



Information and Communication Technologies

# EPIWORK

## Developing the Framework for an Epidemic Forecast Infrastructure

<http://www.epiwork.eu>

Project no. 231807

---

### **D7.5 Yearly activity report with assessment of lesson learned**

---

Period covered:

Start date of project: February 1<sup>st</sup>, 2009

Due date of deliverable: February 20<sup>th</sup>, 2010

Distribution: public

Date of preparation:

Duration:

Actual submission date: February  
20<sup>th</sup>, 2010

Status: Draft

Project Coordinator: Alessandro Vespignani

Project Coordinator Organization Name: ISI Foundation

Lead contractor for this deliverable: ISI Foundation

## Work package participants

The following partners have taken active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

- ISI
- FGC-IGC
- TAU
- MPG
- AIBV
- SMI
- FFCUL

## Change log

<b>Version</b>	<b>Date</b>	<b>Amended by</b>	<b>Changes</b>
<b>1</b>	<b>20/02/10</b>		

# Table of contents

<b>1 Project objectives and implementation strategy</b> .....	<b>4</b>
<b>2 Research and Development activities during the first year</b> .....	<b>5</b>
Modelling and theoretical foundations.....	<b>6</b>
Data-driven computational platform .....	<b>11</b>
ICT monitoring and reporting system .....	<b>17</b>
<b>3 Discussion and outlook</b> .....	<b>20</b>
<b>4 List of publications and scientific talks</b> .....	<b>20</b>
Publications.....	20
International Conferences and seminars .....	22

# List of figures

FIGURE 1: CONTACT NETWORKS BETWEEN SELECTED INDIVIDUALS WITH SWINE FLU IN ISRAEL	8
FIGURE 2: EFFECTIVE GEOGRAPHIC BORDERS IMPLICITLY ENCODED IN MULTI-SCALE MOBILITY NETWORKS (RED). THE DISCOVERY AND COMPUTATION OF THESE STRUCTURES HEAVILY RELIED ON THE USE OF THE MOBILITY DATABASE SERVER DESCRIBED IN THE DELIVERABLE 2.1	9
FIGURE 3: THE STRUCTURE OF THE MOBILITY DATABASE SERVER AND ITS OPERATION WITHIN THE PROJECT	12
FIGURE 4: INCIDENCE CURVES FOR US AND SPAIN FOR DIFFERENT INTERVENTION SCENARIOS. THE GRAY BAR INDICATES THE TIME PERIOD DURING WHICH THE IMMUNIZATION TAKES EFFECT.	15

## 1 Project objectives and implementation strategy

A fully operational, accurate and reliable epidemic forecast infrastructure nowadays faces problems related to the lack of appropriate models to understand how an infectious disease spreads in the real world, lack of extensive and accurate epidemiologically relevant data (from societal data to epidemic surveillance data), lack of understanding of the interplay among the various scales of the problem (from the host-pathogen interaction, to human-to-human transmission, to the interaction with the environment) and, most importantly, lack of communication among the different areas of research which proceed almost independently, crucially hampering a significant progress in a highly interdisciplinary field of research. The present project intends to fill this gap. *Through computational thinking, complex systems concepts and data integration tools relevant for epidemiological understanding at all levels, it will provide a set of radical, paradigm-changing results enabling a novel approach to the modelling, forecast and policy making approach to infectious diseases.* The projects overarching goals are:

- The identification of general principles and laws that characterize complexity and capture the essence of *complex epidemiological systems*.
- The development of a *collaborative information platform* enabling the production of knowledge, understanding and models from the novel *abundance of digital data in epidemic research*.
- The development of an *open, data driven, computational modelling platform* to be used in epidemic research as well as in policy making for the analysis of global epidemics, integrating and leveraging on transnational data.
- The development, deployment and validation of an *Internet-based Monitoring System (IMS) producing real time data on disease incidence and epidemic spreading*.

The project aims at exploring the following *work areas* as the major research themes directly matching the objectives of this proposal:

- ***Modelling and theoretical foundations***
- ***Data-driven computational platform***
- ***ICT monitoring and reporting system.***

The work plan is organized around six distinct scientific work packages (WP1-WP6) whose parallel scheduling of the work packages is necessary to jump-start the cycle and the Inter-WP validation. WP1 and WP2 are aimed at exploring theoretical issues in the area of epidemic modeling in complex, multi-scale systems, structured populations and in the presence of the dynamical interplay between social and technological factors, seasonality and climate, health policies implementations, WP3 and WP4 are devoted to the collection and sharing of data on a computational platform and have a two-way continuous exchange of data and algorithms with WP1 and WP2. WP5 and WP6 is aimed at the development, set-up and deployment of innovative web monitoring and data gathering tools that should provide a continuous stream of data to WP3-WP4 and be informed by constant feedback on the modeling needs in terms of data gathering by WP1 and WP2. The common research agenda of the consortium teams, which work in a coordinated way on the various tasks, favors a closer interchange of ideas and knowledge among the groups and the various components of the project in a truly interdisciplinary collective effort.

## **2 Research and Development activities during the first year**

In its first year of life the project has progressed according to the planned timeline and in many cases ahead of schedule. The level of coordination of the research activity across all work packages has been extremely high with frequent exchanges of information and results. The development of computational and data management tools has been accelerated by the unfolding of the H1N1 pandemic and the involvement of most of the teams of the project in the International effort of monitoring and anticipating its course. All teams of the consortium have been actively involved in the data gathering, computational analysis and monitoring in close contact with national and International agencies, including the JRC crisis unit, the ECDC and Institute or Ministry of Health of most of the countries represented in the project. On a side, the H1N1 pandemic has represented an exceptional strain and workload for the teams while, on the other side, it has forced the implementation and test of theoretical models, computational tools and the IMS infrastructure ahead of time, thus inducing great momentum to the entire project. All Epiwork teams involved in the Epiwork Internet-based Monitoring Systems for Influenza surveillance have carried on an enhanced surveillance during the whole pandemic. The IMS has been readily implemented in UK in July to cope with the emergency

of the rising number of H1N1 cases. The results of the IMS activity in the different countries are collected on the Influenzanet page and being analyzed. The comparison of the data from the IMS with the traditional surveillance is very encouraging and a full retrospective data analysis will result to be extremely valuable to improve the IMS and as a test of its reliability. Some of the results and publications done for the realistic modelling of the H1N1 pandemic has to be considered as a major breakthrough that has shown for the first time, in a real world situation, the potential of computational methods in providing anticipations and forecasts that can be used in the support of the policy making and public health decision-making processes. The results obtained show a very good agreement between predictions and real data, providing a strong and remarkable test of the quantitative level of the prediction offered by computational methods.

