Global Challenge Governance: Time for Big Modelling?

Tibor Toth
Executive Secretary Emeritus
Comprehensive Nuclear-Test-Ban Treaty Organisation PC
Vienna, Austria
glocalecubator@gmail.com

Georgios Theodoropoulos*
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
theogeorgios@gmail.com

Samuel Boland
Department of Global Health and Development
London School of Hygiene & Tropical Medicine
London, UK
samuel.boland@lshtm.ac.uk

Ibad Kureshi
Inlecom Systems
Brussels, Belgium
ibad.kureshi@inlecom.com

Adam Ghandar
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
aghandar@sustc.edu.cn

***Abstract*— Global emergencies such as epidemics present immense governance challenges to national, political and operational decision-makers. Modelling and Simulation has been identified as a crucial force multiplier in the development and implementation of preparedness and response measures for epidemics and pandemics outbreaks. Recent years have witnessed an explosion in modelling and simulation tools for this domain while emerging technologies such as IoT and remote sensing enable data collection as an unprecedented scale. However fragmentation and siloing of these efforts hamper their effectiveness. This paper argues that the complexity and scale of the challenge calls for an integrated "Big Modelling" approach which would bring all the different elements together to enable a holistic view and analysis and outlines a computation framework that can act as a catalyst in this direction.**

Keywords— disaster management; epidemics; global governance; agent based models; dynamic data driven application systems; distributed simulation;

I. INTRODUCTION

Governance of global emergencies is one of the Grand Challenges society faces. The stages and required actions in the event of an emergency are outlined in a 2006 report by Atkinson et. al, commissioned by the EU [1]. Keeping the user central, the report details the processes of identifying and understanding the amalgamation of users, processes, services, existing tools and the data required to firstly mitigate and secondly to manage a disaster (conceptualised in the Disaster Risk Management Life Cycle shown in Figure 1). Amongst global emergencies, epidemics and pandemics are the most critical for national, political and operational decision-makers.

The threat epidemics and pandemics entail, be it naturally-occurring, accidental and deliberate, can be illustrated from the not too distant history by the devastation the 1918-20 Influenza Pandemic brought about, leaving behind 500 million people infected and 50-100 million dead. Expressed in today's population numbers that would mean 2.5 billion people infected and 250-500 million dead. While by some estimates it was more devastating than World War I (with a tragic death toll of 60 million people), unlike WWI, the Spanish Flu has been conspicuously missing from collective memory and conscience rituals.

A more contemporary naturally-occurring disease reference point is the 2013-2016 West Africa Ebola epidemic. Authoritative assessment reviews like the Report of the Ebola Interim Assessment Panel [1] and the Harvard-LSHTM Independent Panel on the Global Response to Ebola [3] have identified important preparedness, response and recovery flaws in the performance national and international institutions displayed (Figure 4). Among governance problems identified was the functioning of the UN Mission for Emergency Ebola Response (UNMEER). In essence, UNMEER's governance inadequacy [4] can be summarized conversely as "The whole is less than the sum of the parts" [sic] and that is an important indicator of Complex Adaptive Systems-type of behavior.

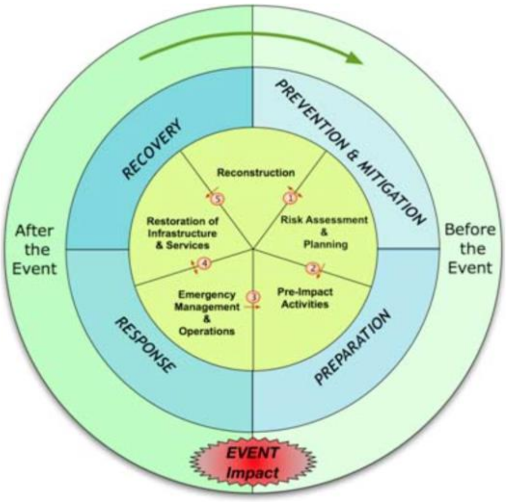

Figure 1: European Commission Lifecycle for Disaster Risk Management Cycle [1].

Ebola did not leave much time for implementing the 2013-2016 West Africa lessons learned [5]. In light of the 2018 outbreak of Ebola in the North Kivu Province of the Democratic Republic of the Congo (DRC) the recurring questions are: "*what progress has been made*" and "*where the challenges remain in preparedness, response and recovery*"?

*Corresponding author

Three progress-highlights are worth mentioning: (a) availability of Ebola vaccine candidates; (b) ICT applications on vaccination status of individuals and spatial vaccination information; and (c) traction in implementing the WHO 2005 International Health Regulations (IHR) through the 2016 launch of the Joint External Evaluation Tool to measure and baseline the country-specific preparedness and response status [6].

Conversely, three challenge-lowlights should be emphasized: (a) lack of focused global political and public health attention; (b) inadequate financial and human resources; and (c) the complex security situation in the DRC provinces generating further potential dwindling of human resources.

In light of the persisting and new challenges it is clear that force multipliers are needed, both for the present DRC Ebola epidemic and for future epidemics in order to prevent the wrong epidemic dynamics recurring again, as it had happened by autumn 2014 in the West Africa Ebola epidemic.

As a promising force multiplier WHO identified simulation exercises as a key component in the validation of core capacities under the IHR monitoring and evaluation framework (2016) and in identifying the strengths and gaps in the development and implementation of preparedness and response measures [7].

In its 2015 report "Managing the Risk and Impact of Future Epidemics" the World Economic Forum invoked technology creators along with other public-private partnership players to develop not just innovative data management software and hardware, but required standardisation mechanisms and identification of "must have" priorities [8]. In a parallel development, the European Union has sponsored a challenge on early warning for epidemics with 5M Euro prize in 2018 [19] (a smaller scale initiative was undertaken by the US CDC in 2013: the Influenza prediction challenge [18]).

Epidemic spread has emerged as a central domain in the field of complex systems that study the connectivity patterns of real-world networks and the emergent properties and behavior of dynamical processes. As a result, it has attracted substantial interest from the modelling and simulation community.

