Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

http://www.epiwork.eu

Project no. 231807

D8.5 Final non-technical report on the project

Period covered: Month 1 - 54	Date of preparation: 08/15/13
Start date of project: February 1st, 2009	Duration:
Due date of deliverable: Month 54th	Actual submission date: 08/25/13
Distribution: public	Status: Final

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:

- FGC-IGC
- TAU
- MPG
- BIU
- FBK
- FFCUL

Change log

Version	Date	Amended by	Changes
1	**30-08-13**	**ISI**	

Table of contents

Final non-technical report on the project

Introduction

EPIWORK is a project sponsored by the Future and Emerging Technologies program of the European Community proposing a multidisciplinary research effort aimed at developing the appropriate framework of tools and knowledge needed for the design of epidemic forecast infrastructures.

The huge flow of quantitative social, demographic and behavioural data becoming available nowadays motivates the development of innovative technologies that can improve the traditional disease-surveillance systems, providing faster and better-localized detection capabilities. For the first time, ICT and computation enable the study of epidemic in a comprehensive fashion addressing the complexity inherent to the biological, social and behavioural aspects of health related problems. In this context, the EPIWORK project proposed a visionary research aimed at giving scientific foundations to the development of the needed modelling, computational and ICT tools to predict the disease spreading in complex social systems.

Epiwork has gathered more than 100 researchers from areas as diverse as epidemiology, medicine, ICT, computer science, physics and mathematics. It has carried out multidisciplinary research effort along the following main research thrusts

i) the foundation and development of the mathematical and computational methods needed to achieve prediction and predictability of disease spreading in complex social systems;

ii) the development of large scale, data driven computational models endowed with a high level of realism and aimed at epidemic scenario forecast;

iii) the design and implementation of original data-collection schemes motivated by identified modeling needs, such as the collection of real-time disease incidence through innovative web and ICT applications.

iv) the set up of a computational platform for epidemic research and data sharing that will generate important synergies between research communities and countries.

The project developed the basic elements for a large scale infrastructure and it has delivered all the planned important components (http://www.epiwork.eu/resources/): the first version of the Epidemic Marketplace that is the first publicly available data repository of this type and a major milestone for the project (http://www.epimarketplace.net); a prototype modeling computational multiscale platform integrating models ranging from ABM to large structured metapopulation approaches (http://www.gleamviz.org/simulator); the Internet based monitoring (IMS) system that comprises ten platforms in different countries aimed at real-time assessment of disease incidence and large-scale surveillance of population (http://www.influenzanet.eu). The theoretical works has produced major advances in epidemic network analysis as well as in the definition of new computational techniques that through an agent based model approach allows the estimate of contact matrices by age in different social settings: household, school, workplace and general community. The project has also produced several exemplary cases of contributions to/from complexity science ranging from computational tools for the analysis of complex multiscale networks to foundational understanding on non-linear effects and tipping points in dynamical processes in heterogeneous/structured populations. The various components of the projects have been actively involved in the H1N1 pandemic emergency by performing data gathering, computational analysis and monitoring (http://www.epiwork.eu/2009-h1n1-flu/) in close contact with national and International agencies, including the JRC crisis unit, the ECDC and Institutes or Ministry of Health of most of the countries represented in the project.

Project context and objectives

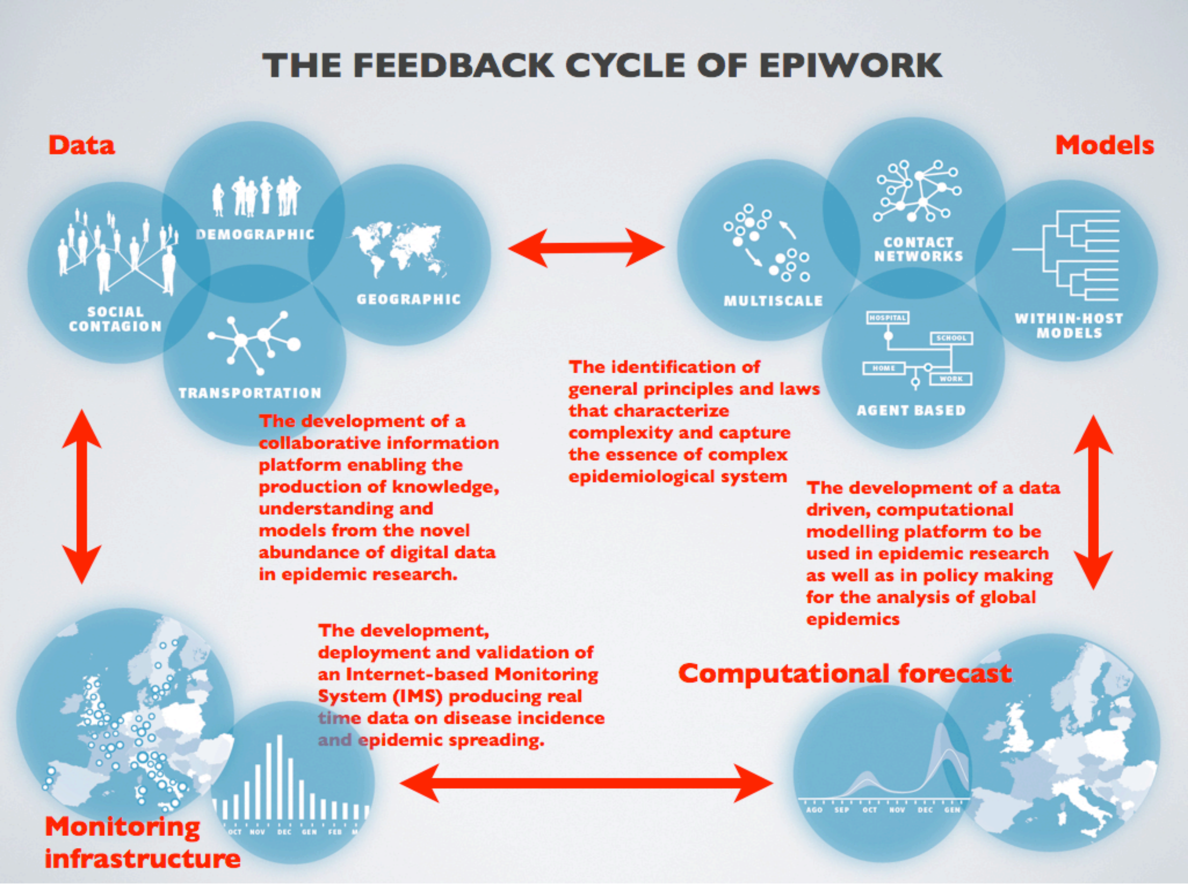
A fully operational, accurate and reliable epidemic forecast infrastructure nowadays faces problems related to the lack of appropriate models to understand how an infectious disease spreads in the real world, lack of extensive and accurate epidemiologically relevant data (from societal data to epidemic 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 in particular at exploring the following *work areas* as the major research themes directly matching the objectives of this proposal: i) ***Modelling and theoretical foundations; ii) Data-driven computational platform***; ***iii) ICT monitoring and reporting system.*** The work plan is organized around six distinct scientific work packages (WP1-WP6) aimed at providing a virtuous feedback cycle between tool development, data collection, analysis and modelling.

Parallel scheduling of the work packages is necessary to jump-start the cycle and the Inter-WP validation. The research plan is structured so as to foster a fruitful interplay between the various components of the project. WP1 and WP2 are aimed at exploring theoretical issues in the area of epidemic modelling 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 with

WP1 and WP2 of data and algorithms. WP5 and WP6 is aimed at developing, set-up and deployment of innovative web monitoring and data gathering tools that provide a continuous stream of data to WP3-WP4 and is informed by constant feedback on the modelling needs in terms of data gathering by WP1 and WP2.



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 identification of general principles and laws that characterize complexity and capture the essence of complex epidemiological system

The development of a data driven, computational modelling platform to be used in epidemic research as well as in policy making for the analysis of global epidemics

The development, deployment and validation of an Internet-based Monitoring System (IMS) producing real time data on disease incidence and epidemic spreading.

The common research agenda of the consortium teams, which work in a coordinated way on the various tasks, favours a closer interchange of ideas and knowledge among the groups and the various components of the project in a truly interdisciplinary collective effort.

Each WP includes several core disciplines expertise and it is anchored to the epidemiology area by the presence of mathematical biologists, epidemiologists and public health experts. These groups will provide the main research questions, the basic disease parameters choices and the relevant complex features of epidemiological systems as well as their contribution in the development of cross-fertilized and novel approaches targeted in the WPs. WP7 is management. This project was initiated by a group of senior scientists, working at the best research institutions in Europe. The Institute of Scientific Interchange (ISI), Turin, provides the management of the project and the coordination of the consortium. Finally, WP8 is devoted to outreach and dissemination activities.

The project description readily conveys the need for a wide variety of skills competencies, ranging from complex systems theory and computational modeling to computer science, statistical physics and epidemiology. The consortium consists of **12 teams in 8 different countries** that provide these competencies and that have skills and expertise that are documented by numerous publications and participation and leadership roles in previous European network projects:

1. ***ISI – Institute for Scientific Interchange Foundation (Italy)***
2. ***FGC-IGC – Fundaçao Calouste Gulbenkian - Instituto Gulbenkian de Ciência, (Portugal)***
3. ***TAU – Tel Aviv University (Israel)***
4. ***MPG – Max Planck Institute for Dynamics and Self-Organization (Germany)***
5. ***AIBV – Acquisto Inter BV (The Netherlands)***
6. ***London School of Hygiene and Tropical Medicine (UK)***
7. ***SMI – The Swedish Institute for Infectious Disease Control (Sweden)***
8. ***KULeuven – Katholieke Universiteit Leuven (Belgium)***
9. ***BIU – Bar Llan University (Israle)***
10. ***FBK – Fondazione Bruno Kessler (Italy)***
11. ***CREATE-NET – Center for Research and Telecommunication Experimentation for NETworked communities***
12. ***FFCUL – Faculty of Sciences University of Lisbon***

Main S&T results/foregrounds

Summarizing the work, achievements, results of a project of the extent and duration of Epiwork it is a challenge per se. While at first instance one is tempted to go along the path of the WP structure of the project, we believe that a more compelling presentation can be achieved by focusing on the *work areas* that constitutes the major research themes directly matching the objectives of this proposal:

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

A specific section will be devoted to the work unexpectedly carried out during the H1N1 pandemic. Furthermore we will highlight the specific contributions to and from the science of complex systems that was one of the objectives of the COSI-ICT call to which the Epiwork project is responding The hope is that in this way we will avoid to parcel the results and achievements of the projects in a way that makes hardly understandable the coherent effort beyond the work carried out along the years and specifically avoid duplicate information that can be found in deliverables and other reporting items of the project.

- *Modelling and theoretical foundations*

Looking back, at the completion of the Epiwork project, it would appear that most, if not all, of the original theoretical goals of the project have been completed. Through annual Epiwork meetings, workshops and communication a symbiosis between the different teams developed in time. The groups synchronized activities throughout the project and were working on similar problems in different ways creating an unusual diversity in terms of the possible solutions. The appearance of the pandemic in 2009, at the start of Epiwork, also helped the groups focus on the common goal of pandemic modelling. The work performed by the teams involved in the theoretical and algorithmic foundational part of the project are clearly defining the state of the art in several areas related to contact networks, seasonal forcing, mitigation and control measures, multi-scale mobility networks. It is worth noticing that the work of the Modelling and Theoretical Foundations area paid off in other areas of the project. This is a sign that the theoretical work impacted the more applied activities and that the foundational and applied part of the project latched in. The most spectacular advances have been obtained in the following specific topics:

- Analysis and modelling of contact patterns and the definition of models with specific age structured contact matrices that have been progressed across the board of the project's WPs. This work has spurred from the theoretical analysis in WP1 (Katriel et al. 2011), and has

reverberated in the work Brooks-Pollock and Eames, 2011 made in collaboration with the activities of WP5 and found a computational application in the context of the epidemic modeling platform (WP4) PLoS Computational Biology, 6(12): e1001021, 2010. They have indeed developed a new computational and modelling tools for studying the dynamics of an epidemic in its initial stages that use both available incidence time series and data describing the population's contact network structure **(Katriel et al. 2011)**. The paper describes new techniques for parameter fitting that is based specifically on the contact structure mentioned above for Israeli 2009 swine flu. The basic model can be extended to include age–class structure, and a maximum likelihood methodology allows us to estimate the model's next-generation matrix by combining two types of data: (i) the incidence series of each age group, and (ii) contact network data that provide partial information of 'who-infected-who'. Unlike other approaches for estimating the next-generation matrix, the method developed here does not require making a priori assumptions about the structure of the next-generation matrix. Using a simulation study, it was found that even a relatively small amount of information about the contact network greatly improves the accuracy of estimation of the next-generation matrix. The method was applied in practice to estimate the next-generation matrix from the Israeli H1N1 pandemic data. In the paper "Pigs didn't fly, but swine flu" **(Brooks-Pollock and Eames 2011)**, use a relatively simple system of differential equations, to explore how changes in human behaviour and social mixing influenced the H1N1 pandemic. In collaboration with partners in WP5, they used Internet surveillance to calculate contact matrices (see figure below) and developed an age-structured model that incorporated changing contact rates over the year.

