Introductions

 George Hurlburt of STEMCorp

 Former federal IT professional with Navy

 Ran a network for DoD Test and Evaluation for 20 years

 Now publishes frequently for IEEE

 Dr. Eva Lee, PhD of Georgia Tech

 Systems Modeling

 Proficient in applied graph theory

 Tony Schaefer of Linguamatics

 Specializing in converting voice and text to structured data

 Working with National Cancer Institute

 Tim Barr of Cray

 Global Intelligence Division of Cray

 Supercomputer analytics for Government

 Tim White of Cray

 Heads the Global Intelligence Division

 Operates below the software layer

 ESRI Integration

 Daniel Pedraza – Entrepreneur

 Aerospace Engineer

 Computational Fluid Dynamics

 Operating in Net Science Space

 Looking at adapting Natural Language Processing (NLP) to Net Science

 Dr. Chris Fulcher, PhD – Missouri University

 Community Commons website

 Web GIS specialization with ESRI and Buddy Press, 12 FTE Staff

 Custom learning

 Background in agricultural economics

Initial Assumptions

 Data elements

 Clinical evidence

 Laboratory confirmation

 Contact tracing

 Epidemiological data

 Food security

 Animal transmissivity/disease vectors

 Sanitation

 Water

 Energy

 Transport/Logistics

 Human

 Material

 Politics

 Cultural more’s

 Financial/economic

 Initial technological scan

 2G GSM telephony

 Geographic Information System (GIS)

 Many available overlays

 Geo-location accuracy

 With GPS

 Without GPS

 Reverse query issue in mapping tools

 Common core data sets – Topic for group 2

 Advanced analytics

 IBM – DKAN/Va. Tech clinical initiatives

 Community Commons

 Taxonomic baseline

 GIS Linkage via ARC/ESRI

 WADPI

 DTRA Dashboards – available for use

 Linguamatics

 Cray

 Initial options to consider

 1) Baseline via common core data sets, OR

 2) Direct graph analytics

 SMS messaging

 Mining

 Academia

 NGO

 Private Sector

 App feeds

 What’s App

 Trace and go

As-Is Environment (~1 hour session)

 Use Cases

 British Telco-based population migration tracking study in West Africa

 Pre-Ebola study

 Telcos cut data flow, citing privacy concerns

 Not yet restored to researchers

 Supposition

 Can be done without the carrier’s cooperation

 Involves mobile topographical operations

 Aggregation within and between communities

 Requires design where data are not readily available

 Extremely large-scale data collection effort

 Exploit shadow networks

 Rich access in a disaster but accuracy may be an issue

 Interpretation of what exists must be accurate

 Must model appropriately to accommodate errors in interpretation

 Author Sourcing via Cray

 Using what exists in slivers

 Can extract author data at 98% accuracy using

 RSS/News

 Social Networks

 SMS messaging

 Fukishema Radiation leak

 SMS is weirdest where the interpretation is relevant but not well expressed

Leaders emerge with some 20K zombie tweets, but most tweets proved unusable

 West Nile Study at U.C. Irvine using predictive analytics

 Found statistically significant leading indicators based on twitter feeds USG, fearing public panic, requested the study results be suppressed

Oxford animal transmission studies suggesting that the forests of Central Africa could harbor Ebola for some time to come

 Technological analysis of data analytic approaches

 Spatial Data Analytics using Geographic Information Systems (GIS)

 Leading exploratory techniques

 Spatial data analysis is useful

 Statistical data analysis is also useful

 Clusters

 Centrality

 Between-ness

 With 11K significant data layers, what must get pared away for sense making?

 Contextual issues

 Where and what is the role of the outlier?

Where does a given phenomenon fall within the graph in terms of significant relationships?