However these efforts tend to be ad hoc and fragmented with no real coordination. Siloed approaches to modelling isolated processes and phenomena at fixed macro-scales are not sufficient to understand the dynamics of complex heterogeneous systems such as epidemics. Instead, what is needed to understand the overall system dynamics, gain insights and develop the required predictive capacity is the consideration and modelling of the full context and the entire set of actors and factors whose interplay at finer spatio-temporal scales gives birth to the overall system dynamics. This integration and coupling should involve a range of models, including bio physical processes, agent-based models of populations, infrastructure models, atmospheric, climate and weather models, earth processes etc. Indicatively, insufficiently holistic modelling in the 2013-2016 West Africa Ebola epidemic resulted in the predicted need for hundreds of treatment beds: Britain alone committed to building 700 beds in bespoke treatment centres built by the British Army's Corps of Royal Engineers [61]. Many Ebola treatment

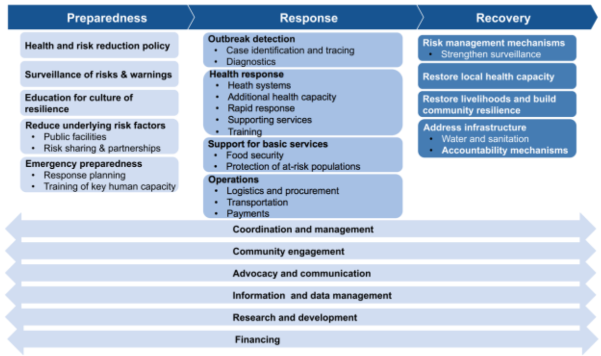

Figure 2: Preparedness, Response and Recovery During the Ebola Outbreak, 2013-2016 [9].

centres received few if any patients, and the financial costs and time burden of facility construction took resources and focus away from more tried and true containment methods, such as the efficient identification and isolation of cases, effective contact tracing, and robust community engagement to underlie these activities [59].

In addition to contextual analytics, the value-added an integrated approach to modelling and simulation could provide stems from the need to be able to reconfigure modelling & simulation inputs and outputs. Each epidemic creates a new mix of challenges:

- new or different subcontinent, or new countries, or new country regions;
- new or different environment: rural vs. urban, high-density vs. low density;
- new or different transportation patterns;
- new or different inflow and outflow patterns, migration, refugees;
- new or different security environment.

Depending on disease virulence and the scale of an epidemic, and in particular the ways in which an epidemic can stress and threaten community and political structures, these challenges may also change quickly and significantly over the course of any one outbreak.

A new challenge configuration thus requires the ability to aggregate, disaggregate, re-aggregate, mix and remix modelling and simulation inputs and outputs prior to and as a process throughout the course of an outbreak. Agent-Based models are more suitable to capture and represent such heterogeneity and disaggregation.

Due to each epidemic posing a new mix of challenges, each epidemic necessitates a new mix of intervention or an adjusted intervention configuration. The 2018- North Kivu DRC Ebola epidemic is not a 'standard' naturally occurring disease outbreak, but one in a highly insecure environment, with Ebola response staff and infrastructure being directly and regularly targeted in the outbreak's epicentre. For instance, a recently announced alliance between the local terrorist group Allied Democratic Forces (ADF) and the Islamic State would only serve to exacerbate risks to Ebola responders and response activities. By comparison it is a *mutatis mutandis* combination of the 2013-16 West Africa Ebola epidemic and the 2013-18 Syria chemical weapons' use situation. It calls for a different mix of IGO, NGO, plurilat-

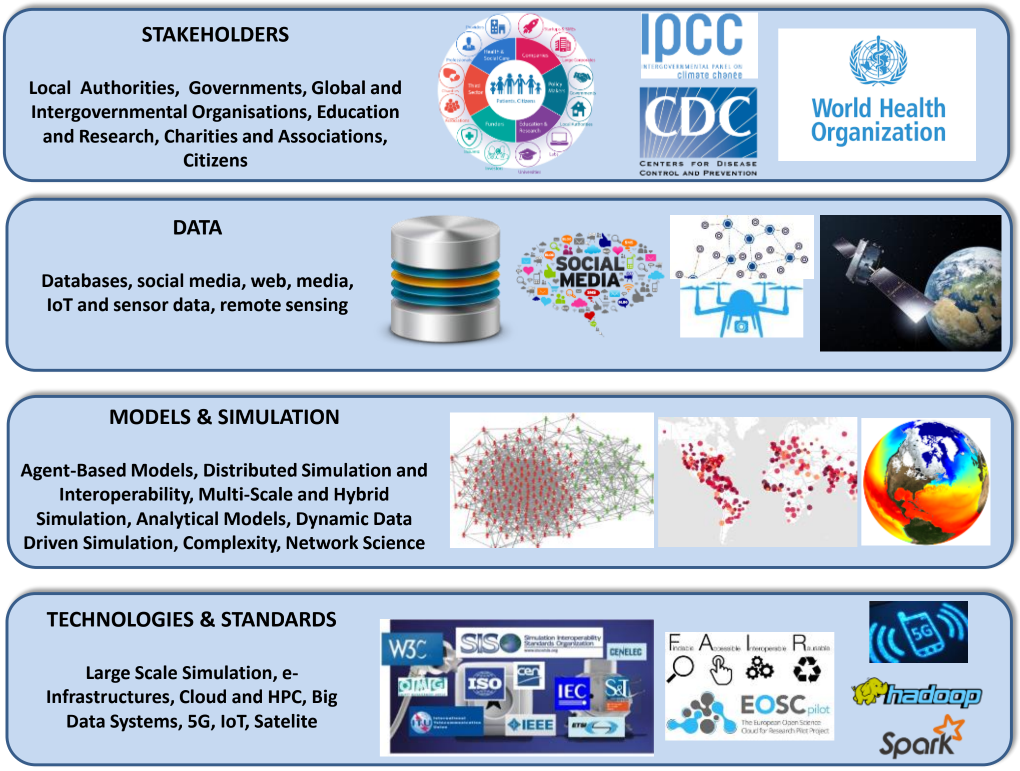

Figure 3: Actors and Elements of a Holistic Approach to "Big Modelling".

eral, bilateral and national intervention. It also invokes co-operation across boundaries of different types of diseases, disciplines, types of intervention and organisations. A whole-of-society problem requires a whole-of-government response. Again, that raises the need and ability to aggregate, disaggregate, re-aggregate, mix and remix modelling and simulation inputs and outputs, including those not normally anticipated or considered in standard epidemiological analyses.