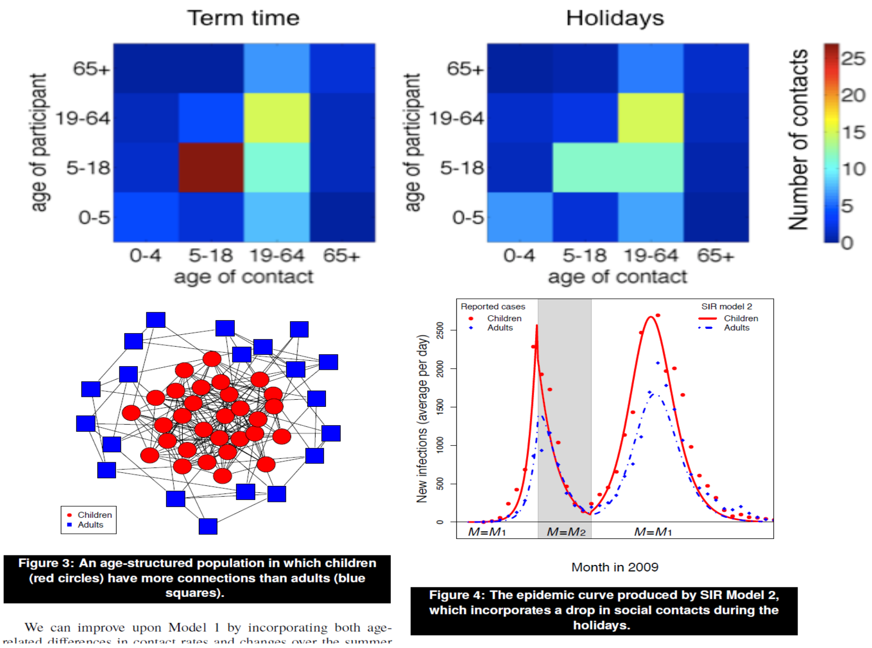


Figure 3: An age-structured population in which children (red circles) have more connections than adults (blue squares).

We can improve upon Model 1 by incorporating both age-

Figure 4: The epidemic curve produced by SIR Model 2, which incorporates a drop in social contacts during the holidays.

Fig1 : The LHSTM model analysis of swine flu using contact matrices and an age-class model.

- Seasonal and external environmental drivers

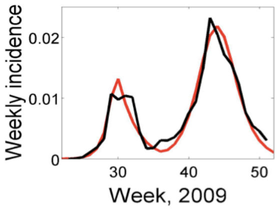


Fig.2 A comparison of the per-capita weekly incidence (black) and the best-fitting model output (red) from the age structured model

The work performed in this area has also touched upon some foundational issues such as the seasonal and external environmental drivers and the graph properties of multi-length scale mobility networks. The LSHTM team has been able to explain the epidemiological H1N1 pattern in the UK, resulting in a termination of the summer wave (as the reproduction number was driven below one), and the reactivation of the epidemic in September, by using an age-structured transmission model giving a remarkably good fit to the epidemic data (see Fig.1). Noticeably the model was developed by using data collected with IMS platform about contact patterns, showing the nice linkage between theoretical development and data gathering.

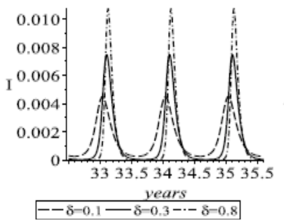


Fig.3 Epidemic curve changes considerably as the strength of seasonality d varies although the final attack rate is found to be mildly varying.

Extremely relevant is the various streams of work concerning time series analysis and the role of weather on the relation between influenza and influenza-like illness (see Fig.2). The results contained in this work will be extremely relevant in the real time analysis of the IMS platform data offered by influenzanet.eu. This is a sign that the theoretical work is impacting the more applied work packages and that the foundational and applied part of the project are finally matching together.

Both the TAU and Gulbenkian groups developed interesting theoretical work that was built around the notion of S0, the initial population susceptibility. Moreover, they both generated new theoretical results on the reinfection threshold, work that was motivated by an Epiwork conference. Many other similar examples of symbiotic interchanges are to be found in the Deliverable 1.5. All groups had a strong interest in modelling seasonal influenza, with Tel Aviv university implementing its forced "age-of-infection" model for Israeli data, while the Gulbenkian group used related techniques on Epiwork internet surveillance data, and the LSHTM group applied MCMC Bayesian sampling techniques for modelling UK surveillance data. All groups accounted for climate forcing in similar consistent ways.

- Multi-scale mobility networks

Considerable progress was made concerning the computation of multi-scale community structure and effective geographic borders. A key result of this study yields a parameter independent link measure termed link salience that is typically bi-modally distributed in transportation networks which means that it can be employed to classify essential links from those that are not. The link salience analysis has been pushed forward in order to provide a new interesting centrality measures in mobility networks. This result along with the analysis of dynamical infection trees paves the way to new clustering algorithms for the definition of surveillance regions.

A multi scale proxy network has been obtained from the geographic circulation for the United States. 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.

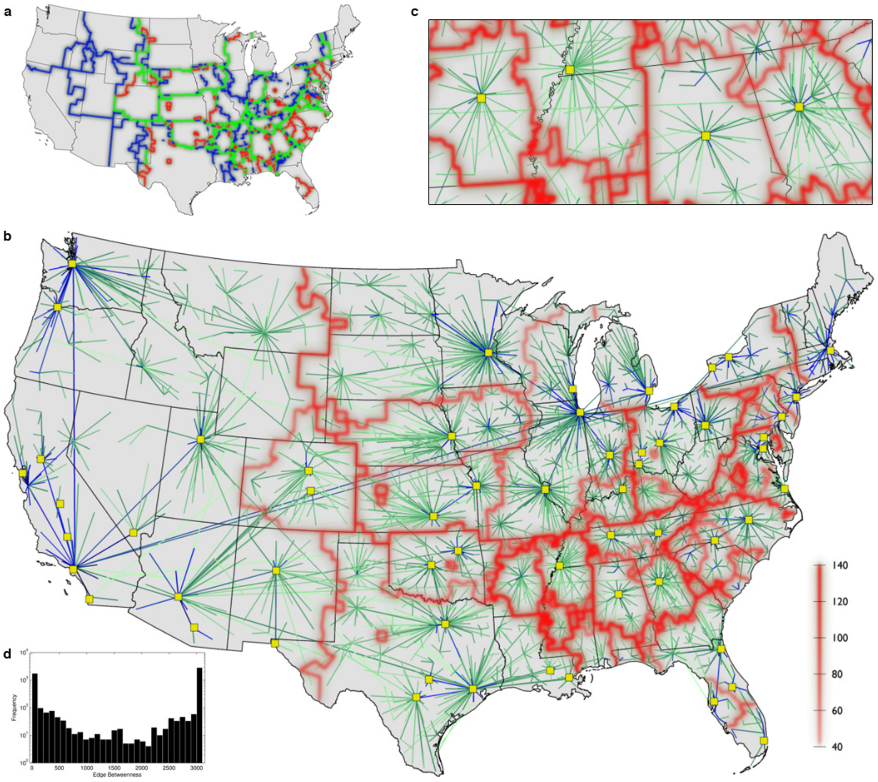


Figure 4: effective geographic borders implicitly encoded in multi-scale mobility networks (red). The discovery and computation of these structures heavily relied on the use of the mobility database server described in the Deliverable 2.1

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 have been published on Nature.

On the impact of **multi-scale mobility networks on spatial epidemic spread**, a paper published on the Proceedings of the National Academy of Sciences by the ISI team 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 modelling of infectious diseases and have already contributed to the improvement of the computational model used to provide estimates and projections of the H1N1 pandemic.

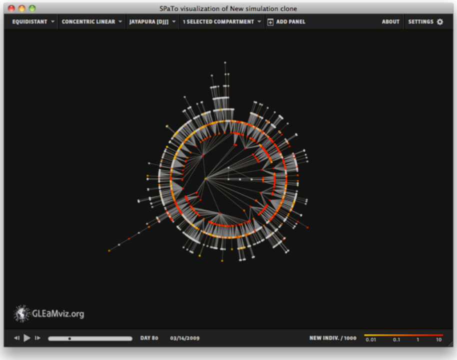


Fig.5 : The SPATO algorithm and visualization approach previously developed within WP2 by the MPG group has been integrated directly into the GLEaMviz client. It is now possible to visualize the results of a simulation by displaying the spreading of a global-scale epidemics upon a shortest-path representation of the airline transportation network, with different rescaling options.

The theoretical works has also provided the theoretical analysis and foundations for the SPATO (Shortest Path tree Tomography) algorithm for the visualization of dynamical phenomena in population networks. The algorithm extracts mobility networks from their geographical embedding and visualizes the network based on shortest paths and effective distances. The interactive tool has been extended and fully integrated in the modeling computational platform of **WP4** and allows the visualization of network characteristics, node specific parameters in an interactive way along with the progression of the epidemic.

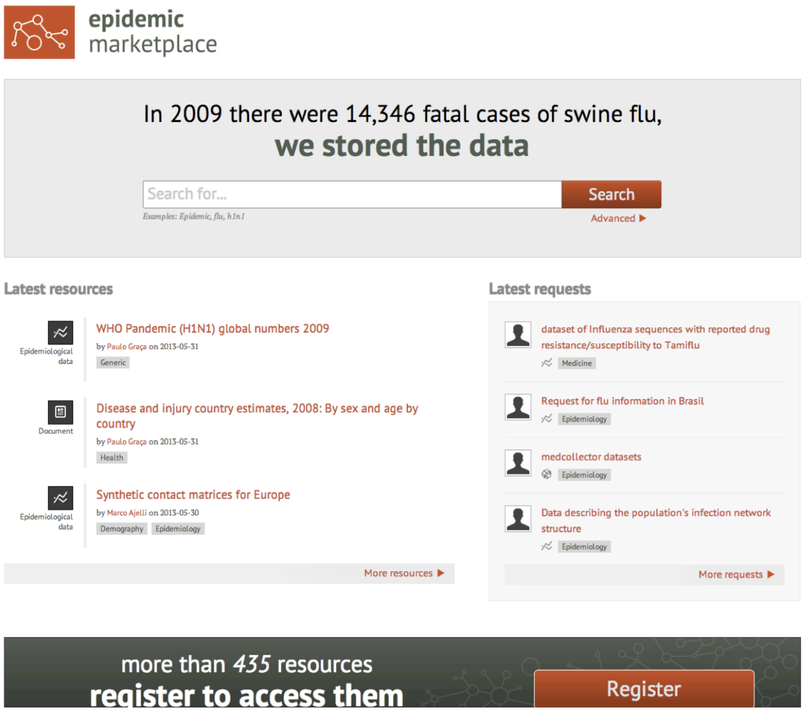
Finally progress has been made in the identification of immunization policies that takes into account mobility patterns and the development of tackled some general theoretical questions that concern the basic understanding of the impact of mobility multi-scale networks and their spatial embedding on the spatial spread of infectious diseases on the large scale. In particular detailed analysis of the effect of travel restrictions on the large-scale spreading of emerging infectious diseases has been carried out. Work from ISI and MPG has been able to map the problem of the global invasion of a pathogen in a collection of populations can be mapped in a critical phase transition of a reaction diffusion process, thus obtaining analytical insight on the possibility and effectiveness of containment at the source of an epidemic outbreak. The consortium has also explored in a real-world system the interplay between complexity features and predictability in computational epidemiology. In particular, the work has made extensive use of the computational modelling platform to explore how the variations in the complex characteristics of a system can produce very different scenarios when simulating a real-world epidemic spreading on a large scale. This very computationally intensive effort has advanced our foundational understanding of how the theoretical complex features of a system can deeply affect the output and thus the decision-making when analysing an epidemic spreading scenario. Moreover, the consortium developed new techniques that based on easily accessible socio-demographic data, developed synthetic contact matrices to be used in a computational approach aimed at revealing contact patterns driving the disease spread in a structured population. The synthetic contact matrices represent a very important data set that will be shared with the scientific community.

