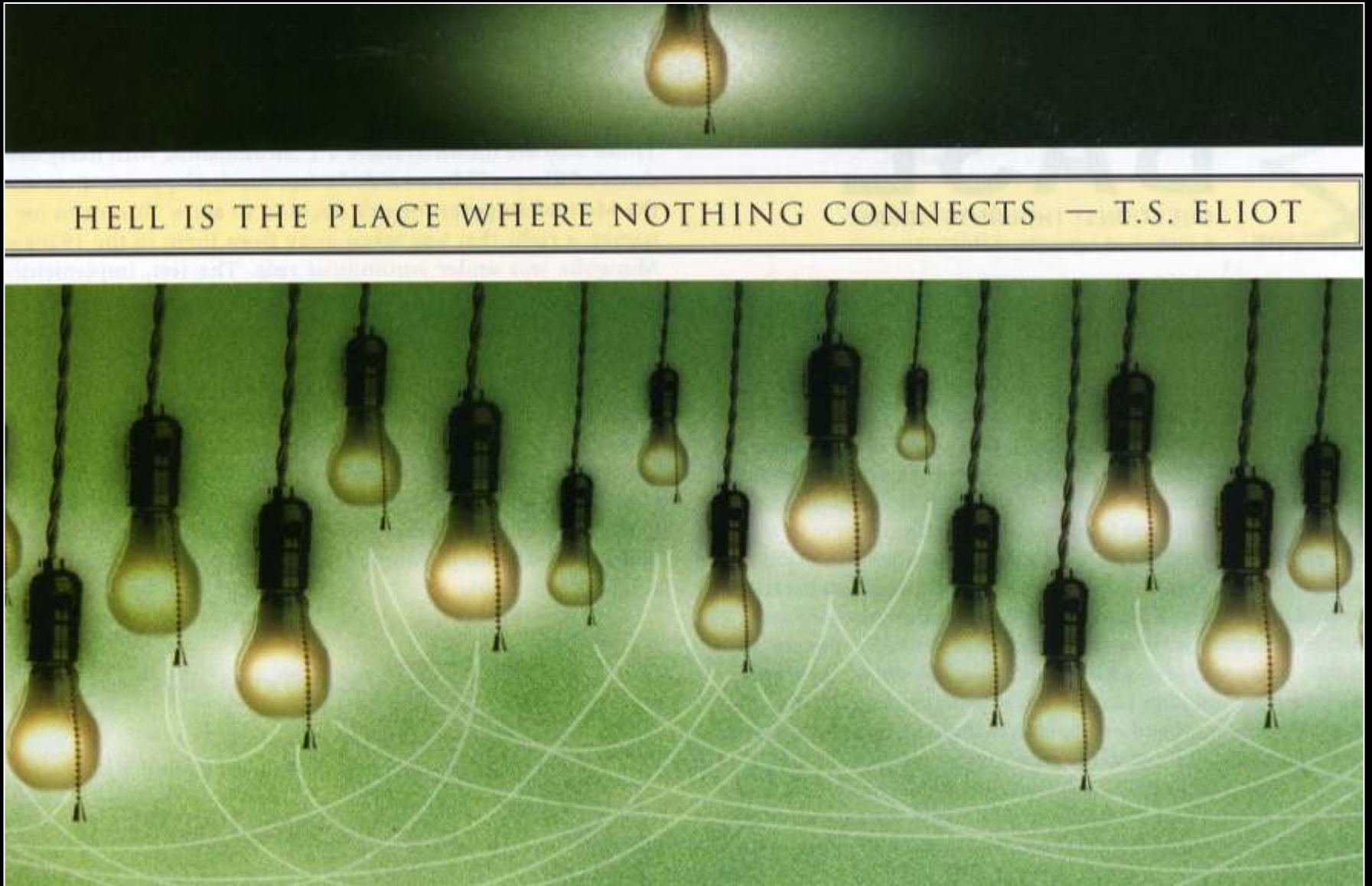
U.S. Influenza Sentinel Provider Surveillance Network



Quarantine Activity Reporting System (QARS).