 The power of the edge (arc) in defining associations

 Social network must be married with geospatial data

 GIS as a means to a visualization end vs. an end in itself

 Ontology

 Best approached bottom-up as an emergent, living phenomena

 No a-priori ontology is required using graph

 Graphs naturally generate contextually relevant many to many

Relationships

 Law of large volume as a moderator of meaningful context

 More than brute force, more of a useful aggregation function

An upper ontology is still potentially useful for tuned precision among common recurring elements

 Geo-location

 Coordinate systems

 Place naming conventions

 Time, including zones

Natural Patterns trump derivation, although both represent models, which only approximate reality

 Geo-location and time vs. behavior

 Geo-location and time vs. group activity

 Layers of data must extend beyond geo-location and time for contextual relevance

 Bottoms up discovery lend significance to geo-location and time

 Cannot normalize all bottom-up data into large scale aggregations

 Facts generally trump ontology from a modeling accuracy standpoint

 Linguistic pattern does, however, vary on meaning, which ontology captures; thus a back-end bottom-up ontology may not be totally immune to some linguistic tuning

Content discovery implies learning from the past (e.g. Ebola Outbreak)

Data standardization

 Cannot standardize data

 To many groups with too much vested interest

 Too much proprietary data

 Moore’s law almost negates time-consuming bottom-up standards building

 Can contextualize unstructured data as filtered for bias

 Goldilocks Principle

 Data is sized just right

 Ebola seems a candidate

 Ebola corpus is large but is likely finite

 Ebola presents as fever as do many other diseases in the West

 African region, and thus the broader field of Public Health applies

Directed graph approach to tacit knowledge

 Limited functionality empirical look at any situation

 Yields a non-partitionable NP Complete data set

 Connections (edges and arcs) extend back to Euhler’s Principle

 Edges are highly significant

 Edge types are also significant when edges in 100’s of billions

 Graph inference requires Subject Matter Experts (SME)

 Connected and multi-disciplinary in nature

 10-100M triples is significant

 100B triples is highly significant

 Embraces GIS data

 Embraces all perspectives

 Geospatial

 Temporal

 Contextual

 Combinational probability supports quality of data

 Planet labs

 Orbital imaging

First session Out-brief

 The value of Network Science – As requested by Dr. McDonald

 Networks are everywhere

 Beehives

 Eco-systems

 Internet

 Social-Network

 Disease Network

 Nodes represent entities within a network

 Arcs represent relationships between networked entities

 As-Is use-cases in point

 Telco in West Africa

 West Nile prediction study on U.S. West Coast

Oxford predictive work on animal transmission of Ebola

 Call to action

Need to broaden use cases and demonstrate:

 Applicability

 Accuracy

Need to get past old paradigms – All systems eventually become legacy

Group went well beyond the As-Is, but that will be reported with the To-Be

To-Be Environment (Time inhibited to ½ hour)

 Going-in topics

 Metrics for:

 Data quality

 Data accuracy

 Applicability of output from a graph discovery engine

 Sources of data

 Methodology

 Role of experimentation (with some level of implied risk)

 Required Research and Development

 Short-term agenda

 Long-term agenda

System of systems approach

 Ebola greater than a single system

 Network of Networks extends the reach a bit far

 System of Systems

 Nested solutions where cause and effect are decoupled

 Not a legacy approach

 Requires design thinking

 Where?

 By whom?

Abstraction requires an analytics pipeline

 Need an “Uber” discovery capacity

 Subject Matter Experts (SME) required

 Multidisciplinary

 Collaborative

 Statistical analysis required for integrated view

Objective: meet stakeholder groups in the network where they are “at”

 Requires community engagement

 Akin to the principles of evidence-based personalized medicine

 Generation of contextually relevant personalized SMS response to all SMS

 Reporters 60-90% of the time

 A difficult task

 An experimental goal

 Context sensitive response, not a normalized reply

 Use of bounded sub-graphs to generate replies

 Supportive of adaptive learning

Noise is not an evil thing. Rather noise is necessary

Garbage collection at all levels

 Feedback depends upon ground input (likely mostly SMS in nature)