All groups published their work often in high rating journals in their field, and over the five years of the project, the groups published a relatively large quantity of original publications. From the perspective of the participating teams, it appears that the project has been very successful scientifically, and has also given the different European groups a basis for fostering future scientific cooperation.

- Data-driven computational and data sharing platform

In this area the project has delivered according to the plan, and beyond, the computational and data sharing platforms. Both of them are intercommunicating and publicly available.

The Epidemic Marketplace available at http://www.epimarketplace.net is now sporting a redesigned front-end with much improved usability. The team hired the services of a professional graphic design to create an Epiwork-specific style. The entire set of interactions on the website have been re-created using the knowledge acquired with the implementation of the previous versions of the EM. On the computational tool side the activity of the Epidemic Modelling Platform has progressed with several releases of the GLEaMviz epidemic simulation software, equipped with the visualizations of additional data layers, improvement to the simulation engine and enhancement to the data visualization settings. GLEAMviz is now at the version 4.2 with an imminent release of a v5.0. The FBK and ISI teams have finalized the alignment of agent based and structured meta-population models. The result is a hybrid simulation approach that combines the resolution of agent-based models with the computational efficiency of structured population models. This has been implemented through the development of open access APIs for the integration of the ABM model developed by the FBK team in the computational platform. The development of epidemic modelling platform is being performed in close contact with national and international agencies, who are testing and monitoring the development of the platform to provide the appropriate feedback aimed at improving its usability in the crisis management and policy making processes. The integration between the epidemic marketplace and the epidemic modelling platform has been improved and fine-tuned and has been presented at the Epiwork final workshop.



In 2009 there were 14,346 fatal cases of swine flu,
we stored the data

Examples: Epidemic, flu, h1n1

Latest resources

Epidemiological data — WHO Pandemic (H1N1) global numbers 2009
by Paulo Graça on 2013-05-31
Generic

Document — Disease and injury country estimates, 2008: By sex and age by country
by Paulo Graça on 2013-05-31
Health

Epidemiological data — Synthetic contact matrices for Europe
by Marco Ajelli on 2013-05-30
Demography Epidemiology

Latest requests

dataset of Influenza sequences with reported drug resistance/susceptibility to Tamiflu
Medicine

Request for flu information in Brasil
Epidemiology

medcollector datasets
Epidemiology

Data describing the population's infection network structure
Epidemiology

more than *435* resources
register to access them

Fig. 6: The landing page of the Epidemic Marketplace platform

The Epidemic Marketplace is an information platform to mediate access to distributed collections of public health data, offering an easy and safe way to share data for those data providers who want to collaborate with epidemiological modellers. Researchers will use this platform in multiple ways: i) as catalogue of data sources containing the metadata describing

existing databases; ii) as a forum to publish information about their own data, seeking modellers to collaborate with, and/or to seek sources of data that could be of interest to their epidemiological modelling efforts; and finally, iii) as the host of mediating software that can automatically process queries for epidemiological data available from the information sources connected to the platform. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that serves researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It supports the exchange of resources as well as user interactions.
The EM is fully built on open source technologies (see Fig.7). The back end is based on a Fedora Commons repository for storing and managing resources and a Lightweight Directory Access Protocol (LDAP) server for user management. The EM provides a set of web services that enable full access and manipulation of the repository content, and a Drupal-based user interface front-end for interactive upload and manipulation of resources. The front-end uses the same web services that are offered to external applications. The whole system has been running on a cluster of Linux machines.

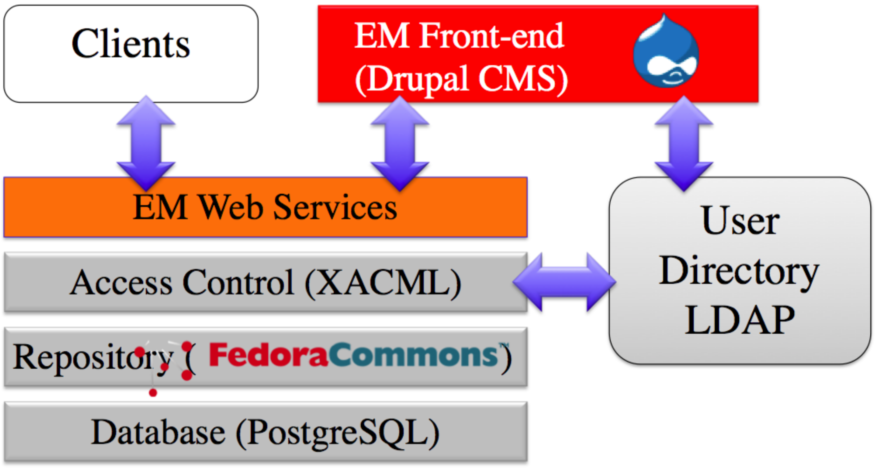


Fig.7: The Epidemic Marketplace architecture.

Based on a Web2.0 approach, users become active participants, sharing information and data, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. We implemented a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset comes with a metadata file, signalling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata includes information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and geographic data,. The Marketplace supports flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users are supported. During the four years of the project, we have developed and extensively validated the meta-model proposed in the paper "Epidemiological resource identifiers and their semantics, where are they?", submitted to PLOS One, on June 2013. We also created a new Ontology, "The Epidemic Ontology" (EO), which is designed to support the semantic annotation of epidemiology resources. The EO is integrated into NERO and complements epidemically relevant aspects that are not described in the existing ontologies. The EO currently contains 130 classes, 53 synonyms, 51 cross-references to seven external resources including MeSH, UMLS, NCI Thesaurus, Pathogen Transmission Ontology and the Infectious Disease

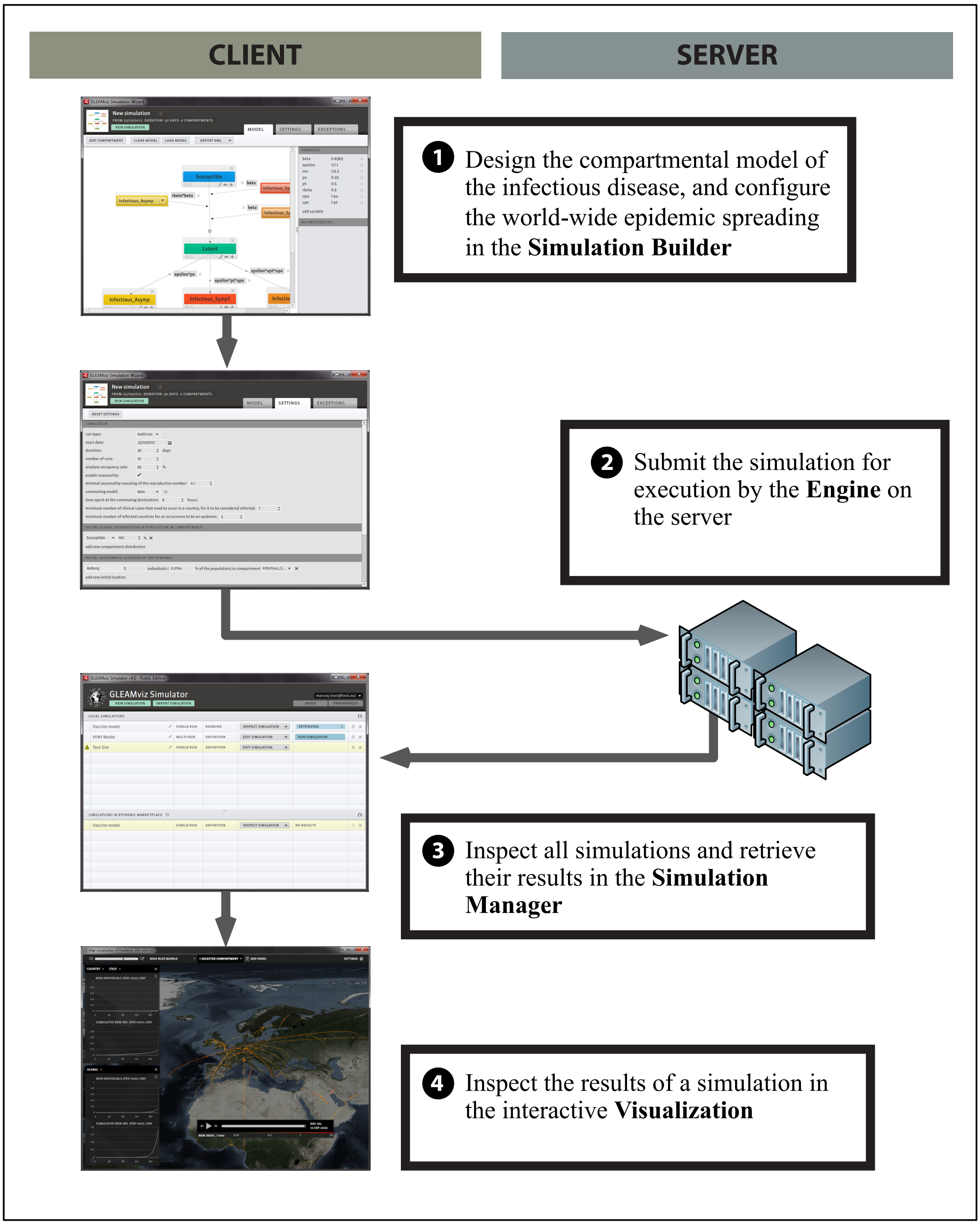
Ontology. It uses the Basic Formal Ontology as an upper ontology, but we expect to integrate it into the mid-level Medical Surveillance ontology. See https://code.google.com/p/epidemiology-ontology/. Finally during the last year of the project

we have redesigned front-end with much improved usability. The team hired the services of a professional graphic design to create an Epiwork-specific style. The entire set of interactions on the website have been re-created using the knowledge acquired with the implementation of the previous versions of the EM. The final version the Epidemic Marketplace was presented in the Epiwork workshop in Digital Epidemiology in Torino (may 2013).

The platform for the computational modeling of infectious disease spread has been designed and implemented. The result is a publicly available software system, called “GLEAMviz” (http://www.gleamviz.org) that simulates the spread of emerging human-to-human infectious diseases across the world. This last release of this software software, as described in detail in the Deliverable 4.6, is a user-friendly tool for the simulation of a case study, test and validation of specific assumption on the spread of a disease, understanding of observed epidemic patterns, study of the effectiveness and results of different intervention strategies, analysis of risk through model scenarios, forecast of newly emerging infectious diseases. GLEAMviz comprises three components:

- the client application
- the proxy middleware
- the simulation engine.

The simulation engine is based on the Global Epidemic and Mobility (GLEaM) framework, a stochastic computational scheme that integrates worldwide high-resolution demographic and mobility data to simulate the spread of diseases on the global scale. The GLEAMviz design aims at maximizing flexibility in defining the disease compartmental model and configuring the simulation scenario; it allows the user to set a variety of parameters including: compartment-specific features, transition values, and environmental effects. The output is a dynamic map and a corresponding set of charts that quantitatively describe the geo-temporal evolution of the disease. The software is designed as a client-server system. The multi-platform client, which can be installed on the user's local machine, is used to set up simulations that will be executed on the server, thus avoiding specific requirements for large computational capabilities on the user side. The communication between server and client is implemented using a standard "XML over HTTP" protocol, following the REST paradigm. The GLEAMviz usage workflow, from the definition of the model simulation to the visualization of the simulation output, is depicted in Fig.8.