Seamless Data Communication Networks: A Key Success Factor in Bioincident Management



Seamless Data Communication Networks: A Key Success Factor in Bioincident Management

HELL IS THE PLACE WHERE NOTHING CONNECTS — T.S. ELIOT

**Welcome to a Neglected and Poorly Integrated
Global Public Health Response System**

**Incomplete, Inconsistent and Incompatible Data Report
Formats and Limited Global Database Integration**

The Troubled State of Too Many Data Sets in Biomedical R&D and Healthcare Delivery

- **sloppy science (the reproducibility problem)**
- **statistics (underpowering, overfitting of small N sample sets profiled by large N panOmics feature sets)**
- **silos (data tombs)**
- **sharing (what's that?)**
- **semantics (limited use of common ontologies)**
- **standards (inconsistent and incompatible data formats and dbase inter-operabilities)**
- **static (episodic snap shots of dynamic systems)**

Critical Challenges in the Generation and Analysis of Robust, Large Scale Data in Biomedical R&D, Public Health and Healthcare Delivery

- **scale (exabyte, zettabyte and beyond)**
- **sensor world (IoT)**
- **signals (weak, noisy environments)**
- **structure (80% unstructured; need for NLP methods)**
- **speed (latency, inadequate infrastructure and few fat pipes)**
- **storage (cost and tiered levels of preciousness/fast access)**
- **search (tools, provenance, authenticity)**
- **sustainable ($V^1 \dots V^n$ agility)**

Critical Challenges in the Generation and Analysis of Robust, Large Scale Data in Biomedical R&D, Public Health and Healthcare Delivery

- **security**

- **data-theft/corruption/deception**
- **compromised personal information**
- **healthcare records most frequently attacked data in 2015**
- **national security (dual-use proliferation)**

- **surveillance**

- **sensors to social media**
- **intrusiveness, privacy, consent**
- **data ownership**

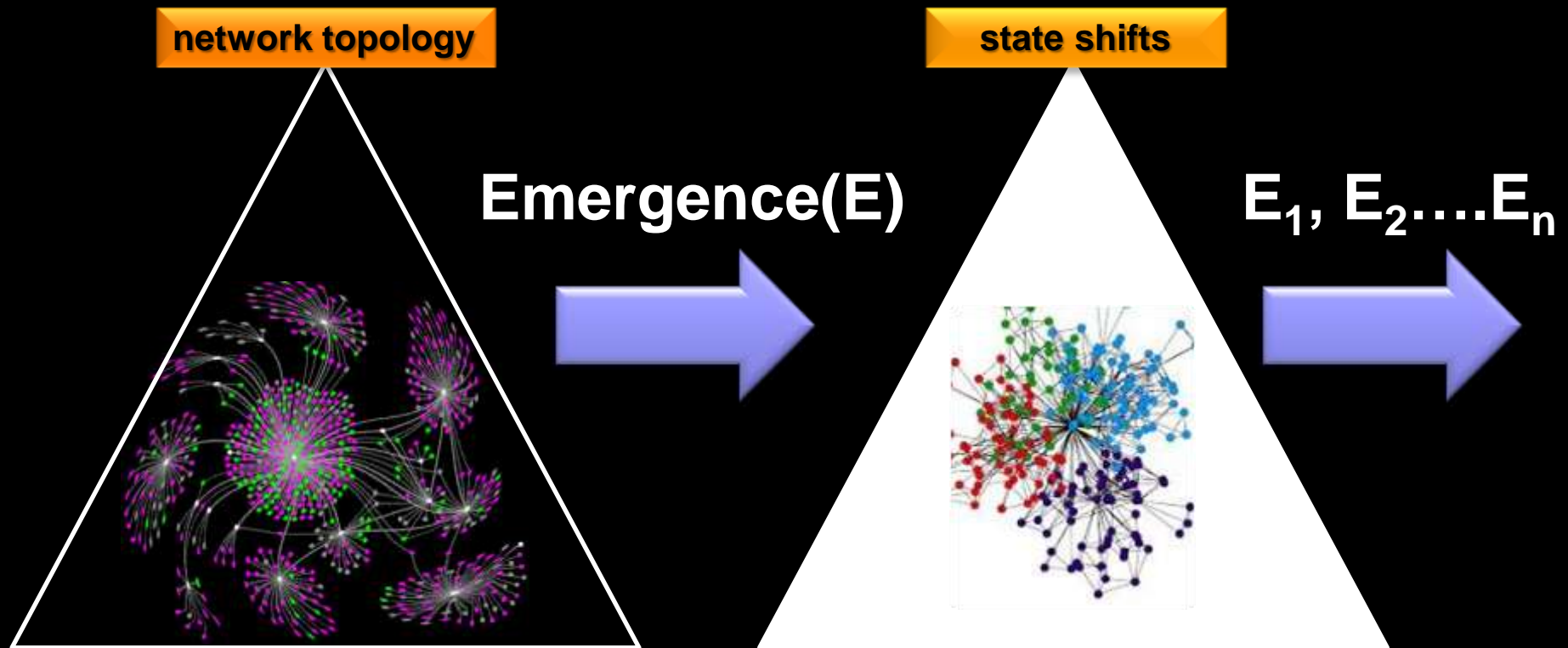
Data Sharing, Transparency and Enforcement