This paper aspires to conceptualise the need for an integrated approach to modelling and simulation that can improve the situational awareness of and the state of interventions in epidemics representing a global challenge. It does this by proposing a computational framework that link together the different elements of this fragmented landscape and support a holistic approach to decision making. The paper coins the term "Big Modelling" to describe such large-scale ecosystems of models, simulations and data, as well as the term "Big Model Platform" referring to the platform as an embodiment of the framework. Big Modelling invokes a new framework since the challenges well exceed the capabilities of conventional analytics approaches and call for an intermingling of scalable data infrastructures and analytics with multi-scale, distributed and agent based simulations engines for the creation of digital twins (Big Model Twins) at a very large scale.

The rest of the paper is organized as follows: Section II discusses the different elements and actors in epidemics modelling and summarizes the gaps along four different dimensions: stakeholders, data, simulation modelling and standardisation. Section III outlines a computation framework that can act as a catalyst for integration and holistic decision support. Section IV epitomizes the conclusions and presents some ideas for future directions.

II. Towards a Holistic Approach

As we are shifting from the "Information Age" to the "Intelligent Age", with the physical and digital worlds rapidly merging, we are dealing with extreme scale complex socio-technical systems where devices, networks, models, data, services, applications and humans are entangled and interconnected. Such systems are associated with the emergent concept of "interaction society" in which the focus of societal issues is increasingly driven by massive and mostly non-hierarchical interactions between individuals and organisations at different levels, whether local, regional, national or fully global. As in other areas of industry and society, intelligent automation and increased interconnectedness can lead to new paradigms for a more proactive approach in forecasting and monitoring, as well as responding in real time to epidemiological disasters.

Epidemics management relies on and involves several elements and actors as illustrated in Figure 3. However the development and activities of these tend to be characterized by fragmentation and isolation, both vertical and horizontal. To deal with the complexity and scale of the challenge in a reliable manner, a holistic approach is required that would bring these different elements together in an integrated effort. The following sections outline the different elements and actors in epidemics modelling.

A. Stakeholders

Models applied for prediction of outbreak target variables for an epidemic include the potential and realistic average number of secondary transmissions resulting from each infected person (referred to as the basic and effective reproduction rates, respectively), the effectiveness of interventions, and context-specific spreading patterns. Each variable shows significant variation across contexts and time, and better accuracy could be obtained in future health crises by increased collaboration and co-ordination of the different stakeholders involved [31].

However, disease outbreak dynamics are driven by a far greater number of variables than those typically included in epidemiological analyses. Through an integrated approach there is a need to overcome the modelling fragmentation among intervention stakeholders and remove the existing silos as defined for the stakeholder organisations and entities by their respective mandate:

- **by disease origin**: naturally occurring diseases, accidental releases and deliberate diseases;
- **by disease hosts**: human, animal, environment;
- **by level of inclusiveness**: global, multilateral, plurilateral, bilateral, national;
- **by wider challenge drivers or megatrends**: climate change, migration, refugees, displaced people, travel, transportation, meteorological forecasting, atmospheric transport modelling, conflicts, peacekeeping operations, humanitarian interventions, etc.

Just to illustrate the model integration rationale for removing existing silos as defined by disease origin: a diverse and complex set of stakeholders arises depending whether an outbreak is natural, deliberate or accidental, whether an affected countries' Ministry of Health, the World Health Organisation, the UN Secretary General's investigation mechanism, a UN Security Council mandated entity, a Biological Weapons Convention organ or an amalgamate entity blending the above institutions will play the leading role under any of those disease origin scenarios. This difficult-to-predict mosaic of actors is further complicated if an outbreak crosses an international border, when it is inevitable that sovereign states will implement different response policies under different political milieus. The difficulty in distinguishing between different types of event during the earliest stages of an outbreak means it is unlikely that medical practitioners will initially consider a deliberate or accidental outbreak [16]. A bioterrorism event can mimic the characteristics of natural outbreaks. Confusion and chaos makes providing effective care difficult [17]. In the case of a deliberate outbreak on the other hand, considerable political and other pressures may be present, and key stakeholders' political focus on containing an outbreak may be secondary to other strategic political interests.

A challenging issue related to stakeholders are affected populations and engagement with them, often referred to as 'community engagement'. As illustrated in Figure 2, community engagement exists at the preparedness, response, and recovery phases. While all elements of the response can be broken down into constituent parts, disentangling community engagement at each phase is crucial if its function is to be considered holistically. At all phases, it includes: a) risk communication and education, b) routinized dialogue mechanisms for community feedback, c) robust participatory functions (not limited to community taskforces, local training and employment, and community contact tracing), and d) external relations and media outreach. The relative focus on each of these constituent functions is weighted depending on which phase the process is in, and transitively, this core

Table 1: Potential Data sources for modelling epidemics.

Category	Datasets
Epidemiological, medical and hospital data	• Susceptible individuals (i.e. plausible contacts of confirmed cases) • Infected individuals • Blood samples and case data during outbreaks • RNA sequencing of virus strains and phylogenic analysis • Available health information systems (HIS) databases (e.g. District Health Information System II databases) • Hospital admissions • Deaths • Phylogenic information (DNA samples) • Control measures effectiveness effect (e.g. transmission rate reduction) • Drug and vaccine clinical trials • Biobank of strains of virus (e.g. Ebola) • Individual cases: patient virus strain, location, district, onset and outcome • Transmission chains
Clinic / Family Doctor	• Online records and medical databases • Surveillance of emerging cases • Interventions • Education
Population	• Demographics • Socioeconomic • UN data and statistics • National bureau of statistics • Poverty levels
Travel mobility	• IATA and passenger databases for air travel • Port and shipping databases
Weather / Climate	• IPCC (Climate Change) • Earth System Modelling Framework
Remote Sensing	• Urbanisation • Deforestation • Land use change
Geographic Information System	• Satellite • Mapping • Air photography • Regional clusters
Digital surveillance	• Social media (e.g. twitter, WeChat, LinkedIn, etc) • Internet search data • Online surveys
Internet of Things (IoT)	• Networked devices • Mobile services (e.g. location services, medical information systems) • City infrastructure, Vehicles

component of outbreak response is inherently dynamic in both type and scale. Furthermore, depending on the context, affected communities can range from relatively homogenous to extremely heterogenous. Particularly in highly dynamic and heterogenous societies, or those without clear and stable trust in existing formal governance, power mapping of affected communities is notably challenging, and community engagement activities are subsequently extremely difficult to appropriately tailor and implement. Nonetheless, accurately anticipating and efficiently identifying these interventions and their relative efficacy has an immense impact on the modelling outcome. Even when response architecture has been established and all key activities are online, poor and ineffective community engagement can and does result in a failure to contain an outbreak. Transitively, even if it took some time to identify an outbreak and mount a robust response, effective community engagement resulting in full consent amongst affected communities for surveillance, isolation, and contact tracing activities could conceivably end a significant outbreak such as the (ongoing at the time of writing) 2018- North Kivu, DRC Ebola outbreak after only two or three waves of transmission [59].