① Design the compartmental model of the infectious disease, and configure the world-wide epidemic spreading in the **Simulation Builder**

② Submit the simulation for execution by the **Engine** on the server

③ Inspect all simulations and retrieve their results in the **Simulation Manager**

④ Inspect the results of a simulation in the interactive **Visualization**

Fig.8: GLEAMviz usage workflow.

The **GLEAMviz server** is composed by the proxy middleware and the simulation engine. Users interact with the GLEAMviz system by means of the client application, which provides graphical user-interfaces for designing and managing the simulations, as well as visualizing the results. The clients, however, do not themselves run the simulations. Instead they establish a connection with the GLEAMviz proxy middleware to request the execution of a simulation by the server.

The **GLEAMviz simulation engine** uses a stochastic meta-population approach (already described in D4.1 – Simulation Framework) that considers data-driven schemes for the short-range and long-range mobility of individuals at the inter-population level, coupled with coarse- grained techniques to describe the infection dynamics within each subpopulation. The basic mechanism for epidemic propagation occurs at multiple scales. Individuals interact within each subpopulation and may contract the disease if an outbreak is taking place in that subpopulation. By travelling while infected, individuals can carry the pathogen to a non-infected region of the world, thus starting a new outbreak and shaping the spatial spread of the disease. The simulation engine consists of a core that executes the simulations and a wrapper that prepares the execution based on the configuration relayed from the client by the GLEAMviz proxy middleware. The engine can perform either single-run or multi-run simulations (each with the same configuration but with a different random seed, whose results are then aggregated and statistically analyzed by the wrapper code). The simulation engine writes the results to files and uses lock files to signal its status to the middleware component. The engine component is coded as a Python program with additional modules; all the computational expensive tasks are then performed by embedded FORTRAN routines directly callable by the main program.

The **GLEAMviz proxy middleware** is the server component that mediates between clients and simulation engines, handling the user management and the simulation requests, and controlling the engine's execution. The client users are authenticated on an LDAP directory, which is kept synchronized with the Epidemic Marketplace data platform. The middleware component implementation is based on the Django Web framework (https://www.djangoproject.com/) and the PostgreSQL database (http://www.postgresql.org/) and is designed in a way that makes the GLEAMviz server potentially suitable for interfacing with various client types. **The current implementation of the server component of the GLEAMviz computational platform allows for a flexible and scalable infrastructure.** The optimized engine code implements the GLEAM model with **simulations running approximately ten times faster than with the previous version.**

The software system presented above is operated through the GLEAMviz client, which provides the user interface: the part of the tool actually experienced on the user side. The GLEAMviz client integrates different modules that allow the management of the entire process flow from the definition of the model to the visualization of the results. The **GLEAMviz client** is a desktop application by which users interact with the GLEAMviz tool. It provides GUIs for its main functions: 1) the full definition of a simulation including the design of a compartmental models that defines the infection dynamics and the configuration of the simulation parameters; 2) the submission and retrieval of the simulation to the server for execution; 3) the visualization of the simulation results with different options; and 4) the management of the user’s collection of simulations.

The client can be installed on recent versions of the Microsoft Windows and Mac OS X operating systems; a client version not including the 3D globe visualization (see below) is available for the Linux operating system. The Simulation Builder GUI is used to design a simulation, from the compartmental model to the global settings and optional time-dependent exceptions. Those three components are organized by means of tabbed panels. The Model Builder provides a visual modeling tool for designing arbitrary compartmental models, ranging from simple *SIR* models to complex compartmentalization in which multiple interventions can be considered along with disease-associated complications and other effects. This allows the user to model many kinds of human-to-human infectious diseases, in particular respiratory and influenza-like diseases. Models can be exported to XML files and stored locally, allowing the user to load a model later, modify it, and share it with other users. The diagram representation can be exported as a PDF or SVG file for use in documentation or publications. A few examples of compartmental models are available for download from the Simulator website. The Settings panel is used to configure the simulation’s general parameters, which define the basic scenario for the simulation that will be run by the GLEAMviz server. Users have now multiple choices for **implementing local human mobility** (data based models, gravity models, radiation models, etc.). The Exceptions panel, with a new **redesigned interface,** allows the setting of **disease models and interventions that can be geographically dependent** (and for specific time period).

Once the execution of a simulation is finished and the results have been retrieved from the server, the client can display the results in the form of an interactive visualization of the geo-temporal evolution of the epidemic. The geographic mapping involves a zoomable multi-scale map on which the cells of the population layer are colored according to the number of new cases of the quantity that is being displayed. The color-coding of the map represents the number of cases on a particular day. The time evolution of the epidemic can be shown as a movie, or in the form of daily states by moving forward or backward by one day at a time. For single-run simulations it is also possible to show the airline transportation of the 'seeding' individuals by drawing the traveling edge between the origin and destination cities. Beside the geographical map, the Map Window displays a set of charts. For each pair of charts the top one shows the number of new individuals per 1,000 over time (incidence), and the other shows the cumulative number of new individuals per 1,000 over time (size). For multi-run simulations, median values and corresponding 95% confidence intervals are shown. The user can choose the context for which the corresponding charts show incidence and size data.

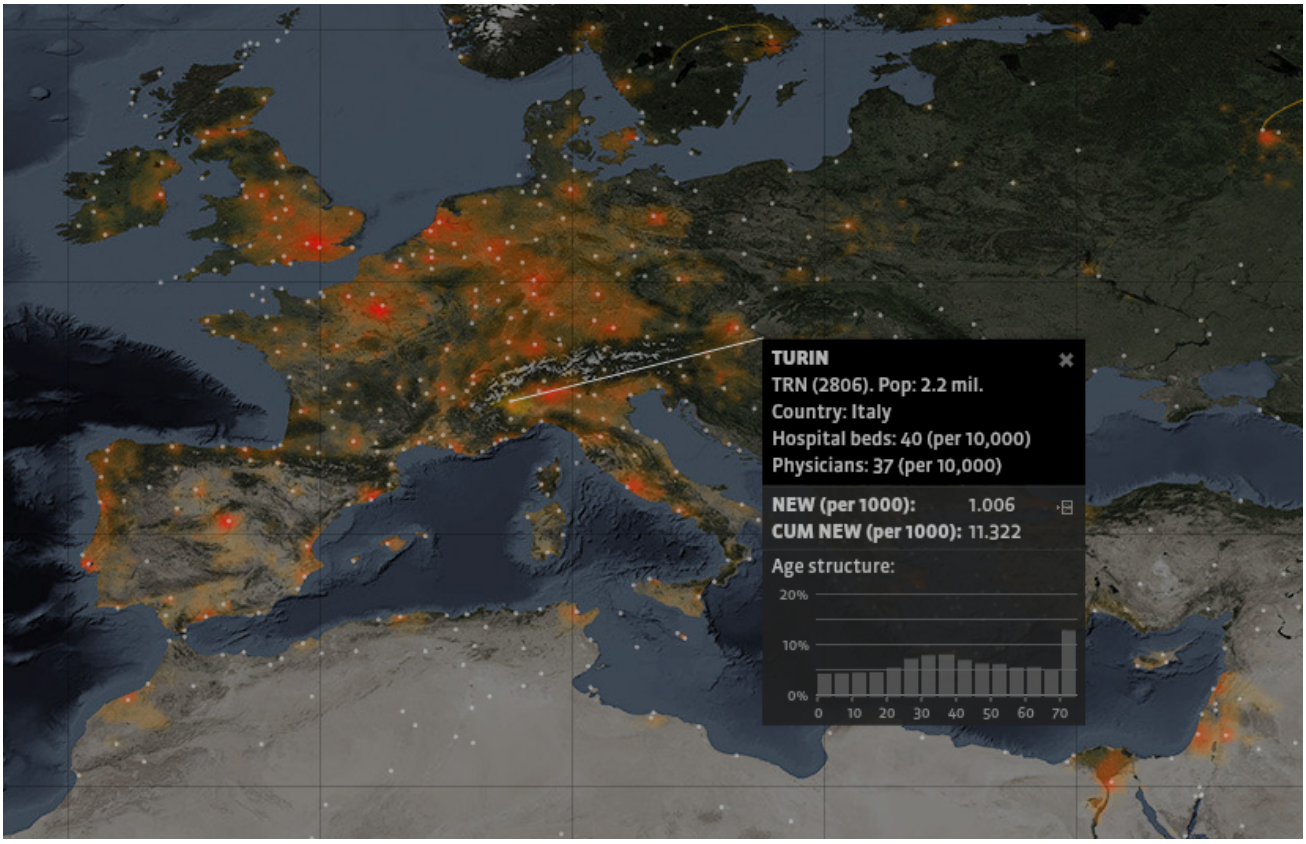

TURIN
TRN (2806). Pop: 2.2 mil.
Country: Italy
Hospital beds: 40 (per 10,000)
Physicians: 37 (per 10,000)
NEW (per 1000): 1.006
CUM NEW (per 1000): 11.322
Age structure:

Fig.9: GUI for data exploration on the GLEAMviz client.

A major result of the project consists in the full integration of the Epidemic Marketplace and the Computational platform GLEaMviz (see Fig.10). The users of the two platforms will be authenticated across a synchronised infrastructure enabling them users to transparently access both systems with the same credentials. This also enables makes the interaction of the

GLEaMviz client with the Epidemic Marketplace repository very smooth, avoiding additional user-credential requests upon interaction and therefore allowing seamless single sign-on access to both platforms. Both platforms will use OpenLDAP software to manage users and credentials.

It is important to stress that the computational platform has already secured funding from private foundations and other international agencies to continue development and support. The Epidemic Marketplace is going to be supported by the Lisbon’s team. In summary all the components of the projects will continue to operate after the end of Epiwork and will be used as the starting point for new projects and tool development.

- ICT monitoring and reporting system.

The project aimed at overcoming the limitation of the state of the art surveillance systems by proposing an innovative ICT approach based on Web2.0 tools. During the whole duration of the project, the consortium had the goal to extend the IMS systems to five new countries as well as manage well operating systems in the four “old” countries. Presently, UK, Sweden, France, Spain, Ireland and Denmark have been added to the Influenzanet network of web-based surveillance platforms (D5.8). The consortium final goal, fully completed, was to develop an easy-to-deploy, fully localizable Internet monitoring system (IMS) platform and a centralized database collecting data from the several participating countries in a harmonized and centrally coordinated fashion. The IMS mobile application for iPhone and Android, to allow the IMS participation of volunteers through smartphones has also been developed (D5.7).

Fig.10: Full integration among the computational modeling and the data repository platforms.

In all the countries taking part in the Influenzanet network, participants are volunteers recruited in the general population who register to the project providing general background information and who are then invited to self-report on their ILI symptoms on a weekly basis. Participation is voluntary and anonymous. Registration takes place on the national websites where participants can create a user account. A valid email address is required to confirm the registration and avoid unsupervised registration attempts. The email address is also used to maintain a communication channel with each participant as a motivation purpose during the influenza season. Each national website contains several pages that are publicly accessible dedicated to the presentation and aims of the project, instructions on how to participate, general information about influenza virus and influenza epidemics and pandemics, and the results of the study provided in real time. Upon registration, the user are asked to complete an intake survey covering demographic factors (age, gender), geographical factors (location of home and work/school expressed at the municipality level), socio-economic factors (household size and composition, occupation, educational level, number of daily contacts with groups of patients, children or elderly, daily transportation means), and health-related factors (height and weight, diet, vaccination status, pregnancy status, smoking habits, major risk conditions, and others). At-risk population is identified as the population eligible for a vaccination voucher. The full list of variables for which data was collected in the intake survey is reported in the "Gold Standard" definition in Deliverable D5.1. Participants are also allowed to register and create accounts for other individuals on their behalf, given consent, thus enabling, for instance, to grant access to the project to individuals less comfortable with Internet tools (e.g. in the older age classes). The intake data are collected only once at the enrollment, however the participant can access her/his intake survey anytime and modify it following any update (e.g. change of residence, of vaccination or pregnancy status). In presence of multiple instances of intake surveys, we considered the last completed. Each week during the influenza season, participants are asked to fill in a weekly symptoms survey, independently of possible ILI symptoms, again reported in the "Gold Standard" definition in Deliverable D5.1. An automatic email message is sent weekly in the form of a newsletter that contains a reminder to fill in the symptoms survey as well. The newsletter is also meant to keep a high motivation in the participation to the study by providing preliminary results on the analysis of the symptoms reported the week before, a general description of the influenza epidemic situation in the participants' country and in Europe,

During the life of the project the IMS has been deployed in ten European countries. Here is the list of teams and countries involved so far in the Influenzanet acvities. After the Netherlands, Belgium, Portugal and Italy, the IMS was deployed in the UK by LSHTM , collecting data on influenza since July 2009. Sweden was the next country, where the IMS was launched in February 2011 by SMI. France, with INSERM, has successfully launched the platform in December 2011 reaching more than 4000 participants in two months. In September 2012, the IMS has been deployed in Spain by a team at the University of Zaragoza. In May 2013, the IMS has been deployed in Ireland and Denmark, where the data collection will begin with the influenza season 2013-14. In Summary, the IMS is now active in the following countries: the Netherlands, Belgium, Portugal, Italy, United Kingdom, France, Spain, Ireland, Denmark. Influenzanet is now the first web-based participatory surveillance network covering almost the whole western Europe. We have plans for establishing a collaboration and data exchange agreement with Prof. Udo Bucholz, in charge of Grippeweb, the web-based monitoring system for influenza active in Germany.