In the following we report an overview of the major results achieved by the project in the main work areas of the project.

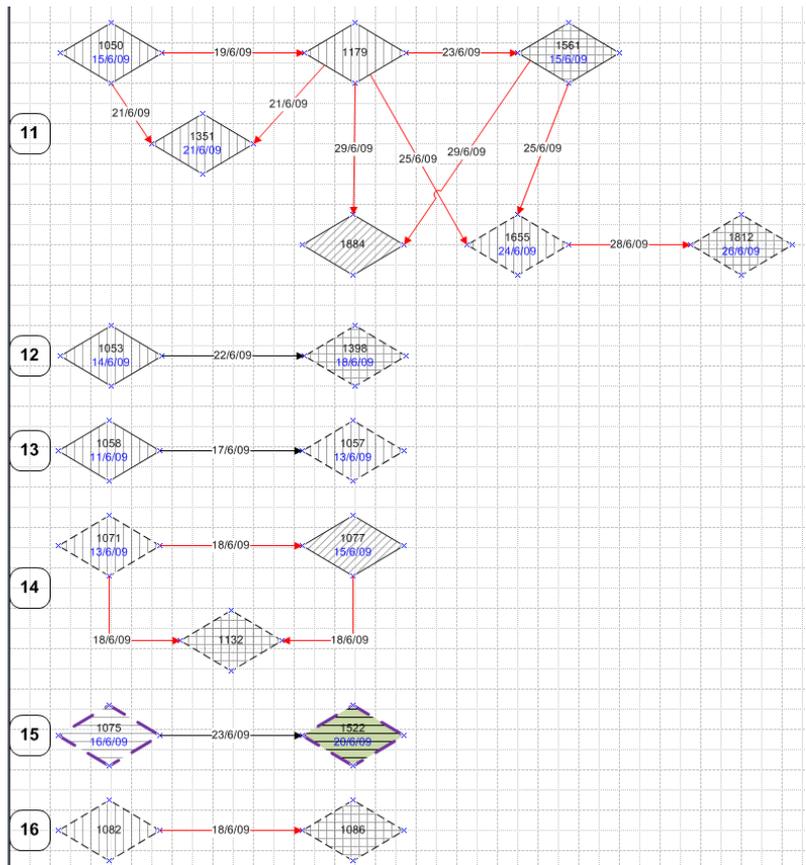
## **Modelling and theoretical foundations**

### **Analysis of Epidemic Dynamics on Clustered Networks**

The theoretical group of the **FFCUL** team has been working on a series of papers investigating spatially restricted contact networks, namely those with power law distributed connectivity to study super-diffusive spreading, as suggested from empirical data of human contact proxies like travel networks and money bill notifications. The influence of such contact networks on epidemic threshold behavior is investigated, especially important for models with reinfection which have, besides the no-growth-endemic threshold, also further thresholds like the one between SIR and SIS type behavior, namely the reinfection threshold.

The Bar-Ilan University (**BIU**) group focused on clustering effect on immunization in epidemic network. First, a clustered network has been constructed in this study. The clustering coefficient gives the probability that two neighbors of an individual are neighbors of each other. By varying the clustering coefficient, it is observed the effect of changing the clustering coefficient on various quantities of interest. Especially, different immunization strategies are studied with different clustered network. As the clustering coefficient of the network increases, the acquaintance immunization is far more effective than random immunization. A manuscript is in preparation.

The TAU team has been working on the emergence and size of the giant component in clustered random graphs with a given degree distribution. The paper published in Physical Review Letters provides a new generating function approach to clustered networks. The group was also active on the study of the early stage swine flu contact networks in Israel based on surveillance of the 2009 pandemic in Israel. Most individuals tested for swine flu were asked by their doctor to identify who they believed they were infected by, thereby providing important information concerning the contact structure of the Israeli population. Of the 713 cases, 183 (or 25.6%) were able to provide information establishing contact links. The data was used to assemble contact networks which map the connections between an infected person (infector) and the different individuals he/she infected (infectees). The networks are visualized by portraying individuals as nodes in a graph and placing a directed edge from any infector to the individuals he/she infects. The networks thus reveal who infected who. Altogether there were 66 separate infection networks totaling 183 nodes, with 123 links between them. The significant mismatch between number of nodes and links is due to the fact that the networks were largely disconnected with many isolated links. The mean number of outgoing links per network was 0.67. Most nodes, i.e. 55.2%, had no outgoing links, 33.3% had one outgoing link, 7.1% had two outgoing links and 2.7% had three outgoing links. In addition, there were three patients having the status of what might be considered "superinfectors" with four, seven and ten outgoing links respectively.