B. Data

Forecasting and monitoring of actual and potential outbreaks relies heavily on the availability of data. The World Health Organisation (WHO), The Global Health Security Agenda (GHSA), the UN, the Institute of Health Metrics Evaluation (IHME), the Global Infectious Diseases and Epidemiology Network (GIDEON), the International Society for Infectious Diseases, the Infectious Diseases Data Observatory (IDDO), and the Centers for Disease Control and Prevention (CDC) are some of the organisations maintaining pertinent databases and that actively monitor and report on potential disease outbreaks [31] .

Surveillance data are crucial in order to rapidly detect, report and respond to outbreaks (see Table 1). Emerging technologies such as satellite remote sensing, IoT, and mobile phone technologies have significantly enhanced surveillance capabilities and have enabled the development of participatory surveillance systems that can provide rich, high resolution crowdsourced data directly from the point of care, whether from local healthcare workers or even from affected individuals or families [30][37] e.g. [62][63][64]. The latter are crucial as they not only augment our understanding of the situation but they can be utilized to dynamically adapt the models enhancing their accuracy and reliability in real time. Improving performance and accuracy of predictive modelling is a force multiplier which allows limited resources to be used more effectively.

However, despite these developments, data collection and management during a crisis remains a profound challenge. Even data management internal to a response that has been operating for many months can be a daily struggle: missing or mis-reported data is common, which is exacerbated by the myriad agencies and individuals who are responsible, or take responsibility for, collecting and reporting data [60]. To aggravate this problem, basic data collection and management architecture and standardized epidemiological analyses often do not exist, and where they do, they are often inadequate [60]. While efforts like the rollout of District Health Information Systems 2 (DHIS2) are slowly improving health facility-based surveillance, in many countries, health facility-based surveillance is still done in a paper-based ad-hoc way, and very few health facilities have efficient, robust, and accessible surveillance databases. As a result, in many countries, the timely and reliable aggregation of surveillance data and subsequent automation of epidemiological analyses is currently not possible with available infrastructure [59]. Often, reliable databases on key components of epidemic analysis and prediction outside of health facilities are even more difficult to source, particularly sociological analyses that underpin the success of many epidemic response interventions.

While in time routinizing the collection and analysis of health facility-based surveillance will become possible due to development of health systems, and will represent foundational importance to the preparedness-phase success of any Big Model Platform, reliable data linkages with responders will remain fundamental to the resolution and breadth of available epidemic-specific data for the response phase. This is particularly true when considering the importance of data sources that exist outside standard epidemiological sources, such as social science research, which are often not collected or understood prior to the declaration of an outbreak and the response to it.

As the data grow in size, resolution and sources so does the need to integrate them to enable contextual analytics. Indeed, any effort towards holistic decision support epidemics should start with data integration and fusion.

C. Models and Simulation

The effectiveness of response to newly emerging events is supported by a rapid targeted deployment of countermeasures [24]. Modelling and Simulation can assist in understanding characteristics of an outbreak and increase the chance of limiting the spread and intensity of its impact. Modelling can be incorporated to guide responses into outbreak events before, during and after the event takes place (Figure 1) providing input into decisions such as intervention types through the use of optimisation and what-if analysis. Key model outputs relate to the intensity and dispersion of an outbreak including the probability an infection will invade a population, spread patterns, expected number of cases, and estimations of the efficacy of interventions [23].

Outbreaks particularly the 2013-2016 West Africa Ebola epidemic which was extremely severe in terms of casualties and scale as well as influenza outbreaks such as SARS and Avian influenza instigated an intensive effort into research of tools and methods to obtain more useful tools that leverage developments in conceptualisation of intelligent and customized modelling using diverse sources of information and big data. Several approaches have been developed based on complexity theory and network science, mathematical approaches and agent-based modelling [23][25][26][27][28][29].

From those the latter is the most promising and suitable to deal with the granularity of data available and the need to aggregate and disaggregate simulation inputs and outputs, as discussed in section I [22]. Agent based modelling techniques provide a capability to develop heterogeneous representations of complex systems such as populations and interactions of different pathogens and other variables such as mobility, diversity of viruses and effectiveness of vaccination. Agent-based models can capture behavioural elements that are difficult to predict. An example of such an element is security. For example, any modeling of the 2018- North Kivu, DRC Ebola response would have to consider the attacks on healthcare workers and health facilities which led to an uptick in cases, namely because two Ebola treatment centres for confirmed case isolation in the outbreak epicentre were temporarily taken offline, and because affected communities were thereafter even more hesitant to be seen to cooperate with the response, or feared for their security were they to be admitted at the later-rehabilitated Ebola treatment centres [59]. Forecasting of insecurity and conflict risks using artificial intelligence and big data is a nascent but increasingly studied field [65][66][67] and some preliminary models have high statistical significance in explaining levels of violence using these mechanisms[68].

Despite these significant developments, the current state of the art of forecasting and predictive modelling infectious diseases unfortunately has failed to deliver the effective outcomes expected. This failure may be largely attributed to two major factors: the first is the grounding of models to data which affects the accuracy and reliability of the simulations; the second is model siloing. We outline these in the next subsections.