The Consortium is also active in maintaining the Influenzanet website: www.influenzanet.eu/. This is the Influenzanet corporate website for health care professionals, researchers, policy makers and everybody else, interested in epidemiology, flu surveillance and modeling. Influenzanet.eu presents the project an its results in a reader-friendly way, in order to promote the concept of Internet-based Monitoring Systems in other countries, to expand scientific cooperation with colleagues all over the world and, last but not least, to show interested visitors what flu, vaccination and epidemiology is. It contains an analysis of the collected Influenzanet data so far is presented, with as main and public elements:

- Graphs and data on flu and cold from all current (10) Influenzanet partners;
- the Netherlands, Belgium, Portugal, Italy and the UK;

- Daily updates from all graphs and data;
- ILI curves, as compared to Google Flu, EISN, Temperature (also interactive);
- Curves of other syndromes and ILI curves within various subgroups;
- Participation data: participants, completed surveys, histograms;
- Week and incidence data, also as CSV files;
- All published articles on IMS and Influenzanet to date;
- A weekly update of modeling and flu surveillance news.

The (weekly) editing is and operational maintenance is part of AIBV's responsibility. The scientific part and cooperation within and outside Epiwork, is the responsibility of the Influenzanet Science Committee, established at our W5 meeting in Amsterdam, May 2011.
This Committee answers request for research cooperation, including data admission requests. Until the end of Epiwork in February 2013, the ISC consisted of five members including Sander van Noort, representing GGM (NL/B) and Gripenet, from the Gulbenkian Institute in Portugal, Dr. Daniela Paolotti, on behalf of Influweb and ISI as Epiwork coordinator, Dr Vittoria Colizza from INSERM, France and Dr Sandro Meloni from Spain, AnnaSara Carnahan from Influensakoll/SMI, Sweden and Professors John Edmunds from Flusurvey/LSHTM and Marc van Ranst from KU Leuven. Secretary is AI BV's director Ronald Smallenburg. On the longer run, it is our intention that the ISC includes one representative per partner. This implies that the Committee will need procedures for decision making on research cooperation and data admission requests.
The central EpiDB is now fully functional. It has been extensively described in the previous report and in the third period Deliverable 5.5. Real-time from the Influenzanet countries during the whole influenza seasons of 2011-12 and 2012-13 have been loaded into the EpiDB finally making possible the building of a European map (Fig.11) created by the shared database.

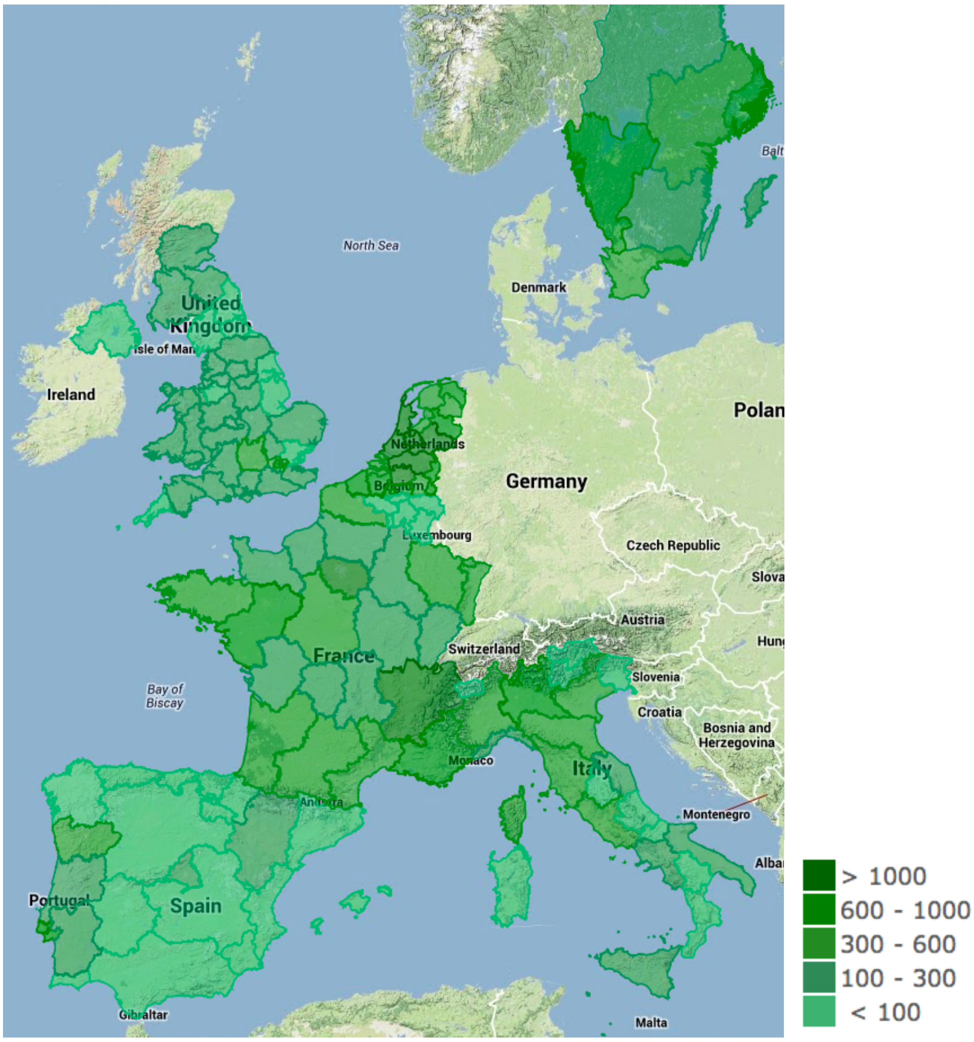


Fig.11: Number of participants in the NUTS2 regions of the Influenzanet Countries

The Influenzanet Consortium is also leading a worldwide effort consortiating all the platforms carrying out web-based influenza surveillance, i.e. the European Influenzanet (www.influenzanet.eu), Flu Near You (www.flunearyou.org) in the United States and Flu Tracking (www.flutracking.net) in Australia. During the 2nd International Workshop on Participatory Surveillance (IWOPS II) held in Amsterdam in April 2013 organized by the Skoll Global Threats Fund in collaboration with AIBV, aimed at mapping infectious diseases in both humans and animals, a Letter of Intent on cooperation and data exchange has been concluded between the three systems with the aim to achieve a world-wide disease radar: https://www.influenzanet.eu/en/news/2013/04/29/international-workshop-on-self-reported.
We also plan to extend the IMS surveillance of other than influenza contagious diseases. Questionnaires for monitoring dengue have been developed in collaboration with a team of epidemiologists in Brazil. In collaboration with our Brazilian colleagues, the dengue monitoring system (Dengue na web) started to operate in the city of Salvador, Bahia, by October last year. This project got a lot of attention in the community of Portuguese speaking countries by a programme from the Portuguese National Radio in 2011: http://www.rtp.pt/icmblogs/rtp/cientificamente/?k=Dengue-na-WEB.rtp&post=36828

Along the lines of creating a world-wide disease radar, in May 2013 we organized the International Workshop "Public Health in the Digital Age", in joint organization with the World Wide Web Conference 2013, Rio de Janeiro (http://www.digitalhealth.ws/). The aim of this interdisciplinary workshop was to bring together public health professionals and computer science researchers in Big Data mining, crowdsourcing and SM user engagement to raise awareness on some critical global applications: public health, user participation and risk communication.

At the end of May 2013 we organized the International Workshop on digital epidemiology (http://www.isi.it/workshop_digital_epidemiology) aimed at exploring how Web-based systems, on line information streams and mobile devices have revolutionized the way we monitor large scale communities on the quest for real time public health signals. The Workshop has seen the participation of experts in on line data mining, crowdsourcing and social media together with public health professionals working in epidemic intelligent services of ECDC, NIH, WHO.

Partner KU Leuven invested part of its EPIWORK budget in preparing virological and epidemiological validation of the method by distributing SWAB kits to volunteers with flu symptoms as registered by the Dutch-Belgian IMS, de Grote Griepmeting. In the 10th GGM season, a maximum of 150 volunteers were selected on and invited for assisted (with an MD) self-sampling by swab kits from KU Leuven and diagnosis with help of the Agreed Influenzanet Questionnaire. KU Leuven collected and analysed all of these SWABs in their own labs. A PhD student is coordinating this additional and innovative part of Influenzanet for which KU Leuven was granted permission by the EC program officer for Epiwork

Partner SMI had the task to validate the IMS data collection scheme. Their work was meant to provide a comparative analysis of three systems in Sweden: (1) the existing GP sentinel system of influenza surveillance with the (2) new internet monitoring system (IMS) and (3) a population based approach, currently being evaluated on a large-scale pilot study basis. In Sweden, these three influenza surveillance systems are the key elements of a comparative analysis with regard to IMS. The H1N1 pandemic in 2009 accelerated the set up of the population based approach and of the infrastructure that collected data about ILI cases among a random representative sample of the Swedish population. In parallel, they set up the IMS in Sweden at the end of the project second year, thus they had more than two years and a half to draw a conclusion about the comparison between the different systems and these conclusions are extensively described in Deliverable 6.5 These conclusions can be summarized as follow. Despite the self-recruited sample, resulting in poor reflection of the Swedish populations' demography, Influensakoll (the Swedish IMS) offered a reasonable representation of the temporal ILI pattern in the community during the 2011-2012 and 2012-2013 influenza seasons and could be a suitable tool for following ILI trends in the community. However, comparison of self-recruited participants and randomly sampled invitees shows that invited participants represent the target population better and complete a larger number of reports than the self-recruited participants. These findings suggest that a personal invitation to a random sample of the population improves the quality and usability of community-based surveillance data.

- The challenge of the H1N1 2009 pandemic

After a few months in the life of the project the consortium has been involved in the effort of the international scientific community to fight the 2009 H1N1 pandemic. 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 and exceptional strain and workload for the teams 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. The project achieved several major achievements related to the H1N1 pandemic. The Tel Aviv University group has obtained relevant results concerning the contact network structure of the Israeli population and the manner it affected the spread of the swine flu pandemic in 2009. Three teams of the consortium have been using computational models to produce projections of the spread of the ongoing H1N1 Flu epidemic. The work 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 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. 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. *It is work remarking that all the activities concerning H1N1 carried out in the consortium do not represent a deviation from the planned work. All Activities have contributed to specific tasks of the project and produced advances (in most cases ahead of schedule) of the planned work.*

In Epiwork 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, R0, 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).

The **consortium developed high 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 km. 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 km 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 and 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.

The project has also used 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.