 Cultural

 Generational

 Refinement through secondary sources

 Library holdings

 Extensive RSS feeds

 Reporting

 NGO

 Private Sector

 Governmental

 International

 Graph optimization for decisions

Thrust of tailored output to support decisions

 International

 National

 Regional

 Local

 Groups/spheres of influence

 Individual

 By gender

 By age

Second Session Out-brief

 Time limited version – Presented to Dr. Nabarro in under 1 minute

 Seeking strategic outcomes based upon demand signals as discovered

 Personalized messaging using evidence based reporting as culturally influenced

 Did not elaborate on methods, but they were considered

 Fully prepared version – An applied System of Systems problem – Not presented

 Standards are un-realizable

 Pace of technology

 Level of organizations with vested interest

 Information Systems in West Africa are under-exploited

 Examples

 Cell telephony

 Internet of Things

 No backlog of old technology hastens rapid adoption

 Fielding legacy systems is retrogressive

Garbage-collect everything

 SMS

 Photo/video

 Cellular and social media data

 NGO data

 Tribal/linguistic/ethnic data

 Formal structured reporting

 RSS

 Pertinent academic holdings

 Etc.

Ingest all data as triples into a graph discovery engine

 Bulk Resource Description Format (RDF)

 Applied:

 Network Science

 Graph Theory

 Allow ontology to naturally emerge bottom-up and grow

 Limited front-end formal methods with a rudimentary “seed ontology”

 Fact based linguistic tuning potential exists for the backend

 Personalized messaging

 Evidence based:

 Reporting

 Decision-making

 Culturally tempered, not merely normalized

 Seeking strategic outcomes based upon demand signals as discovered

Quick response on late July 09, as requested, tempered by all group replies:

1. Data standards, best built bottom up, require long gestation for consensus among many stakeholders. Given that Moore’s Law remains intact for the foreseeable future, technology will likely move faster than even the most recent standards. Thus, pursuit of standards and standardized data sets will remain elusive given the myriad of potential independent stakeholders, each with their own vested interests. Notwithstanding existing pockets of excellence, such as food data standards, enduring standards are likely unobtainable across all affected disciplines.

2. Information systems in West Africa are underexploited. Cell phones are abundant and the Internet of Things will come about rapidly. Underdeveloped countries have a technological advantage, as they need not overcome a backlog of old technology. Thus, building or endorsing Western-style legacy style systems is retrogressive at best.

3. It is necessary to “Garbage Collect” any and all data concerning West Africa from myriads of available worldwide sources, including:

 A. In country SMS texts

 B. Cellular Video and photographs

  C. Existing data repositories and databases

  D. Tribal/ethnic/linguistic data bases

 E. Academic holdings (often extensive and open)

 F. NGO reporting (frequently public)

 G. Private sector holdings (as available)

 H. Governmental reporting at all levels of indenture

                I. International reporting

 J. Structured NGO and other surveys

4. Convert all collected data to triples (Resource Description Format) and inject them into a powerful graph database for pattern and hub behavior analysis. Use collaborative multi-discipline Subject Matter Experts (SME), such as clinicians, virologists, psychologists, sociologists, anthropologists, agriculturists, engineers and economists, to validate data quality and accuracy using statistical methods to authenticate data.

5. Avoid formal methods on the front end, in lieu of emergent backend ontology.

A. Use a rudimentary ontology on the front-end to establish key contexts such as place and time.

B. Allow living back-end ontology to emerge and flourish based upon existing relationships.

C. Use "two-way" SMS texting inquiry/response in lieu of building exhaustive survey mechanisms, apps or standardized collection tools.

6. Given that a broadly defined community, including individuals, relatives, neighbors, spiritual believers, elders and meme sharing groups, must be fully engaged and empowered, ensure that all feedback is personalized and contextually relevant to the community recipient, health care provider, engaged psycho-social worker or provisioner. Use graph patterns to establish memes and key hubs representing spheres of influence.

7. Reinforce strategic outcome based on demand signals as they are discovered. Build training around identified gaps in knowledge or capability. Use stated demand to drive supply side resource allocation decisions. Positively identify all available clinicians, suppliers, supplies and providers by GPS fed geo-location.

8. Develop, endorse and enforce an Ethical Code of Conduct regarding the use and protection of any personal or otherwise sensitive data consumed, selectively shared and/or selectively dispensed.