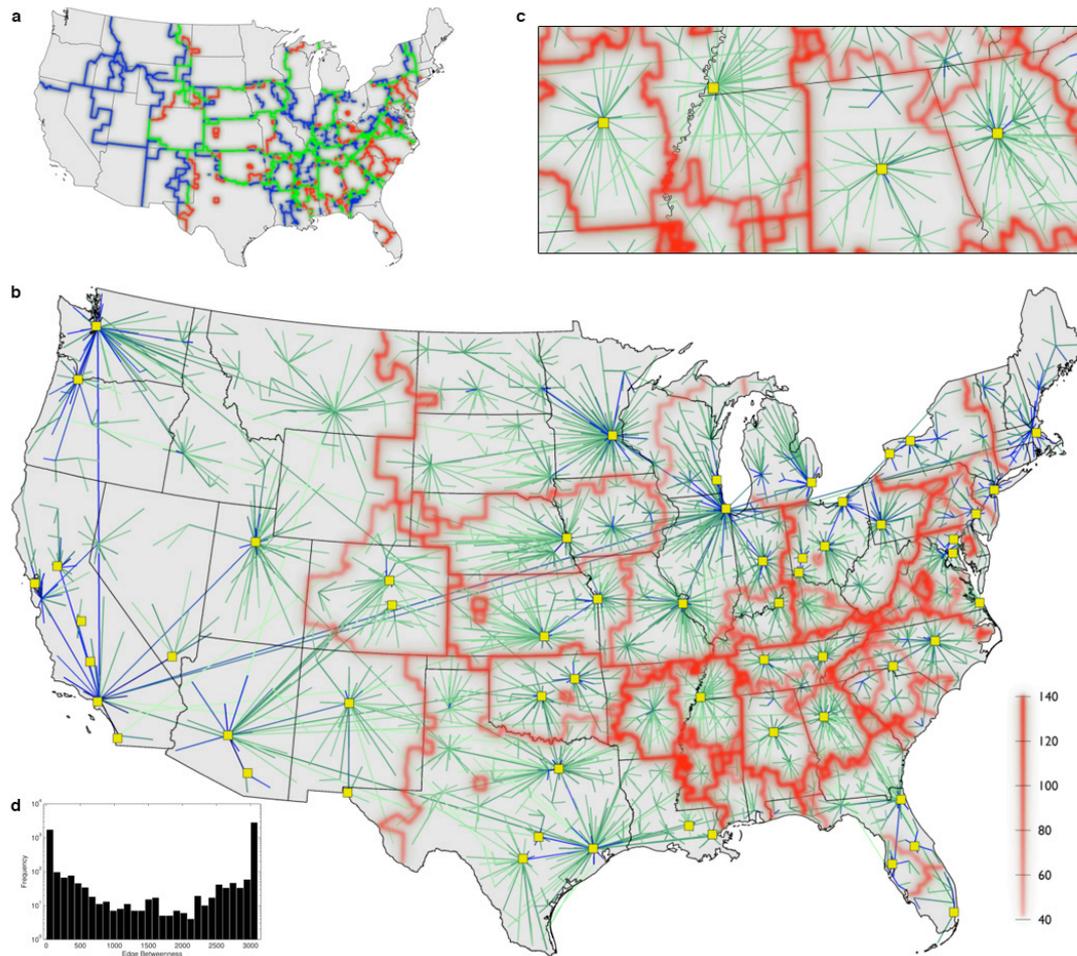
**Figure 1: Contact networks between selected individuals with swine flu in Israel**

Based on the contact networks, it was possible to determine the generation-time interval distribution for swine flu. The mean generation-time was found to be  $m=2.92$  and its standard deviation  $s=1.79$ , based on a generation time distribution of up to seven days. The contact networks also provided revealing information concerning who infected who in terms of the age-class structure of the patients.

### **The structure of multi-scale mobility networks**

Considerable progress was made concerning the computation of multi-scale community structure and effective geographic borders as described in Task 2, Theme 1 of the work package 2 by the **MPG** team. A multi scale proxy network has been obtained from the geographic circulation for the United States. In this project and while performing the research, an intense use has been made of the database server (see deliverable D2.1) that greatly expedited the analysis. In addition to existing methods for community detection in complex networks based on network modularity maximization, a new efficient technique based on the

analysis of sets of shortest path trees in the network has been developed during the research work.



**Figure 2: effective geographic borders implicitly encoded in multi-scale mobility networks (red). The discovery and computation of these structures heavily relied on the use of the mobility database server described in the Deliverable 2.1**

The novel approach not only permits the geographic identification of locations of effective borders but also provides a means to quantify their significance and strength. Furthermore it revealed which parts of multi-scale mobility networks are responsible for the shape and location of effective borders. The results of this project are currently under revision at Nature, a preprint of a manuscript is available at <http://arxiv1.library.cornell.edu/abs/1001.0943v1>. The results represent a promising base for investigations of multi-scale human mobility networks and effective order in Europe, a task that researchers plan to accomplish next within work package 2 based on proxy data for Europe and direct data on human mobility. Once multi-scale mobility data for Europe is incorporated in the database server, the analysis for Europe can be accomplished almost automatically.

The **ISI** team has been working the impact of multi-scale mobility networks on spatial epidemic spread. The paper published on the Proceedings of the National Academy of Sciences discusses the interplay of human mobility patterns like those between local metropolitan commuters and long-range airline travelers during a global epidemic. The work also tackles some general theoretical questions that concerns the basic understanding of the spatial spread of infectious diseases on the large scale: i) is there a most relevant mobility scale in the definition of the global epidemic pattern? ii) at which level of resolution of the epidemic behavior a given mobility scale starts to be relevant and to which extent? In order to fully consider the effect of multi-scale mobility processes the ISI team first integrates data of commuting patterns in five different continents with the airline transportation database and then develops a time-scale separation technique for evaluating the force of infection due to different mobility couplings and simulate global pandemics with tunable reproductive ratios. The results obtained from the full multi-scale mobility network are compared to the simulations in which only the large scale coupling of the airline transportation network is included. The analysis shows that while commuting flows are, on average, one order of magnitude larger than the long-range airline traffic, the global spatiotemporal patterns of disease spreading are mainly determined by the airline network. Short-range commuting interactions have on the other hand a role in defining a larger degree of synchronization of nearby subpopulations and specific regions which can be considered weakly connected by the airline transportation system. It also is possible to show that short-range mobility has an impact in the definition of the subpopulation infection hierarchy. In other words, global disease outbreaks tend to touch down at major travel hubs, generally major airport locations and spread out like a wave that follow local commuting patterns. The findings of the paper open the path to quantitative approximation schemes that calibrate the level of data resolution and the needed computational resources with respect to the accuracy in the description of the epidemics. Furthermore the developed techniques allow for an understanding of the level of data integration required to obtain reliable results in large scale modeling of infectious diseases and have already contributed to the improvement of the computational model used to provide estimates and projections of the H1N1 pandemic.

### **Other theoretical advances**

Other relevant theoretical results have been obtained by the **FGC-IGC** team that developed a generic framework for parameter estimation in dynamic models. The framework builds on different tools from the fields of Bayesian inference and sequential Monte Carlo (SMC), and is aimed at the fitting of dynamic models to available data. The fitting process generates posterior probability distributions for both the model's parameters and outputs. The original contribution of this work is in the way Bayesian and SMC computational methods available in the literature are combined to maximize generality and efficacy in the incorporation of available information into the estimation process. (A manuscript introducing the procedure and its application to influenza is in preparation). The FGC-IGC team has also developed mathematical models to systematically investigate the effect of seasonality on the final size of epidemics. They are currently applying the main concepts and methods to the study of influenza epidemics. Finally they have investigated different mechanisms for re-infection in malaria (manuscript submitted) and tuberculosis (manuscript in preparation). A theoretical exploration of how heterogeneity may confuse measures of re-infection has been performed. These works contribute to different tasks of the WP1 and WP2. It is worth also remarking that the work on data-driven computational modeling (reported in the next section) has also theoretical and foundational components.