1) Dynamic Data Driven Simulation

Given the complexity and contextual diversity of epidemic outbreaks it is intricate to model and predict their emergent properties before the event. During an outbreak the pre-existing models themselves are not void but new rules, initial conditions and parameters need to be constantly provided to dynamically update and calibrate the models to ensure that they accurately reflect the real situation on the ground as it unfolds. It is challenging to update and calibrate models in an outbreak that is ongoing because as well as data limitations and availability, interventions and other parameters are dynamic. Reactive changes that occur during outbreaks represented by dynamic model parameters provide better results than static approaches.

The need for dynamic data driven simulations has been recognised in other domains too and there have been significant work in this direction culminating in the concepts of Info-Symbiotic or Dynamic Data-Driven Application Systems (DDDAS) [10]. DDDAS provides an adaptive feedback loop framework that covers real time collection of data for model adaptation and new initial conditions. As ground realities change-streaming data from sensors and external simulations can be fed back to continuously refine the models and the simulations. Conversely, simulation outputs can guide data collection. These techniques enable advanced data-driven simulation capabilities that can provide more accurate analysis and prediction through dynamic augmentation of models with dynamic data inputs and can enhance understanding of how social systems respond to policy interventions [11][12][13][14][15].

In disease surveillance, data driven simulation has an additional benefit in anticipating spread to new locations before less developed areas report occurrence. The 2013-2016 West Africa Ebola epidemic is an example where the infection was not detected in one country for several months leading to its eventual spread across national borders [33]. Because of the infrequency and uniqueness of events obtaining a statistically significant set of validation data is a significant hurdle [34]. However, while it may not be possible to predict the onset of a new outbreak prior to a health facility or epidemiologists' validation, anticipating its spread thereafter even from highly rural and otherwise poorly understood localities. Furthermore, as real time data systems become more ubiquitous and robust, it will become possible to highlight and make targeted assessments of areas without data visibility, simultaneously increasing the chances of identifying previously unidentified transmission, and encouraging the area's integration with growing data systems. Info-symbiotic simulation can support validation and mitigate the impact of missing data. Data driven models that dynamically adapt as the situation unfolds have also been the focus of recent research in epidemics modelling, e.g. [35][36][38]. However more work is required to fully utilize and exploit the benefits of the info-symbiotic simulation paradigm.

Most crucially, info-symbiotic systems presume robust data collection, availability, and accuracy, which, as discussed in section II.B, for epidemics is not currently feasible at the necessary detail or scale. The use of confidence intervals can constitute a solution to this problem, however the existing data quality would make the intervals too wide to reliably or meaningfully inform decision making. It is also important to note that even if confidence intervals are made clear, the political exigency of modelling and the ease with which non-expert decision makers rely on modelling outputs suggests care must be taken until a degree of accuracy can be tested and assured.

However, despite these challenges, holistic modelling and simulation still has conceptual strengths. Consensus on the utility and appropriateness of core concepts encourages appreciation of the sheer complexity of disease outbreaks. It can thereby help inform improved responses to outbreaks now and in the near future, as well as guide ongoing improvements to global surveillance architecture until robust data is reliably available for Big Model Platforms as it will inevitably become [59].

2) Model Coupling and Distributed Simulation

Epidemics occur in a complex system of systems with dynamic environments, evolving and adapting pathogens, and complex interactions between human and natural systems. As discussed in Section I, siloed approaches are not sufficient to understand the dynamics of complex heterogeneous systems such as epidemics. While genetic algorithms can be used to predict the growth of the pathogens, and transport and logistics models can be used to study the movement of people and animals without integration the results will be simplistic. Systems-of-Systems modelling require integration of different models so that the entire set of actors and factors whose interplay creates the emergent dynamics of the system can be analysed in context (butterfly effect [38]). For epidemics, integration would allow the inclusion of fac-

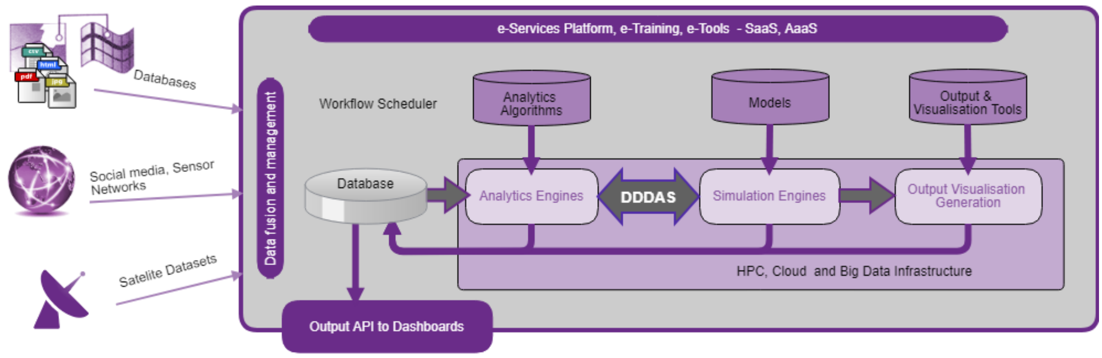

Figure 4: Components of the Proposed Decision-Support Framework (Big Modelling Platform).

tors that relate to the interdependence between society, biological systems, environment, locations, populations, governance, policy etc. Furthermore, integrating different modelling approaches can be managed to promote desirable properties in the resultant ensembles, in addition to quality and standardisation through following best practices, the diversity and correlations between models can enhance the combined forecasts. Ensemble methods that combine outputs from several other models have been utilized for epidemics modelling, e.g. [40][41][42].

However, model coupling is a challenging task. Subcomponent models are of different types (numerical, discrete event) and each is likely to operate at different scales (both spatially and temporally) and may even operate in different dimensions. This requires matching of models at syntactic and semantic levels as well as synchronisation and coordination of simulation components as different spatio-temporal scales. Coupling models also introduces challenging problems with regard to uncertainly related to the interaction of variability from the different model sources and the propagation of uncertainty through the models [20][21].

Distributed Simulation does not only enable integration but would also support high simulation performance that is crucial in large heterogeneous population models [43][44][44] [45].

D. *Technologies and Standards*

Standardisation is the key to data and simulation integration and interoperability. For the former significant progress has been made in the last decade in the context of big data systems and web technologies [47][48][49][50]. Important initiatives have also been undertaken in the health domain (e.g. WHO SDMX-HD), climate and environmental modelling (e.g. OpenMI) and manufacturing (e.g. Industry 4 RAMI).