- **International Committee of Medical Journal Editors (ICMJE, 01/16)**
- **NEJM editorial by D. Longo and J. Drazen and inflammatory comments on data scientists as ‘research parasites’ (01/16)**
- **EU, NIH and FDA transparency initiatives for deposition of clinical trial data**
- **increasing importance of access to source code and raw data for reproducibility of reported findings**

Data Sharing and Transparency: A Laudable Concept But Fraught With Many Unresolved Issues

- **who sets the standards for data quality, dbase design and inter-operabilities?**
- **who enforces deposition?**
- **who protects the data?**
 - **patient anonymity, IP**
 - **hacking and data corruption**
- **who ensures legal compliance with different national policies on data transfer about individuals?**
- **who pays?**

Understanding State Shifts in Complex Adaptive Systems and Identification of Triggers of Emergence



- Black Swans
- dislocations
- tipping points
- irreversible cascades
- phase shifts
- perturbations
- inflection points
- unintended consequences
- critical thresholds
- bifurcations
- trigger points

“Digital Darwinism”

- **a pending digital divide**
 - **growing imbalance in sophistication of different end users and their ability to embrace data scale and complexity**
 - **institutions unable to access and analyze large data sets will suffer ‘cognitive starvation’ and relegation to competitive irrelevance**
- **understanding data structure and its productive application/customization for actionable decisions will emerge as a critical institutional competency**
- **major skill gaps in biomedicine and training a new cadre of data scientists**

The Pending Era of Cognitive Computing, Machine Learning and Decision-Support Systems: Overcoming the “Bandwidth” Limits of Human Individuals