## **Data-driven computational platform**

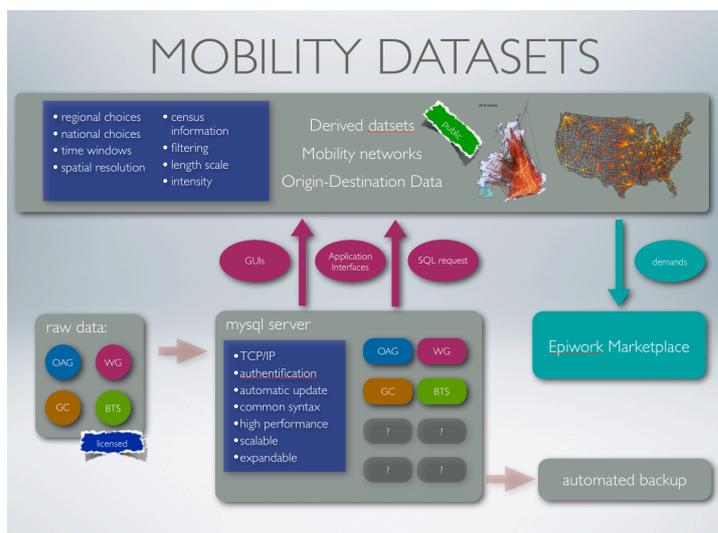
### **Data sharing and management tools**

A dedicated high performance MySQL database server was instantiated and runs on a MacOS Server platform (see Deliverable 2.1). This server manages and maintains large-scale databases of records of mobility related data employed in this project. The server maintains data integrity, is integrated in a backup system and guarantees data integrity. The server is operated on a user/password basis and can be accessed, updated and managed via TCP/IP protocol from anywhere and from all the participating partners, easing access to teams of the project. In order to manage mobility data from heterogeneous sources, a mobility specific data syntax was developed to which all dataset comply. This will ease the integration of additional dataset to be integrated in the system and zero friction development of analysis tools and software that builds on these datasets and implements automatic access to the server. So far three large scale database have been integrated:

1. Proxy data for mobility based on the geographic circulation of bank notes.

2. The worldwide air transportation network.
3. GPS data on trackable items in geocaching.com.

The raw data obtained from these sources was processed, compiled and embedded the database server. The server is maintained and managed by the MPI-DS team. They are in the process of finishing a web based GUI that will be made available to the project partners for issuing queries to the database and obtaining multi-scale mobility networks that can be incorporated into modeling frameworks.



**Figure 3: the structure of the mobility database server and its operation within the project**

Key to the design of the database and the common syntax for datasets managed by the server is a design that provides maximal flexibility in terms of regional and national parameters and the spatial resolution underlying derived origin destination data structures requested from the server as well as important census related data on the specified spatial resolution.

The *Epidemic Marketplace Platform* has progressed towards the implementation of a platform based on the integration of grid technology and services and software publicly available on the web to support the sharing and management of epidemic datasets and resources, as well as their rating, annotation, and selection. The **FFCUL** team is mainly involved in the development of this platform. The architectural requirements of the Epidemic Marketplace are directly related to the objectives of the Epiwork project and have been defined according to the feedback from its partners. It can be defined as a *distributed virtual repository*, a platform supporting transparent, seamless access to distributed, heterogeneous and redundant resources. It is a *virtual* repository because data can be stored in systems that are external to

the Epidemic Marketplace and it provides *transparent* access because several heterogeneities are hidden. The Epidemic Marketplace is composed of a set of interconnected data management nodes geographically distributed, sharing common canonical data models, authorization infrastructure and access interfaces. Data can be either stored in one or more repositories or retrieved from external data sources using authorization credentials provided by clients. Data can also be replicated among repositories to improve access time, availability and fault tolerance. Researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings. The Epidemic Marketplace Platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. Each dataset will come with a metadata file, signaling the date of submission, the last update, the source of the dataset, a basic profile (e.g., transportation network – Origin-Destination matrix), and a more thorough description of the dataset and the classification used. The data collection activity by the consortium started at M13, after the deployment of first functional prototype and its release to the consortium. All groups have been worked in integrating new data sets at various resolutions in the computational modeling tools. The **ISI, FBK and MPG** groups have gathered data from various sources of proxy networks, the census bureau of 40 countries in 5 continents and collected detailed data on the socio-demographic structure (e.g., age-structure, frequencies of household type and size, employment rates) of European population as provided by the Statistical Office of the European Communities (Eurostat) and integrated with data provided by the National Statistical Offices for countries not covered by Eurostat. Along those, data at the worldwide scale concerning multimodal transportation have been collected and cured for their integration in the simulation software. The data have been integrated in both the large scale, stochastic, spatially explicit, individual-based simulation model developed by the FBK team and the Global Epidemic and Mobility model platform developed by the ISI team.

### **Data-driven simulations for case studies analysis and epidemic forecasts.**

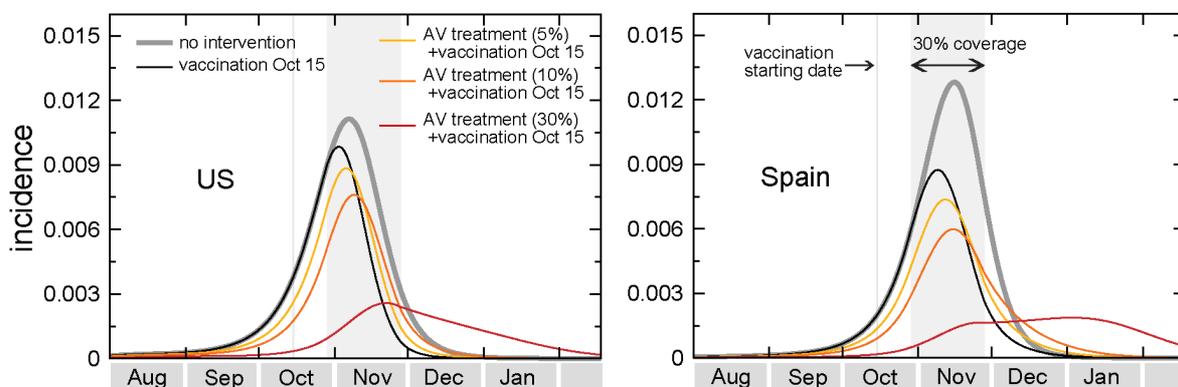
The FBK, ISI and MPG teams have been used different simulation and computational approaches to provide real time, quantitative analysis of the spatial spreading of the H1N1 pandemic.