On the other hand, simulation interoperability remains the holy grail of the Distributed Simulation community and despite significant progress that has been made in this direction (e.g. through the initiatives of SISO, the Simulation Interoperability Standards Organisation) solutions tend to be ad-hoc, domain specific or linked to specific technologies.

III. A Model Platform for Holistic Epidemics Management

To support “Big Modelling” and managing of epidemics a computational framework should be able to interlink people, models and simulations, data, and different types of services including training across borders and scientific disciplines, including social science disciplines. The system should assist stakeholders to monitor and predict epidemic events by providing insight into the risks, vulnerabilities and intervention scenarios. Predictive capacity can and should include not only preparedness-phase risk indicators but also response-phase predictions, whereby geographic risk mapping is updated in real time as surveillance data is made available. The platform should be based on a modular yet integrated design philosophy, allowing the integration of additional models and data sources as they become available (space, sky, sea, ground earth and citizens observation data, new social science research and anthropological information); thus, providing a powerful distributed collaborative environment linking models and data and engaging researchers and stakeholders with different domain knowledge and expertise from different geographic locations.

Such a platform goes beyond a conventional monolithic HPC/Cloud platform. It should encompass features of a service oriented e-infrastructure to support collaboration on a global scale. The ability to scale up or down based on demand alongside support for continuous sustainable operation.

Figure 4 provides a bird’s eye view of the proposed platform with its components. The system provides all necessary support for the coupling of models at different spatio-temporal scales and their efficient execution, data integration and analytics and fosters interoperability through standardisation of interfaces.

A workflow engine drives the system and automates the various stages of data management and processing as specified by the user. A Graphical Workflow Programming environment will enable users a drag & drop interface to create complex layered workflows. These workflows a then able to run as one of instances and output reports, or through the engine can be configured for continuous operation over an area of interest.

A data management layer will perform all necessary data ingestion, cleansing, transformation, fusion, and co-registration. The workflow engine will ensure the movement of the data to and from each component, snapshotting it within the database for provenance and data management. An integrated suite of custom tools built to match the output specifications of data from the different simulation scenarios will facilitate the dynamic visualisation and analysis of data in meaningful, interactive and accessible ways.

The system leverages Cloud and HPC-in-the-Cloud technologies to deliver the required computational power and configurations dynamically to power end users or stakeholder workflows. This goes beyond a naive utilisation of existing computational environments. System-of-systems simulations and holistic analytics form highly heterogeneous memory and computationally intensive big data workloads. They require computational platforms that are highly autonomic, adaptive, elastic, resilient, cost-effective, responsive, and scalable.

Different stakeholders are working to interpret and operationalize response plans and policy with potentially slightly different, though complimentary, objectives. The platform should be able to support the different stakeholders through a continuous integration of services co-developed and co-operated by researchers, technology providers, and commercial vendors. These include modelling (Modelling-as-a-Service), simulation (Simulation-as-a-Service), data (Analytics-as-a-Service) and training.

IV. Conclusion and Future Work

The aim of the Big Modelling framework proposed in this paper is not to replicate past efforts or develop new models and standards, but to leverage existing work and develop a methodology, capability and platform for system-of-systems-wide integration of existing and future models, simulations and data sources to facilitate decision making and governance-support. As new data systems and capabilities come online, they can be integrated into the platform, which by design will adapt and grow according to changing contexts. In that respect at the early stage the framework and the platform is postulated with a restricted focus on

- naturally-occurring, accidental and deliberate disease epidemics being a sub-system of global challenges that are conceptualized as complex adaptive systems;
- governance-support as one aspect of epidemics interventions from the wider toolbox of disease preparedness, response and recovery;
- modelling, in particular agent-based modelling, as one aspects of system-of-systems and complex-adaptive-systems science and applied science.

The novel systems-of-systems approach will support decision-making to a diverse stakeholders and users in relation to potential epidemic outbreak events (real time as events occur and through what-if analysis and prediction of likely outcomes in different scenarios). The Big Model Platform enables the seamless coupling of the constituent dynamic components of a complex adaptive system. The Big Model Twins, as cyber-physical systems modelling both the challenge and the intervention aspects are crucial enablers throughout all the stages of modelling, simulation, validation and testing. They could be extremely relevant in applying "stress testing" as an emerging new tool more and more frequently used in other system-of-systems areas like finance (post 2008 financial melt-down) [54][55][56] and nuclear safety and security (post 2011 Fukushima nuclear accident) [56]. The 360° system-of-systems-wide integration and coupling will lead to innovative governance support approaches in developing risk-management strategies and more broadly in assisting epidemic preparedness, response and recovery as an important subset of global challenge governance.

A source of inspiration for the feasibility of enhanced modelling collaboration should be the Functional Mock-up Interface standardisation project run by the German and European automotive industry [51]. Interestingly the automotive industry, a manufacturing characterized by strong competition, managed to launch and sustain an association, ProSTEP iViP that currently has 180 members from 20 countries. ProSTEP iViP maintains and constantly expands a network of like-minded organisations [52].

There are different options for achieving the integrated approach advocated and envisioned in this paper with a distinct possibility of blending different option elements as well, e.g.:
1. Anchor the framework in WHO's Health Emergencies Programme or its Joint External Evaluation Secretariat as a Big Modelling Interface in collaboration with and relying on the Global Health Security Agenda initiative.
2. Initiate a collaboration project similar to the automotive industry's Functional Mock-up Interface standardisation project.
3. Leverage on the vision of the European Commission for large scale Research Infrastructures [58] and e-infrastructures (European Open Science Cloud, EOSC [51]).

Future work will pursue these options and will seek to develop and operationalise the proposed framework.

V. References

[1] Martin Atkinson, Sara Bouchon, Anna-Mari Heikkila, and Jean-Pierre Nordvik. Report on user requirements supporting risk and vulnerability management of systemic and industrial risks, 2006, EU, Institute for the Protection and Security of the Citizen, Action no. 4335

[2] Final Report of the Ebola Interim Assessment Panel, WHO, July 2015

[3] Harvard-LSHTM Independent Panel on the Global Response to Ebola, The Lancet, November 28, 2015

[4] Fidler, David P, The Ebola outbreak and the future of global health security, The Lancet, May 9, 2015

[5] WHO Secretariat response to the Report of the Ebola Interim Assessment Panel, August 2015

[6] Joint External Evaluation Tool, WHO, 2016

[7] WHO Simulation Exercise Manual

[8] Managing the Risk and Impact of Future Epidemics: Options for Public-Private Cooperation, World Economic Forum, 2015

[9] Managing the Risk and Impact of Future Epidemics: Options for Public-Private Cooperation, WEF, 2015 and Hyogo Framework for Action, "Recovering from the Ebola Response", UNDP Expert Interviews, Project Partner Organisations, Desk Research, BCG.