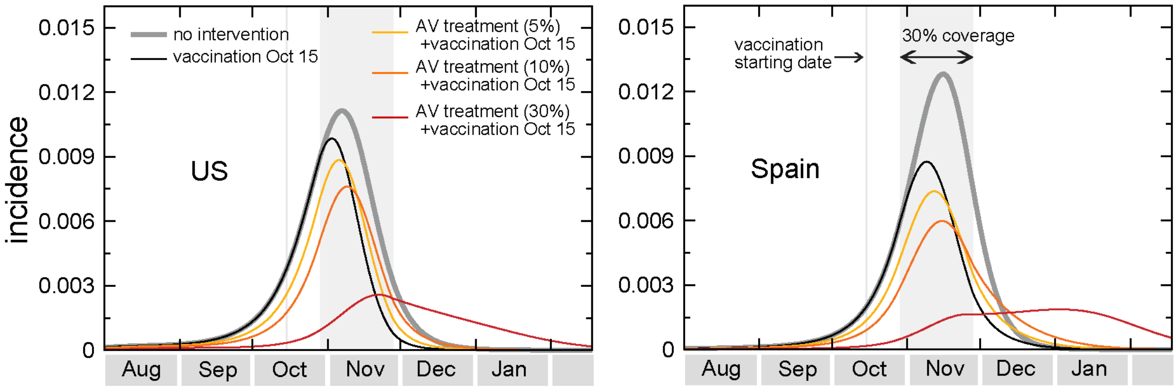


Figure 12: Incidence curves for US and Spain for different intervention scenarios. The gray bar indicates the time period during which the immunization takes effect.

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 R0 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 2009 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. The results from the computational model have been finally validated and published in 2012 in BMC Medicine as part of D4.5. In fig. 13 we report a comparison of the computational and field data about the epidemic peaks in the Northern hemisphere.

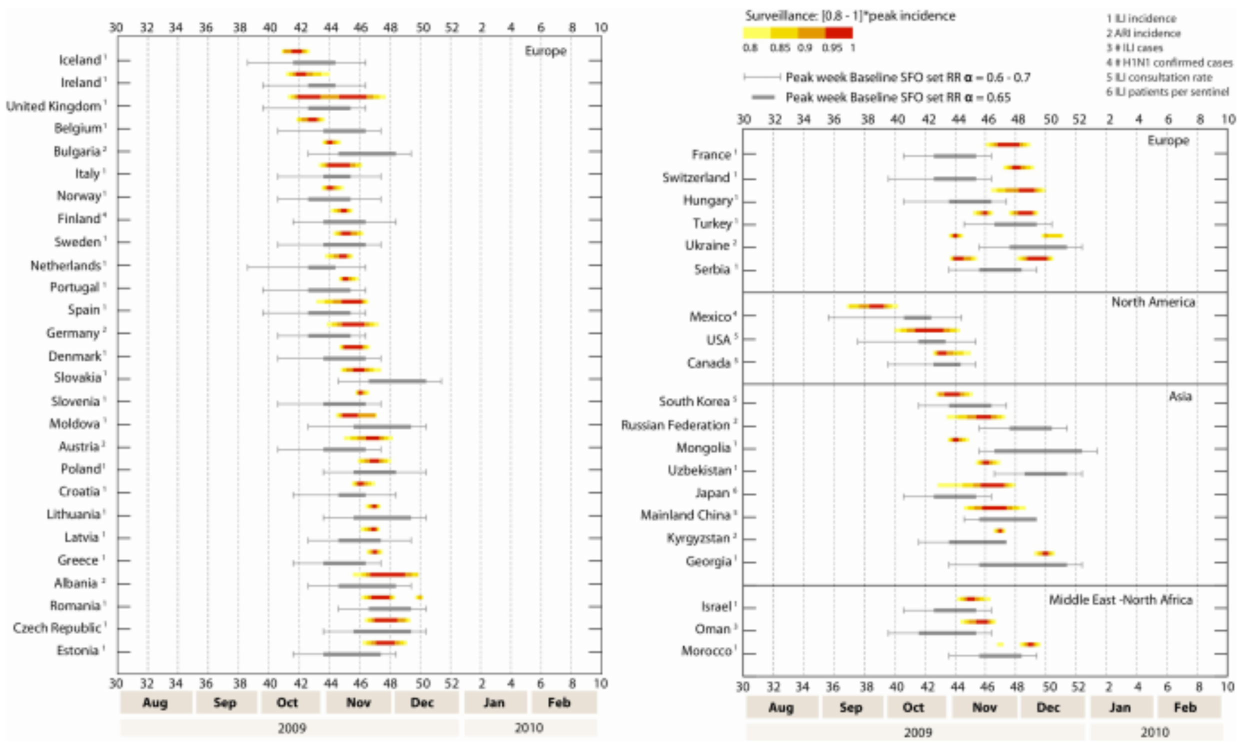


Figure 5 Peak timing in the northern hemisphere: simulations and real data. Peak weeks of the epidemic activity in the baseline stochastic forecast output (SFO) (gray). The reference ranges of the simulated peak week were obtained by analysis of 2,000 stochastic realizations of the model for three different values of the seasonal rescaling factor, α_{min}, of 0.6, 0.65, and 0.7. The peak weeks reported by the surveillance for the fall/winter wave are shown as color gradients, whose limits correspond to the time interval at which an incidence of greater than 80% of the maximum incidence was observed. The numbers 1 to 5 indicate the type of data provided by the surveillance of each country, and the numbered weeks of the year correspond to the calendar used by the US Center for Disease Control and Prevention.

The arrive of the new influenza H1N1 in Europe encouraged all the researchers of the consortium 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 2009, the second draft has been tuned in August 2009, before the striking of the Autumn wave of H1N1 pandemic. In December 2009 a draft for the design of the European centralized database has been prepared, again brought forward by the occurrence of the H1N1 pandemic. The database would 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. 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) 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.

- Exemplary cases of contributions to/from complexity science

The second year review panel has explicitly remarked the need *to spell out the contribution to/from complexity science of the projec*t. Here we provide a few exemplary cases of contributions toward the objective of the call at the origin of the Epiwork funding. All the following cases reports results obtained because of the exploitation of a systemic approach to epidemic system, computational thinking the use of techniques and modeling in complexity science.

- ***Travel restrictions, invasion threshold and critical phase transitions.***

The human mobility flows that determine the spreading of infectious diseases and the control measures based on limiting or constraining human mobility are considered in the contingency planning of several countries. The target of these control measures is the decrease of travel to/from the areas affected by the epidemic outbreak and the corresponding decline of infected individuals reaching countries not yet affected by the epidemic. Most numerical methods and empirical evidence show that even sever regimes of travel reduction would lead to delays on the order of a few weeks even in the optimistic case of early intervention. It is unlikely that given the ever-increasing mobility of people travel restrictions could be used effectively in a future pandemic event.

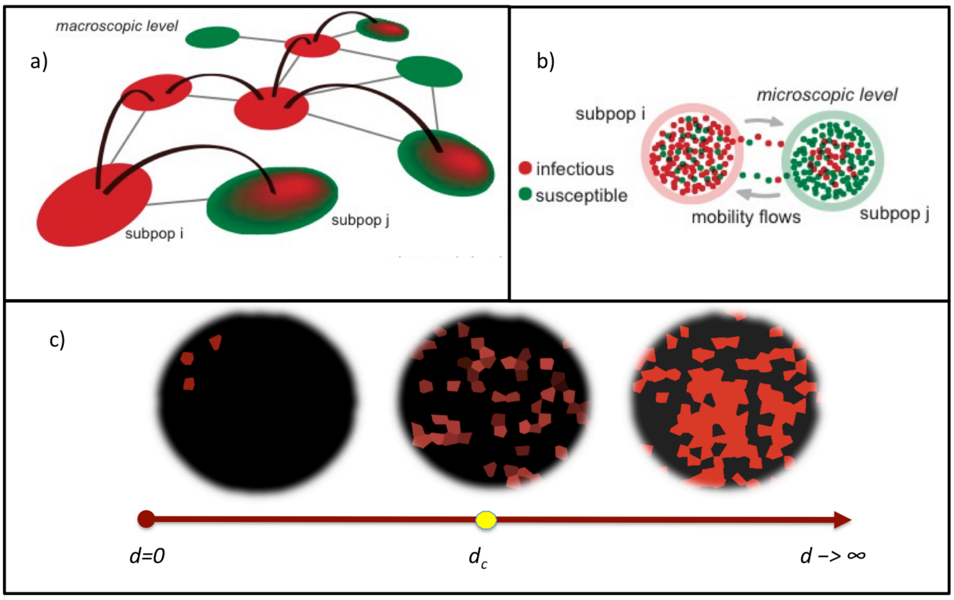


Fig.14: a) At the macroscopic level the system is composed of a heterogeneous network of subpopulations. The contagion process in one subpopulation (marked in red) can spread to other subpopulations because of particles diffusing across subpopulations. b) At the microscopic level, each subpopulation contains a population of individuals. Within each subpopulation, individuals can mix homogeneously or according to a subnetwork and can diffuse with rate d from one subpopulation to another following the edges of the network. c) A critical value d_c of the individuals/particles diffusion identifies a phase transition between a regime in which the contagion affects a large fraction of the system and one in which only a small fraction is affected (see the discussion in the text).

This non-linear and somehow paradoxical feature of global epidemic spreading, on one side depending on people mobility and on the other side insensitive to its magnitude, can be understood mathematically a particle-network framework that extends the heterogeneous mean-field approach to reaction diffusion systems in networks with arbitrary degree distribution. Let us consider for instance a simple epidemic process such as the SIR model in a meta-population context. In this case each node of the network is a sub-population or urban area connected by a transportation system (the edges of the network) that allows individual to move from one subpopulation to the other (see Fig.14). If we assume that the single population reproductive number R_0>1 of the SIR model and a diffusion rate p on the network for the individuals we can easily identify two different limits. If p=0 any epidemic occurring in a given subpopulation will remain confined; no individual could travel to a different subpopulation and spread the infection across the system. In the limit p very large we have that individuals are constantly wandering from one subpopulation to the other and the system is in practice equivalent to a well-mixed single population. Since the epidemic has R_0>1 it will spread across the entire system. A transition point between these two regimes is therefore occurring at a threshold value p_c of the diffusion rate. The threshold is defined as the diffusion rate that allows a sufficient number of infected individuals from each infected subpopulation to visit other subpopulations and trigger a new epidemic. In this perspective we can consider the subpopulation network in a coarse-grained view and provide a characterization of the invasion dynamics at the level of subpopulations, defining a subpopulation reproductive number R^* that identifies the number of subpopulation epidemics triggered by each infected subpopulation. As for the basic reproductive number, only if R^*> 1 we have that the epidemic will spread globally. The number R^* is depending on the diffusion rate p and therefore define a global invasion threshold p_c for the diffusion rate. Furthermore the global invasion threshold is affected by the topological fluctuations of the metapopulation network. In particular, the larger the network heterogeneity, the smaller the value of the diffusion rate above which the epidemic may globally invade the metapopulation system. The relevance of this result stems from the fact that the complexity and heterogeneity of the present time human mobility network is responsible for the fact that travel restrictions, reducing the rate at which individuals leave an infected region, would not be able to considerably slow down the global spread unless 90% (or more) effective. This result has been originally derived in the context of the project and extended to recurrent mobility patterns by the MPI and ISI. In addition and interestingly this threshold cannot be uncovered by continuous models as it is related to the stochastic diffusion rate of single individuals. It defines a new critical phase transition class that can be found in several multi-particle systems in complex networks. On one side this result is providing answers to a specific problem in epidemic modeling and scenario analysis by using a complex systems approach. On the other hand the result can be applied and generalized to other systems and opens a new window on the effect of stochastic fluctuations on phase transitions and emerging tipping points. This result indeed defines a general class of models generally applicable to complex techno-social systems. While the presented result is anchored upon the example of disease spread, the meta-population approach can be abstracted to the phenomena of knowledge diffusion, online community formation, information spread, and technology.

- ***Systemic network approach to epidemic diffusion dynamics***

In complex networks embedded in the geographical space, diffusion processes often exhibit a disordered pattern that does not resemble the classic wave-like behavior we are used to observe in regular systems. In general, the relation of geographic distance and arrival time correlate well only in the regular lattice while the presence of long range interactions destroys these correlations (see Fig.15). The algorithm SPATO, developed within the project, reduces the network to the weighted shortest-path tree of a selected root node, obtaining a local but simpler view of the network that can be easily visualized. This representation provides immediately the right relational distance among nodes so that the diffusion dynamics can be rationalized and visualized according to our physical intuition of a wave like process (see Fig.15).