**Agent based models.** The **FBK** team developed a stochastic, spatially structured individual-based model, considering explicit transmission in households, schools and workplaces, to simulate the spatiotemporal spread of an influenza pandemic in Italy and to evaluate the efficacy of interventions based on age-prioritized use of antivirals in terms of cumulative attack rate and excess mortality reduction under different scenarios. Results suggest that governments stockpile of influenza antiviral drugs suffice to treat approximately 25% of their populations. In countries with limited antivirals stockpile, providing prophylaxis to younger individuals is an option that could be taken into account in preparedness plans. In countries where the number of antivirals stockpiled is well below 25% of the population, priority should be decided based on age-specific case fatality rates. However, late detection of cases (administration of antivirals 48 hours after the clinical onset of symptoms) dramatically affects the efficacy of both treatment and prophylaxis (academic paper published on BMC Infectious Diseases). The FBK team has then extended the model to the entire European populations leveraging on the integration of air and railway transportation data. The analysis has shown that the impact of the epidemic in the European countries is highly variable because of marked differences in the socio-demographic structure of the European populations. The cumulative attack rate,  $R_0$ , and the peak daily attack rate depend heavily on socio-demographic parameters, such as the size of household groups and the fraction of workers and students in the population (academic paper published on Proceedings of the Royal Society B).

**Multi-Layer Epidemic Model.** The **MPG** team employed performance computational techniques and multi-layer, large-scale computer simulations to project the time course of the H1N1 flu epidemic in the United States. The simulations yielded projections and risk assessments of the epidemic outbreak in a worst-case scenario, in which no containment measures are taken to mitigate the spread. The approach was based on the current knowledge of the disease parameters and took into account the backbone of spatial spread: a precise estimate of human mobility on spatial scales between a few and a few thousand kilometers. The projections resolved the expected dynamics down to the county scale (3,109 counties in mainland United States). Details of the modeling approach are not yet published but are available online at <http://rocs.northwestern.edu>. The key factors in the MPG modeling

approach are very accurate human mobility datasets on scales from a few to a few thousand kilometers based on human mobility proxies that included small scale daily commuting traffic, intermediate traffic, and long distance travel by air. The simulations consisted of multiple layers, each layer possessing an increasing degree of accuracy and complexity. The final projections are done with a fully stochastic model that incorporates the inherent randomness in disease dynamics that is particularly important at the onset of an epidemic when the number of infected individuals is small compared to the whole population.

**Global epidemic and mobility model.** The ISI team have been using the Global Epidemic and Mobility (GLEaM) model to provide real time forecast on the unfolding of the H1N1 epidemic worldwide. This modeling effort has been unique as it has been the only one attempting to obtain quantitative results worldwide. The necessity to provide new way to obtain real estimates for the disease parameters have pushed the team to work on a new methodology that perform a likelihood analysis of the model with respect to chronological data of the diffusion processes. This methodology allowed early estimates of the transmission potential of the H1N1 virus by taking advantage of the multi-scale diffusion processes defined by the population mobility networks. This is the only model coupling countries worldwide and this feature is extremely relevant in evaluating the time pattern of emerging infectious diseases. For instance, given a set of initial conditions for a local outbreak of new strain of influenza, the timeline of the arrival of the epidemic in each country and the ensuing activity peak is mainly determined by the human mobility network that couples different region of the world. By looking at individual countries in isolation, any estimate of the epidemic timeline is going to be based on assumptions about imported cases from the rest of the world.



**Figure 4: Incidence curves for US and Spain for different intervention scenarios. The gray bar indicates the time period during which the immunization takes effect.**

GLEaM, instead, has the human mobility built-in, allowing to consistently simulate the mobility of infectious individuals on the global scale, thus providing ab-initio estimates of the epidemic timeline in each country or urban area without assumption on mobility and case importation. The model uses a maximum likelihood analysis on the GLEaM model parameters fitting the data of the chronology of the H1N1 epidemic. This is done by generating one million simulations on the worldwide scale of the pandemic evolution and finding the set of parameters that best fit the actual evolution of the pandemic. This has allowed us to estimate the transmission potential of the disease and the seasonality features. The algorithm models the seasonality by rescaling the  $R_0$  value in the Northern and Southern hemisphere taking into account seasonal changes. The free parameter to estimate is the value of the scaling factor during summer months,  $a_{\min}$ , representing the degree of dependence of the observed swine flu epidemic on the seasonality effects ( $a_{\min}=1$  if no seasonality is in place). By using the seasonality scaling, GLEaM provides an early assessment of the future unfolding of the epidemic in the different hemispheres. The model has been used to provide short-term predictions of the early unfolding of the epidemic in the USA and Europe. The results for the US have been found to be in very good agreement with the approach by the MPG team, thus providing evidence that the modeling strategies proposed in this project offers stable and consistent results. The obtained results have been also validated with a posteriori analysis with the real data collected by the CDC in the months of May and June. The agreement between the predictions and the actual unfolding of the pandemic has been proven to be remarkable. The GLEaM approach has then been used to provide in the month of June and July long-term prediction of the occurrence of the epidemic activity peak in the Northern hemisphere countries in the winter. The method anticipated an early peak occurring in October/November in most of the countries. The predictions, of a quantitative nature (peak week and relative 95% reference range), have been published in early September on BMC Medicine. This is the only paper so far that has attempted a quantitative forecast of the activity peaks. The predictions contained in the paper have been validated against the real data provided by agencies of more than 40 countries. The results show a very good agreement between predictions and real data with offset of at most two weeks. These findings provide a strong and remarkable test of the quantitative level of the prediction offered by computational methods. The GLEaM model has been used also for the analysis of vaccination campaigns, the effect of antiviral systematic use, the backtrack estimates of the actual number of cases in

the Mexico, the projections for ICU occupancy and antibiotic usage. All papers have been published in disciplinary journals (Emerging Health Threats, Plos Currents Influenza).

### ICT monitoring and reporting system

The new influenza H1N1 or swine flu pandemic in Europe encouraged all the researchers in the project Work Package 5, under the direction of AIBV team, to speed up the design of a central database, including hardware and software development, the layout of the new Influenzanet website, the first outlines of the mobile phone data and contact pattern collection and the introduction of IMS in the UK (July) and starting preparations in Sweden (Autumn '09). First of all, the development of a “gold standard” questionnaire, crucial for the project’s aim of gathering cross-country surveillance data that do not suffer from national-related biases and methods, has been brought forward by the need of having an appropriate survey best describing the different countries requirements in the data collection method during a public health emergency period. The first draft has been developed in July, the second draft has been tuned in August 2009, before the striking of the Autumn wave of H1N1 pandemic.