[10] Frederica Darema, Dynamic data driven applications systems: new capabilities for application simulations and measurements, Proceedings of the 5th international

conference on Computational Science, May 22-25, 2005, Atlanta, GA
[11] Ibad Kureshi, Georgios Theodoropoulos, Eleni Magina, Gregory O'Hare, John Roche, "Towards An Info-Symbiotic Framework For Disaster Risk Management", the 19th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2015), China, October 14 – 16.
[12] Kennedy, C., Theodoropoulos, G., Sorge, V., Ferrari, E., Lee, P., Skelcher C., "AIMSS: A System for Data Driven Simulations in the Social Sciences", Proceedings of the Dynamic Data Driven Application Systems - DDDAS 2007, Beijing, China May 27-30, 2007, Pages: 1098 - 1105, ISBN:978-3-540-72583-1.
[13] Kennedy, C., Theodoropoulos, G., "Intelligent Management of Data Driven Simulations to support Model Building in the Social Sciences", Proceedings of the Dynamic Data Driven Application Systems - DDDAS 2006, The International Conference on Computational Science 2006 (ICCS 2006), UK May 28-31, 2006,
[14] C. Kennedy, G. Theodoropoulos, V. Sorge, E. Ferrari, P. Lee and C. Skelcher, "Data Driven Simulation to Support Model Building in the Social Sciences", Journal of Algorithms and Computational Technology (JACT), Special Issue on DDDAS, Multi-Science Publishing Co Ltd, ISSN 1748-3018, December 2011.
[15] Adam Ghandar, Georgios Theodoropoulos, Bowen Zheng, Shijie Chen, Yue Gong Miner Zhong, "A Dynamic Data Driven Application System to Manage Urban Agricultural Ecosystems in Smart Cities", IEEE International Conference on Universal Village 2018 - UV2018, October 21-24, 2018, MIT, Boston, USA.
[16] Dembek, Z.F., Kortepeter, M.G. and Pavlin, J.A., 2007. Discernment between deliberate and natural infectious disease outbreaks. Epidemiology & Infection, 135(3), pp.353-371.
[17] Cieslak, T.J., 2001. Medical consequences of biological warfare: the ten commandments of management. Military medicine, 166(12), p.11.
[18] Centers for Disease Control and Prevention, 2013. Announcement of requirements and registration for the predict the influenza season challenge. Federal Register, 78(227), pp.70303-70305.
[19] EU Challenge: Early Warning for Epidemics, 2018. https://ec.europa.eu/research/eic/index.cfm?pg=prizes_epidemics
[20] W. Chen, L. Baghdasaryan, T. Buranathiti, and J. Cao, 'Model validation via uncertainty propagation and data transformations', *AIAA J.*, vol. 42, no. 7, pp. 1406–1415, 2004.
[21] A. Patil and Z. Deng, 'Analysis of uncertainty propagation through model parameters and structure', 2010.
[22] Sameera Abar, Gregory O'hare, Georgios Theodoropoulos, Pierre Lemarinier "Agent based Modelling and Simulation Tools: A Review of the State-of-Art Software", Computer Science Review (CSR), Elsevier.
[23] Walters, C.E., Meslé, M.M. and Hall, I.M., 2018. Modelling the global spread of diseases: A review of current practice and capability. Epidemics
[24] Morens, D.M., Fauci, A.S., 2013. Emerging infectious diseases: threats to human health and global stability. PLoS Pathog. 9 (7), e1003467.
[25] Nepomuceno EN, Resende DF, Lacerda MJ, A Survey of the Individual-Based Model applied in Biomedical and Epidemiology Research. J Biomed Res Rev Vol: 1, Issue: 1 (11-24). (2018)
[26] Sara Del Valle, Agent-based Modelling Approaches for Simulating Infectious Diseases, Los Alamos National Laboratory, LA-UR-13-2651 2013
[27] Elizabeth Hunter and Brian Mac Namee and John Kelleher, A Comparison of Agent-Based Models and Equation Based Models for Infectious Disease Epidemiology
[28] Willem L, Verelst F, Bilcke J, Hens N, Beutels P. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). BMC Infect Dis. 2017;17(1):612.
[29] Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review, Phys Life Rev. 2016 Sep;18:66-97.
[30] Chunara R, Smolinski MS, Brownstein JS. Why we need crowdsourced data in infectious disease surveillance. *Curr Infect Dis Rep*. 2013;15(4):316-9.
[31] Chowell, G., Simonsen, L., Viboud, C. and Kuang, Y., 2014. Is West Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia. PLoS currents, 6.
[32] Rachel Reel, How to Find Data on Infectious Disease Outbreaks, Global Health Governance Programme, 2017
[33] Carroll, M.W., Matthews, D.A., Hiscox, J.A., Elmore, M.J., Pollakis, G., Rambaut, A., Hewson, R., García-Dorival, I., Bore, J.A., Koundouno, R. and Abdellati, S., 2015. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature, 524(7563), p.97.
[34] Wong, Z.S.Y., Bui, C.M., Chughtai, A.A. and Macintyre, C.R., 2017. A systematic review of early modelling studies of Ebola virus disease in West Africa. Epidemiology & Infection, 145(6), pp.1069-1094.
[35] Shaman, J., Karspeck, A., Yang, W., Tamerius, J. and Lipsitch, M., 2013. Real-time influenza forecasts during the 2012–2013 season. Nature communications, 4, p.2837.
[36] Ong, J.B.S., Mark, I., Chen, C., Cook, A.R., Lee, H.C., Lee, V.J., Lin, R.T.P., Tambyah, P.A. and Goh, L.G., 2010. Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. PloS one, 5(4), p.e10036.
[37] Thompson, R., Morgan, O. and Jalava, K., 2018. Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. bioRxiv, p.485821.
[38] Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M.A., Flasche, S., Baguelin, M., Pollington, T., Carney, J.R., Glover, R., Smout, E. and Tiffany, A., 2015. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. PLoS currents, 7.