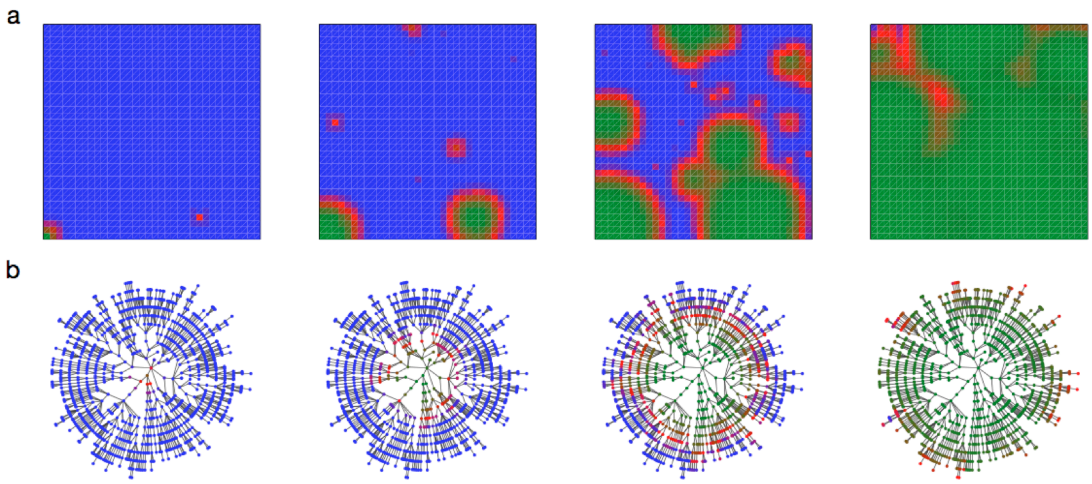


Fig.15: radial distance is shortest path distance in a lattice with long range connections. The SPATO visualization permits the computation of effective speeds, based on effective distance, which can be used as a predictor.

This results has been integrated in the Computational modeling platform offering a novel perspective on the large scale spreading of infectious diseases worldwide. However the results is much more general and applies to all diffusive and contagion processes in complex networks. Also in this case we have provided the development of methodology and algorithmic tools to understand the impact of large-scale complex features (scale invariance, extreme heterogeneity, unbounded fluctuations) of interaction, mobility and communication networks in the behavior of spatially extended models.

- ***Computational thinking, non-linearity and population heterogeneity***

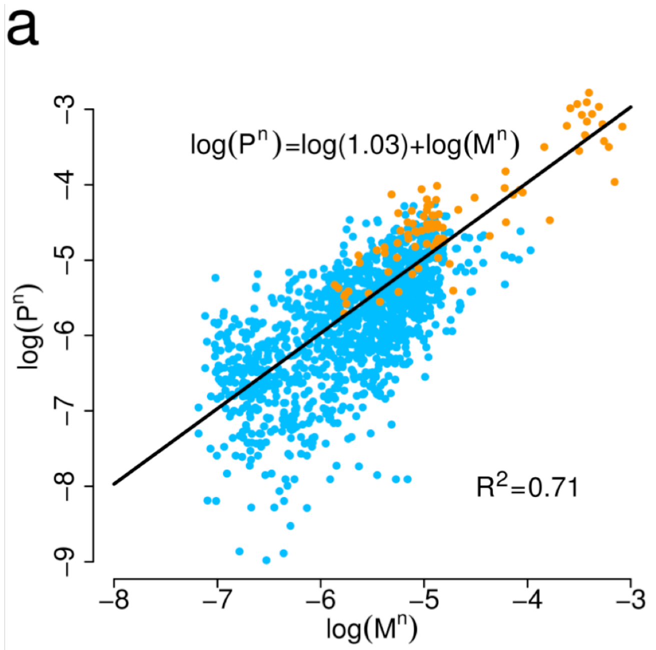


Fig.16: Linear regression model with zero intercept for POLYMOD matrices P^n against those from the synthetic model, M^n.

The accurate characterization of the structure of social contacts in mathematical and computational models of infectious disease transmission is a key element in the assessment of the impact of epidemic outbreaks, and the accuracy in the evaluation of the effective control measures. For instance, the transmissibility potential of a disease and the final epidemic size strongly depend on the mixing patterns between individuals of the population, which in turn depend on socio-demographic parameters (e.g. household size, fraction of workers and students in the population).}. For this reason, several efforts have been recently carried out in order to obtain contact data with the aim of quantifying ``who meets whom (where, when, how long and how often)'. Although a vast majority of studies assumes the population as homogeneous –all individuals are equal with same average contact rate– the social and demographic structure of the population is generally reflected in heterogeneous contact patterns among individuals. Age is obviously one of the main determinants of the mixing pattern of individuals: children tend to spend more time with children and members of their household, active adults mix with individuals in their workplace etc. Mixing patterns by age are generally defined by a contact matrix whose elements M_{ij} represent the average frequency of “adequate” contacts that an individual of age i has with individuals aged j.

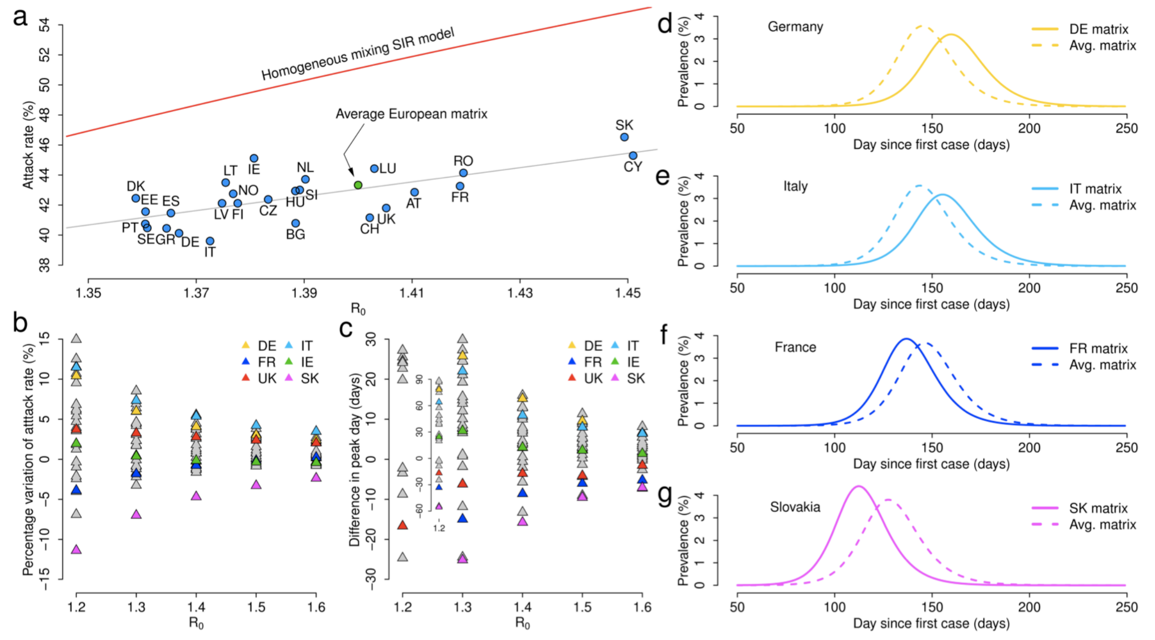


Fig 17 : a Final infection attack rate as a function of the basic reproduction number R_0 in the different countries (blue dots) by adopting country-specific matrices and by assuming the same probability of transmission q in all countries – specifically, the value resulting in $R_0 = 1.4$ by adopting the average European matrix (green dot). The attack rate corresponding to the average European matrix is computed by assuming the average European age structure in the model. Red line represents the attack rate of the homogeneous mixing SIR model for values of R_0 in the range of variability of the basic reproduction number of country-specific matrices. Grey line represents the best fit of the linear model to data points related to the use of country-specific matrices. **b**Percentage variation of infection attack rate for increasing values of R_0 of models based on country-specific matrices with respect to models based on the average European matrix (with country-specific age structure). **c** As **b** but for the variation of the peak day. **d–g** Daily prevalence over time of models with $R_0 = 1.4$ based on either the country-specific matrix (solid lines) or the average European matrix (dashed lines, with country-specific age structure) in Germany, Italy, France and Slovakia respectively.

The project has developed a general computational approach to derive mixing patterns from routinely collected socio-demographic data. In particular the project has generated the contact matrices by age of 26 European countries for which we are in the position to construct a synthetic society in the computer by integrating available social and census data. The approach is based on the simulation of a virtual society of agents that allows the estimate of contact matrices by age in different social contexts: household, school, workplace and general community. Those matrices are appropriately combined in order to obtain the overall ``adequate" total contacts matrix for influenza like illness. The method has been validated by comparing the obtained contact matrices by age with the results of the Polymod study the first large-scale survey on social mixing patterns relevant to infectious disease transmission (see Fig.16). The proposed method is extremely general and can be readily exported to other countries in the world for which the necessary social and demographic data can be gathered. We consider this approach an important step in order to overcome the current difficulties in real data gathering. Furthermore the computational path to the estimate of contact matrices represents a convenient scheme for the introduction of detailed individual based information in a wide range of modeling approaches working at the population level. In this case we have used a computational thinking approach to develop a novel technique for the generation of hybrid models integrating synthetic and real world data. We have then considered a standard SIR model where all the basic parameters and scaling factors are set on the baseline yielding a basic reproduction number for the European average. For each country we used the synthetic contact matrices aggregated by one-year age brackets up to 84 years of age (so that all matrices have the same dimension). All other parameters being equal, the different contact matrices in each country define different values of R0 and different epidemic behaviors in each country. By applying to every country the average European contact matrix, large differences in terms of attack rate and peak day can be observed compared to the results obtained with the country-specific mixing patterns, especially for values of the basic reproduction number consistent with influenza epidemics as shown in Figure 17. Once more we see important non-liner effects induced by considering the heterogeneity of the population. These effects are non simply linearly related to the reproductive numbers, and once more indicate that the complex features can give rise to non-intuitive behavior. It is very important to stress that these results can be generalized to social contagion processes such as the spreading of rumors, opinion formation etc., thus having an impact in the analysis of complex systems well beyond the epidemic spreading research area.

- ***Non-linear effect of data integration.***

The sensitivity analysis of most epidemic models focuses only on the parameters describing the disease. However, in a large-scale computational model, the integration and assimilation of data on census, mobility, and other demographic factors has to deal with issues related to the quality and completeness of the data. The sensitivity analysis of the model results with regard to the incompleteness or poor quality of those 'structural' data is thus extremely important. We tested this aspect by assessing whether the full complexity of the real mobility data considered in the computational platform GLEAMviz would be essential to obtain the reliable forecast of epidemic spreading and how much considering incomplete data would be affecting the model output. In particular, previous approaches in the literature have considered only one transportation mode (air travel) and included a limited number of airports, ranging from 52 to 500. Although we agree with previous studies which state that considering partial datasets is informative for the overall theoretical analysis of general spreading features, we tested the performance of partial datasets in providing results concerning the evolution of the H1N1pdm according to the calibration of the GLEAM model executed in previous studies.