During the first WP5 meeting (25-26 May 2009) in Amsterdam and Epiwork project meeting (15-18 November 2009), the discussion among the partners has led to the preparation in December 2009 of a draft for the design of the European centralized database, again brought forward by the occurrence of the H1N1 pandemic. The presentation will be in March 2010, well ahead of schedule. The database will provide easy-to-access, reliable, uniform and standardized epidemiological data. Database contents of the local IMS will be consistent and have standard cross-country information. Moreover, WP5 partner FGC-IGC developed a platform that downloads daily the latest data from all Influenzanet systems and converts those different data formats into one standardized format. During the early phases of the H1N1 pandemic, database contents of the already existing local IMS platform have been collected in a centralized repository that provided an extensive source of epidemiological data. The FGC-IGC team has confirmed the consistency of the Influenzanet system over its 7 years of existence in four different countries by applying linear regression model on Influenzanet data and EISN data and Google Flu Trends data, respectively. They have identified individual variables associated with increased risk of ILI, by applying logistic regression models to explore the association between individual and household-level covariates and the occurrence

of at least one ILI episode during a flu season, In the Influenzanet cohort, having diabetes or asthma, living with a child, being female, belonging to young age group, and being a heavy smoker, were all independent predictors of the risk of having at least one ILI episode during a flu season. These findings are in line with the influenza literature.

The pilot phase of IMS in the new countries, foreseen for the winter season 2010/2010, was forcedly anticipated. The IMS was deployed by **LSHTM** in UK in the early summer of 2009. The UK IMS ([www.flusurvey.org.uk](http://www.flusurvey.org.uk)) has been collecting data on influenza since July. The implementation of this was brought forward due to the H1N1 pandemic. Information is currently being collected from over 5500 participants across the UK (making the UK survey the third largest survey within the network). Hence, implementation of this was ahead of schedule. Two papers are in draft form from the results of the ICT monitoring during the pandemic of 2009. The first documents changes in the way individuals accessed treatment over the course of the H1N1v epidemic. Importantly, there were major changes in access to health care during the epidemic, with individuals being far more likely to contact health services during the early part of the epidemic than later on. This was partly because in the National Pandemic Flu Service (NPFs) was introduced (a service that allowed individuals to collect antivirals without visiting a general practitioner). This highlights the utility of this internet-based collection methods, as all traditional forms of data collection require individuals to contact the health service in some way. If individuals change the way they access health services, then the traditional surveillance data will be biased. The ICT allowed the UK to monitor and correct for this bias in the epidemic in real-time. The data from the ICT monitoring system were used to inform pandemic-related decision-making in the UK on a range of topics from the period of absenteeism to the physician consultation and hospitalisation rates. John Edmunds is a member of UK Government's pandemic scientific advisory committee, and the results from flusurvey were used to inform the modelling committee on epidemic progress and changes to access care patterns etc, during the epidemic. The data were also used to inform a real-time model of pandemic influenza in the UK and an associated economic evaluation of alternative vaccination policies. This paper will shortly be available.

Starting from May 2009, **CREATE-NET**, **AIBV** and **ISI** started to cooperate with Dr. Ben Reis from the Harvard Medical School (Boston, US) on the implementation of a social data

gathering application. At the first WP5 meeting in May 2009, Dr. Reis presented his experience in developing healthcare applications over the Facebook platform, thus exploiting the power of social networking to engage a wide number of users into a healthcare application. In particular, he reported on healthy social (<http://www.healthysocial.org/>), a Facebook application for running health surveys and for promoting healthy behavior. During the meeting in May, the WP5 partners agreed to collaborate with Dr. Reis in order to adapt the existing healtysocial application to the requirements of the EPIWORK survey, and to jointly launch the initiative during the 2010 winter season. The work was completed by December 2009, with the platform being ready in English, Italian and Portuguese. A Dutch version is about to be finished. Researchers are currently working at official launch of the application, and preparing the media campaign that they plan to use to advertise the launch of the application. Most of the costs are working hours, covered by Dr. Reis and his team. AIBV will soon sign a subcontract with the Harvard Medical School on the coverage of graphics and editing of the Facebook Influenzanet application.

Due to the intervening new influenza pandemic the set-up and implementation of the a new population-based approach (PBA), recently developed and evaluated at the Swedish Institute for Infectious Disease Control (Smittskyddsinstitutet – SMI) was expedited. This dataset will provide statistical analysis and validation of the IMS and the existing General Practice-based sentinel system of influenza surveillance. The ethics approval for the PBA was obtained already before the end of the summer, and so was the agreement with Linewise, an IT subcontractor small business. In collaboration with researchers at SMI and the department of Medical Epidemiology and Biostatistics at Karolinska Institutet (MEB/KI), Linewise has built and tested the reporting system, including the Interactive Voice Response telephone service, the secure website, and the response database. Over and above the prescribed tasks for this first year, the project has also drawn a random population sample and included approximately 5500 active participants who started real live reporting in early September 2009. Data from the PBA were quite reasonable, and in terms of numbers much more reasonable than the existing sentinel reporting. These data have been extremely helpful in SMI's monitoring of the pandemic. Efforts to validate the PBA are already in progress. Criterion validity is studied using concurrent postal questionnaires that inquire about disease onsets in the previous week. The first publication, studying the comparability of data generated with Internet-based

reporting and data from Interactive Voice Response telephone contacts, is presently under review at an epidemiological journal.

### 3 Discussion and outlook

The first year of the project was extremely productive and the consortium has harvested a large number of scientific results and pushed forward the development of the data and computational tools. The H1N1 pandemic has undoubtedly represented an extra challenge strain to the consortium but at the same time has been a motivation to the consortium to work ahead of schedule and speed up in all work areas. Data collected during this year will result invaluable in the future to test the tools and concepts developed in the next years of the project. The level of communication among the consortium partners has been very high since the beginning. In most of the cases, results achieved in one WP would have been impossible without the work of other WPs. Teams have been collaborating in the set up of new IMS services, in data and theoretical results sharing. The scientific standing of the consortium and its output is witnessed by the list of papers and talks reported below. In summary we believe the project is moving ahead with the enthusiastic support of all the teams and we believe that the project has already poised itself in the best position to achieve and deliver the results promised in the initial plan.