[39] Grenfell BT, Kleczkowski A, Gilligan CA, Bolker BM, Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. Stat Methods Med Res. 1995 Jun;4(2):160-83.
[40] Evan L. Ray, Nicholas G. Reich, Prediction of infectious disease epidemics via weighted density ensembles, PLOS Computational Biology, Edited by Cecile Viboud, vol. 14, issue 2, p. e1005910, Vol 14, 2018
[41] Moore SM, Andrew M, Griffith KS, Titus A, Mead PS, Eisen RJ. Improvement of disease prediction and modelling through the use of meteorological ensembles: human plague in Uganda. PLoS One. 2012;7:e44431.
[42] Thomson MC, et al., Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature. 439:576–9. 2006
[43] Keith R. Bisset, Jiangzhuo Chen, Xizhou Feng, V.S. Anil Kumar, and Madhav V. Marathe.. EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In Proceedings of the 23rd international conference on Supercomputing (ICS '09). ACM, New York, NY, USA, 430-439. 2009
[44] Lees, M., Logan, B., Theodoropoulos, G., “Distributed Simulation of Agent-based Systems in HLA”, ACM Transactions on Modelling and Computer Simulation Journal (TOMACS). Volume 17, Number 3, July 2007.
[45] Vinoth Suryanarayanan, Georgios Theodoropoulos, Michael Lees, "PDES-MAS: Distributed Simulation of Multi-agent Systems", International Conference on Computational Science, ICCS 2013, Barcelona, Spain.
[46] Minson, R., Theodoropoulos, G., “Distributing RePast Agent-Based Simulations with HLA”, Computation and Concurrency Practice and Experience Journal, Wiley, Vol. 20, Issue 10, pp 1225-1256 ,2008
[47] Michel Girard, Big Data Analytics Need Standards to Thrive: What Standards Are and Why They Matter, Centre for International Governance Innovation, Paper No 209, January 2019
[48] Bonner, S., Kureshi, I., Brennan, J., Theodoropoulos, G., “Exploring the Evolution of Big Data Technologies”, in “Software Architecture for Big Data and the Cloud”, edited by Ivan Mistrík, Bahsoon, R., Nour Ali, Maritta Heisel and Bruce Maxim, Elsevier, June 2017, Chapter 14, Pages: 253-285.
[49] Big Data: Big today, normal tomorrow ITU-T Technology Watch Report November 2013
[50] Recommendation ITU-T Y.3600,: Global Information Infrastructure, Internet Protocol Aspects And Next-Generation Networks - Cloud Computing, 2015
[51] Prompting an EOSC in practice, Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud, 2018
[52] Christian Bertsch; Jonathan Neudorfer; Elmar Ahle; Siva Sankar Arumugham; Karthikeyan Ramachandran; Andreas Thuy. "FMI for physical models on automotive embedded targets". Proceedings of the 11th International Modelica Conference, September 21-23, 2015, Versailles, France
[53] ProSTEP iViP homepage, https://www.prostep.org/en/
[54] Ajay Katara, Manoj Reddy, An integrated approach for cross-border stress testing compliance, Global Banking & Finance Review,09/01/2019,
[55] Philipp Härle, Andras Havas, Andreas Kremer, Daniel Rona, Hamid Samandari.The Future of Bank Risk Management, McKinsey Working Papers on Risk, 2015, p.5
[56] Beverly Hirtl. The past and future of supervisory stress testing design. the 2018 Federal Reserve Stress Testing Research Conference, Federal Reserve Bank of Boston, Boston, 9 October 2018.
[57] Communication from the Commission to the Council And The European Parliament on the comprehensive risk and safety assessments ("stress tests") of nuclear power plants in the European Union and related activities, COM/2012/0571 final.
[58] ESFRI Roadmap 2018, Strategy Report on Research Infrastructures, EU, roadmap2018.esfri.eu
[59] David Heymann, Centre on Global Health Security, Chatham House, London, personal communication, April 03, 2019
[60] Owada, K., Eckmanns, T., Kamara, K. B., & Olu, O. O. Epidemiological Data Management during an Outbreak of Ebola Virus Disease: Key Issues and Observations from Sierra Leone. Frontiers in public health, 4, 163, 2016.
[61] Freeman, C. (). British Ebola clinic opens for patients in Sierra Leone. The Telegraph. 2014, November 5.
[62] Robertson, C., Sawford, K., Daniel, S. L. A., Nelson, T. A., & Stephen, C.. Mobile Phone–based Infectious Disease Surveillance System, Sri Lanka. Emerging Infectious Diseases, 16(10), 1524–1531, 2010.
[63] Yang, C., Yang, J., Luo, X., & Gong, P, Use of mobile phones in an emergency reporting system for infectious disease surveillance after the Sichuan earthquake in China. Bulletin of the World Health Organization, 87, 619–623, 2009.
[64] Zaidi, S. M. A., Labrique, A. B., Khowaja, S., Lotia-Farrukh, I., Irani, J., Salahuddin, N., & Khan, A. J.,Geographic Variation in Access to Dog-Bite Care in Pakistan and Risk of Dog-Bite Exposure in Karachi: Prospective Surveillance Using a Low-Cost Mobile Phone System. PLoS Neglected Tropical Diseases, 7(12), e2574, 2013.
[65] Guo, W., Gleditsch, K., & Wilson, A. Retool, AI to forecast and limit wars. Nature, 562(7727), 331–333, 2018.
[66] Rost, N. Will it happen again? On the possibility of forecasting the risk of genocide. Journal of Genocide Research, 15(1), 41–67, 2013.
[67] Schneider, G., Gleditsch, N. P., & Carey, S. Forecasting in International Relations: One Quest, Three Approaches. Conflict Management and Peace Science, 28(1), 5–14, 2011.
[68] Pavlik, M. Network Position as a Contours of Conflict: Network Position as a Predictor of Political Violence Level and Type in Afghanistan. Presented at the Households in Conflict Network (HiCN) Workshop 2019, Paris, October 17, 2019.