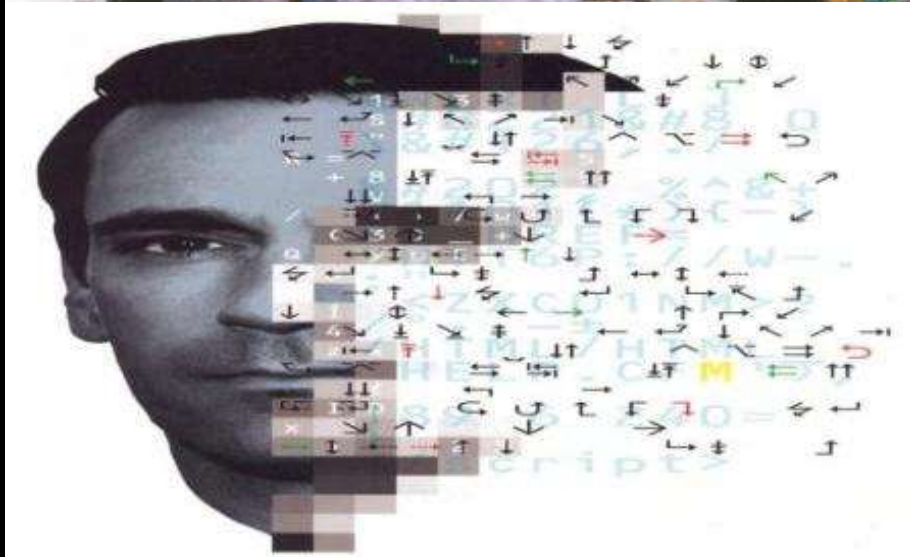
- **limits to individual expertise**
- **limits to our multi-dimensionality**
- **limits to our sensory systems**
- **limits to our experiences and perceptions**
- **limits to our objective decision-making**

Technology Acceleration and Convergence: The Escalating Challenge for Professional Competency, Decision-Support and Future Education and Training Curricula

Data Deluge



New Science and Cognitive Bandwidth

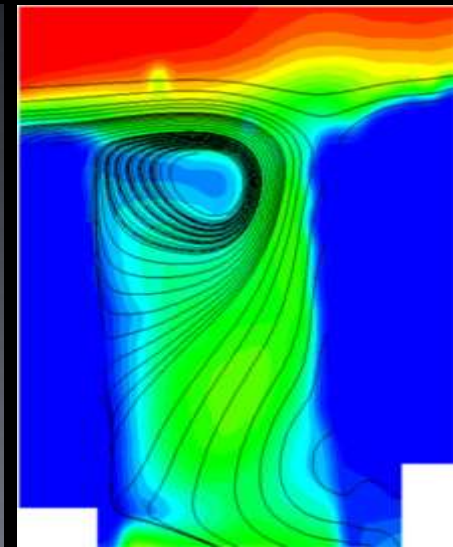
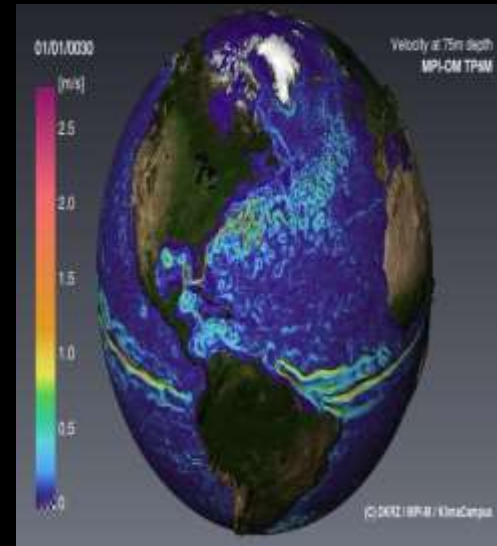


Automated Analytics and Decision Support



Facile Formats for Actionable Decisions

Managing Big Data in Public Health and Biomedicine: Learning Precedents from Other Research Domains and Corporate Capabilities



Advanced Computing and Artificial Intelligence: The Rise of 'Learning Machines' in the Analysis of Massive Datasets and Decision Algorithms



Data-Driven Knowledge, Intelligence and Actionable Decisions

- **changing the nature of discovery**
 - unbiased analytics of large datasets (patterns, rules) versus traditional hypothesis-driven methods
- **changing the cultural process of knowledge acquisition**
 - large scale collaboration networks, consortia and open systems versus individual investigators and siloed data
- **changing the analysis and application of knowledge**
 - real time intelligence, deeper insights and faster improved decisions
- **changing education, training, research, healthcare delivery and public policy formulation**
- **changing the critical competencies and infrastructure required for institutional relevance and competitiveness**

Slides available @ <http://casi.asu.edu/>