### 4 List of publications and scientific talks

The project puts great emphasis on publication in high impact scientific communication channels such as papers and major conferences. It also envisions communicating results at top international conferences. During the first year of life of the project the scientific outreach is simply stated by the sheer numbers of publications (28 papers in peer reviewed journals) and presentations at conferences (30 talks, lectures and seminars).

#### Publications

M. Aguiar, N. Stollenwerk, B. Kooi, *Torus bifurcations, isolas and chaotic attractors in a simple dengue fever model with ADE and temporary cross immunity*. Intern. Journal of Computer Mathematics, in press (2009).

N. Bacaër, M.G.G. Gomes, *On the final size of epidemics with seasonality*, Bulletin of Mathematical Biology. Bull Math Biol 71:1954-66, 2009.

S. Ballesteros, A. Camacho and B. Cazelles, *Introducing gradual antigenic drift in co-circulating cross reactive antigenic cluster models*, Proceedings of 9<sup>th</sup> Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2009, edited by Jesus Vigo Aguiar et al., Salamanca, 2009, pp. 1471—1482.

Y. Berchenko, Y. Artzy-Randrop, M. Teicher, L. Stone, *Emergence and size of the giant component in clustered random graphs with a given degree distribution*, Physical Review Letters, vol. 102 (13) 138701, 2009.

J.P. Boto and N. Stollenwerk (2009). "Fractional calculus and Levy flights: modelling spatial epidemic spreading", Proceedings of 9<sup>th</sup> Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2009, edited by Jesus Vigo Aguiar et al., Salamanca, pp. 177—188.

A. Huppert, G. Katriel, R. Yaari, U. Roll, R. Balicer, L. Stone, *Mathematical models as a tool for facing the influenza pandemic*, HaRefuah (Hebrew) 149, p 4-8, 2010.

H. Katriel, L. Stone, *Pandemic influenza dynamics and the breakdown of herd immunity*, PLoS Curr Influenza, October 1: RRN1046, 2009.

J. Martins, A. Pinto, N. Stollenwerk, *A scaling analysis in the SIRI epidemiological model*, Journal of Biological Dynamics, 479 - 496, 2009.

P. Rodrigues, A. Margheri, C. Rebelo, M.G.M.Gomes, *Heterogeneity in susceptibility to infection can explain high reinfection rates*. J Theor Biol 259:280-90, 2009.

N. Stollenwerk, J.P. Boto, *Reaction-superdiffusion systems in epidemiology, an application of fractional calculus*, Proceedings of the International Conference on Numerical Analysis and Applied Mathematics, ICNAAM, 2009.

D. Brockmann, V. David and A. Morales Gallardo, in Diffusion Fundamentals III, C. Chmelik, N. Kanellopoulos, J. Kärger and D. Theodorou (eds.), Leipziger Universitätsverlag (2009).

D. Brockmann, *Human Mobility and Spatial Disease Dynamics*, in Reviews of Nonlinear Dynamics and Complexity, H. G. Schuster (ed.), Wiley-VCH (2009).

D. Balcan, V. Colizza, B. Goncalves, H. Hu, J.J. Ramasco, and **A. Vespignani**, *Multiscale mobility networks and the spatial spreading of infectious diseases*, Proceedings of the National Academy of Sciences USA. **106** 21484-21489 (2009). [Featured on the journal cover](#)

D. Brockmann, *Following the money*, Physics World, Feb 2010.

L. F. Lopes, J. M. Zamite, B. C. Tavares, F. M. Couto, F. Silva and M. J. Silva. *Automated Social Network Epidemic Data Collector*. Inforum, September 2009.

M. J. Silva, F. A. B. da Silva, L. F. Lopes, F. M. Couto. *Building a Digital Library for Epidemic Modelling*. Invited Paper. ICDL 2010. The 3rd International Conference on Digital Libraries. February, 2010.

S. Merler and M. Ajelli, *The role of population heterogeneity and human mobility in the spread of pandemic influenza*, Proceedings of the Royal Society B, 277: 557-565, 2010

S. Merler, M. Ajelli, C. Rizzo, *Age-prioritized use of antivirals during an influenza pandemic*, BMC Infectious Diseases, 9:117, 2009.

M. Ajelli and S. Merler, *An individual-based model of hepatitis A transmission*, Journal of Theoretical Biology, 259(3):478-488, 2009.

P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Goncalves, J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W V. den Broeck, V. Colizza, and A. Vespignani, *Modeling vaccination campaigns and the Fall/Winter 2009 activity of the new A(H1N1) influenza in the Northern Hemisphere*. Emerging Health Threats Journal. **2**, e11, 2009.

D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W V. den Broeck, V. Colizza, and A. Vespignani, *Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility*, BMC Medicine, 7 45, 2009.

V. Colizza, **A. Vespignani**, N. Perra, C. Poletto, B. Goncalves, H. Hu, D. Balcan, D. Paolotti, W V. den Broeck, M. Tizzoni, P. Bajardi, and J.J. Ramasco, *Estimate of Novel Influenza A/H1N1 cases in Mexico at the early stage of the pandemic with a spatially structured epidemic model* Public Library of Science Currents: Influenza. RRN1129, 2009.

D. Balcan, V. Colizza, A.C. Singer, C. Chouaid, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J.J. Ramasco, N. Perra, M. Tizzoni, D. Paolotti, W V. den Broeck, A. J. Valleron, and **A. Vespignani**, *Modeling the critical care demand and antibiotics resources needed during the Fall 2009 wave of influenza A(H1N1) pandemic* Public Library of Science Currents: Influenza. RRN1133, 2009.

V. Colizza, A. Vespignani, *The Flu Fighters*, Physics World, Feb 2010.

A. Godinho, *Epiwork: Developing the framework for a European forecast infrastructure*. Research Review Magazine 9: 49, 2009.

S.P. van Noort, N. Stollenwerk, L. Stone, *Analytic likelihood function for data analysis in the starting phase of an influenza outbreak* Proceedings of CMMSE 2009, ISBN 978-84-612-9727-6, 2009.

I.H. Friesema, C.E. Koppeschaar, G. A. Donker, F. Dijkstra, S.P. van Noort, R. Smallegenburg, W. van der Hoek, M.A. van der Sande, *Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands*. Vaccine 27:6353-7, 2009.

The project has been also involved and promoting two scientific events:

- [Workshop ‘Frontiers in the computational modeling of disease spreading’ @ ICCS2010](#), Amsterdam, The Netherlands, May 31-June 2, 2010
- [White Workshop on Mathematical Biology](#), Trento, Italy, December 17-19, 2009

## International Conferences and seminars

Gabriela Gomes (FGC-IGC)  
Workshop Design and Analysis of Infectious Disease Studies  
November 2009, Oberwolfach, Germany.  
*High rates of reinfection tuberculosis: The selection hypothesis.*

Gabriela Gomes (FGC-IGC)  
Workshop on Modelling, Computation, and Measurement of Multiple Carriage  
December 2009, Saariselkä, Finland  
*Interpreting genetic variation in pathogen populations.*

Gabriela Gomes (FGC-IGC)  
Workshop on Theoretical Epidemiology, Centro Internacional de Matemática,  
January 2010, Coimbra, Portugal  
*Integrative Epidemiology.*

J.P. Boto and N. Stollenwerk (FFCUL)  
9<sup>th</sup> Conference on Computational and Mathematical Methods in Science and Engineering,  
June 30, July 1-3, Gijón, Spain  
*Fractional calculus and Levy flights: modelling spatial epidemic spreading.*

Nico Stollenwerk, J.P. Boto, (FFCUL)  
International Conference on Numerical Analysis and Applied Mathematics,  
18-22 September 2009, Rethymno, Crete, Greece.  
*Reaction-superdiffusion systems in epidemiology, an application of fractional calculus*

Shlomo Havlin (BIU)  
International Conference on Complex Systems  
Feb. 23 - 25, 2009, Shanghai, China  
*Percolation and immunization of complex networks*

Shlomo Havlin (BIU)  
Complex Networks: Topology and Dynamics  
June 4, 2009, Tel-Aviv, Israel  
*Statistical Physics and Complex Networks*

Shlomo Havlin (BIU)  
International Workshop on Coping with Crises in Complex Socio-Economic Systems  
June 8 - 12, 2009, Zurich, Switzerland  
*Efficient Immunization Approaches to Avoid Epidemic Spreading*

Shlomo Havlin (BIU)  
NETSCI 09 International Workshop and Conference on Complex Networks and their Applications  
June 29 - July 3, 2009, Venice, Italy  
*Novel Percolation Models in Complex Networks*

Roni Parshani (BIU)  
International Conference on Complexity and Interdisciplinary Sciences,  
July 12-18, 2009, Chengdu, Sichuan, P. R. China,  
*SIS epidemic threshold on random networks*

Stefano Merler (FBK)  
NETSCI 09 International Workshop and Conference on Complex Networks and their Applications  
June 29 - July 3, 2009, Venice, Italy  
*Comparing large-scale computational approaches to epidemic modeling: Agent based versus structured metapopulation models.*

Marco Ajelli (FBK)  
NETSCI 09 International Workshop and Conference on Complex Networks and their Applications  
June 29 - July 3, 2009, Venice, Italy  
*Factors affecting the spread of an epidemic in Europe: population heterogeneity and human mobility.*

Marco Ajelli (FBK)  
White Workshop on Mathematical Biology  
December 17-19, 2009, Trento, Italy  
*A dynamic individual based model of hepatitis A transmission.*

Stefano Merler (FBK)  
White Workshop on Mathematical Biology  
December 17-19, 2009, Trento, Italy  
*The role of population heterogeneity and human mobility in the spread of pandemic influenza.*

Mario J. Silva (FFCUL)  
The International Conference on Digital Libraries New Delhi, India,  
February, 2010.

Luis F. Lopes, João M. Zamite, Bruno C. Tavares, Francisco M. Couto, Fabrício Silva and Mário J. Silva (FFCUL)

September 2009, INForum – Simposio de Informatica  
*Automated Social Network Epidemic Data Collector*

Alessandro Vespignani (ISI)  
The James Martin 21st Century School Seminar Series  
February 25, 2010, University of Oxford, UK  
*Predicting the Behaviour of Techno-Social Systems: How Informatics and Computing Help to Fight Off Global Pandemics.*

Vittoria Colizza (ISI)  
Global Health Conference – Global Flows in Global Health: Inter-Asian Connections  
Jan 4-8, 2010, United Arab Emirates University, Al-Ain, UAE  
*Epidemic Science in real time: the H1N1 case* [keynote]

Alessandro Vespignani (ISI)  
Symposium ‘Frontiers in Network Science’  
September 28-30, 2009, Berlin, Germany  
*Predicting the behavior of techno-social systems* [keynote]

Vittoria Colizza (ISI)  
Symposium ‘Frontiers in Network Science’  
September 28-30, 2009, Berlin, Germany  
*Epidemic Science in real time: the H1N1 case* [keynote]

ECCS09 - European Conference on Complex Systems 2009  
September 21-25, 2009, University of Warwick, UK  
*Worldwide spread of the unfolding swine flu epidemic: early assessment and predictions* [contributed]  
C. Poletto  
*Complex Techno-Social Networks for Epidemic Forecasting* [invited]  
B. Goncalves

Alessandro Vespignani (ISI)  
NetSci09 - International Workshop on Network Science 2009  
June 29 - July 3, 2009, Venice, Italy  
*Multiscale networks and forecasting techno-social systems: Planning for pandemic outbreaks in real time* [keynote]

Vittoria Colizza (ISI)  
ICCS 2009 - International Conference on Computational Science “Compute. Discover. Innovate.”  
May 25-27, 2009, Baton Rouge, Louisiana, USA  
*Computational epidemiology: a new paradigm in the fight against infectious diseases* [keynote]

Sander van Noort (FGC-IGC)  
Influenzanet: Internet-based monitoring system for ILI.  
Annual Meeting of the Influenza Surveillance Network in Europe.  
ECDC, Stockholm, Sweden.  
June 2009.

Sander van Noort, Vitor Faustino (FGC-IGC)  
July 2009, ISC, Salvador, Brazil.  
*On the use of the internet to monitor infectious diseases.*

Gabriela Gomes (FGC-IGC)  
September 2009, Workshop on Mathematical Modelling of Epidemics, University of Bristol, UK.  
*Influenzanet.*

Ken Eames (LSHTM)

Seminar  
October 2009, Imperial College, London;  
*Weighted networks, mathematical models, and a bit of flu;*

Ken Eames (LSHTM)  
May 2009, University of Georgia;  
*Human social contact patterns and the spread of infection;*  
*Ecology and Evolution of Infectious Disease conference,*

Shlomo Havlin (BIU)  
Seminar  
Dec. 7 2009, Northeastern University,  
*Percolation in interconnected networks*