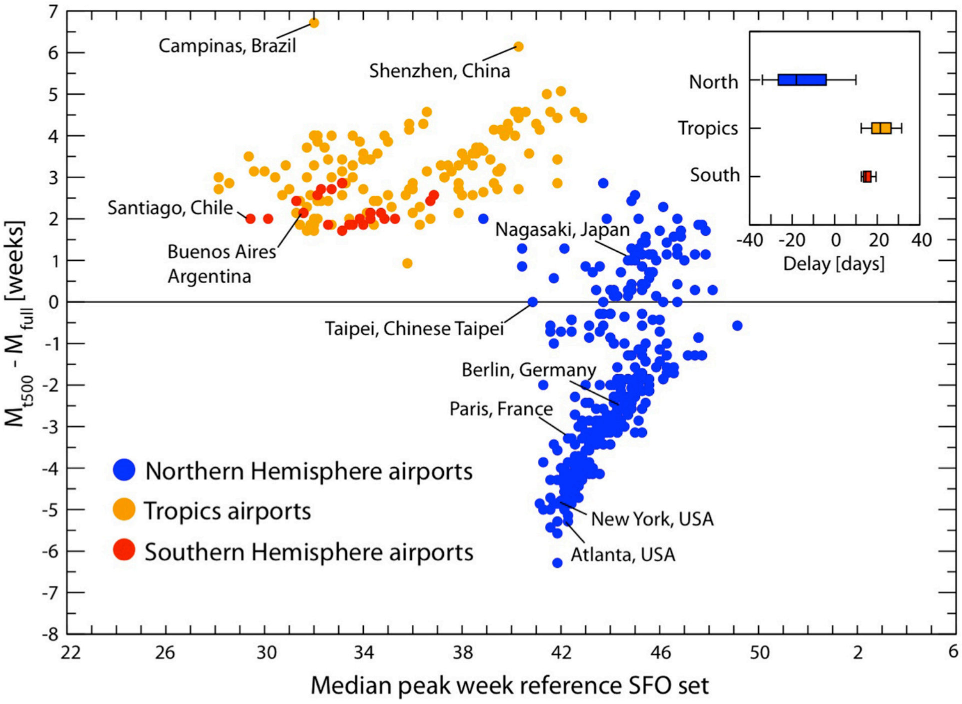


Fig.18: **Difference in the median peak weeks in the reference stochastic forecast output (SFO) set, where the full mobility dataset was considered, and the top 500 scenario, for the 500 busiest airports, as a function of the median peak week in the reference SFO set. Dots are color-coded according to the corresponding airport's climate zone. In the inset, we show the box plot indicating the distribution of the differences (in days) between the peak week of the reference SFO set and the SFO set considering only the top 500 airports. The differences were considerable, with median differences of about 3 weeks.**

We considered the top 500 airports and performed the simulations of the 2009 A/H1N1 pandemic on a version of GLEAM that integrated this partial dataset. We compared these results with the data available from the simulations including the full set of airports available in the database. In particular we compared the median peak time of the epidemic in different cities and compared the results in the full and reduced dataset (see Fig. 18). Although the top 500 airports gather about 80% of the worldwide air traffic, the differences in the median peak times are clearly non-negligible. In specific places the results differs by more than 2 months, showing that a small perturbation to the dataset produces a macroscopic change in the results of the numerical system. A specific application to a real-world epidemic is thus able to show how the global backbone of the epidemic invasion can be strongly affected by the partial sampling of the mobility network, owing to the interplay of different parameters, such as the presence of loops, local connectivity, seasonal effects, and the real and effective (that is, measured on the sampled network) distance of the location from the seed of the outbreak. In addition, a limited version of the model may not be applicable to a specific real epidemic, given its partial coverage of the locations and countries in the world, as would be the case where the initial seed of the outbreak belongs to a region not included in the data integrated into the model.

Potential impact

The self-assessment of the impact of a large-scale project such as Epiwork is not an easy task. We have put in place all the basic elements and pillow of the envisioned epidemic forecast infrastructure and interconnected them (see Fig. 19).

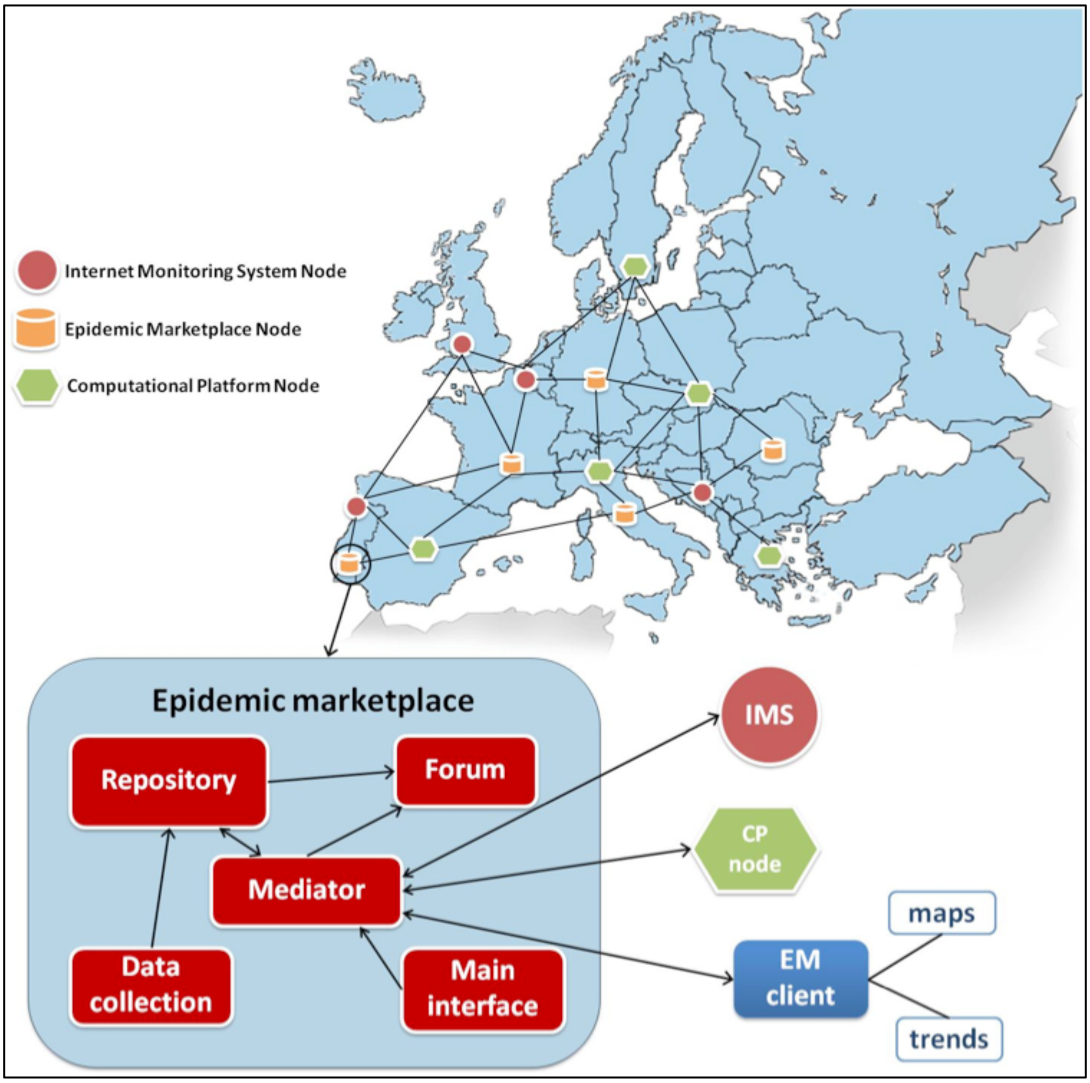


Fig.19: The Epiwork infrastructure model.

We believe that the project has delivered all its promises and has been considered a model in the area of computational and data-driven approach to the analysis and forecast of complex socio-technical systems. Other European projects (see for Instance Forecasting Financial Crisis (FOC)) are mentioning Epiwork as a model of a possible route to develop predictive systems in the area of complex networks and systems. Epiwork was used as a model "framework" in the developing of the FurtureICT concept proposal. Although this flagship proposal has not been funded, it has ranked third in a very hard competition and the proposing consortium is pursuing the FuturICT concept through other initiatives and coordination. We are looking forward to lend our expertise and tools to any future effort in this direction. In the areas of global system science, supercomputing in social systems applications and computational epidemiology the project is often cited as an example of the state of the art.

The computational platform activities have delivered state of the art models and algorithms that are generally considered the forefront of the field of computational epidemiology. GLEAM is the only global scale epidemic model using high performance computing that is available as a plug-and-play to the community. The ABM model developed by the FBK group is certainly the most accurate European-wide level epidemic model available at the moment. The visualizations tools integrated in the computational platform (SPATO, the census tessellation etc.) can be generalized and are being adopted in different contexts pertaining to the analysis of collective social phenomena.

The Epidemic marketplace platform is the first European platform devoted to the community of epidemic researchers. Assessing the impact of the platform implies considering the issues of the research community at large in sharing valuable data. Changing the habits of the epidemic modelers community to adopting data sharing practices through a platform like the Epidemic Marketplace will necessarily take time and persistence. To observe the impact of the information platform we need tools to effectively monitor the usage of the platform. The EM also includes a module for collecting usage statistics, in terms of resources, registered users and user accesses over time.

The experience of designing and operating the EM for several years, with the collected resources that it now includes is motivating the continuation of efforts towards the adoption of a common platform for sharing and describing epidemiology resources. By having their descriptions accessible in a centralized manner, epidemiology researchers can easily cross information and retrieve relevant data, which under the current model of manually searching publications often becomes a time consuming and frustrating task.
Such an endeavour is, however, largely dependent on the adoption of these practices by the community. To have an impact in changing the way epidemiology studies are conducted, both the semantic annotation and the sharing of epidemic resources needs to be encouraged. Although good practices in sharing data, actually even just metadata, of epidemiologically relevant datasets would benefit immensely the advancement of the field, it is clear that the right incentives and policies have to be set in order to initially ignite this process.

Due to the sensitive nature of epidemiological data the sharing of data becomes much more complex. However, data access can be managed, as it is in the EM, under a model where, although the metadata is in principle accessible by all users, access to the data itself can be protected and restricted to authorized users. This ensures that the data can remain private, while some of the knowledge about the dataset can still be shared. In particular, individual patient data meta-analyses, crucial in 21st century epidemiology, can be greatly aided by the adoption of a data-sharing platform.

The adoption of an epidemic data-sharing platform such as the EM would greatly support these studies, as well as more traditional research, by contributing to the reproducibility of research in the field and augmenting the knowledge of the researchers therein. Moreover, they would be expedited by the common annotation of these resources under the same metadata model and ontologies, which would greatly simplify their comparison. By having resources annotated with ontologies, we can compute similarities between resources based, for instance, on geographical location or list of symptoms, supporting near instantaneous retrieval of resources ranked by similarity. This would contribute to new avenues of research based on the analysis of similar resources, be it in terms of diseases, population or geographical location.

Finally it is worth stressing that although the EM design is focused on the scope of an epidemiological information platform, the heterogeneous nature of epidemiological content makes us think that the solution presented in this report could be offered as a package which, given some customization, can be applied to other similar problems where finding and sharing resources is the center of the problem.

In the area of disease monitoring, the project was instrumental in pushing forward the definition of a new scientific areas defined digital epidemiology and digital disease surveillance. The activities and scientists involved in the Epiwork project have contributed prominently in the awareness of the scientific community concerning the fact that the ongoing revolution in the way people communicate has given rise to a new kind of epidemiology. Digital data sources, when harnessed appropriately, can provide local and timely information about disease and health dynamics in populations around the world. At the same time the rapid, unprecedented increase in the availability of relevant data from various digital sources creates considerable technical and computational challenges. The IMS platform is a model that has been adopted in other different countries and the leadership role of the consortium is witnessed by the fact that the Influenzanet Consortium is also leading the worldwide effort networking all the platforms carrying out web-based influenza surveillance, i.e. the European Influenzanet (www.influenzanet.eu), Flu Near You (www.flunearyou.org) in the United States and Flu Tracking (www.flutracking.net) in Australia. During the 2nd International Workshop on Participatory Surveillance (IWOPS II) held in Amsterdam in April 2013 organized by the Skoll Global Threats Fund in collaboration with AIBV, aimed at mapping infectious diseases in both humans and animals, a Letter of Intent on cooperation and data exchange has been concluded between the three systems with the aim to achieve a world-wide disease radar: https://www.influenzanet.eu/en/news/2013/04/29/international-workshop-on-self-reported. Along the lines of creating a world-wide disease radar, in May 2013 we have organized the International Workshop “Public Health in the Digital Age”, in joint organization with the World Wide Web Conference 2013, Rio de Janeiro (http://www.digitalhealth.ws/). The aim of this interdisciplinary workshop was to bring together public health professionals and computer science researchers in Big Data mining, crowdsourcing and SM user engagement to raise awareness on some critical global applications: public health, user participation and risk communication.

We have also to stress the impact in the use of computational and mathematical model in the area of policy making. The consortium has made an effort (outlined in the next sections) in the outreach of agencies, policy makers and stakeholders in the area of health and infectious disease management. The results and the work done during the H1N1 pandemic was pivotal in showing the value of data-driven computational model in scenario analysis and epidemic forecast. Since the pandemic the developing of tools such the EM and GLEAMviz have been focused in making data and resources available to the community. The computational platform has devised always keeping in mind that the results of the computational models must be made accessible and available to a large community of experts that in most of the cases do not have the time or skills to work under the hood of the computational infrastructure. We believe that we were successful in providing the community tools which have the scientific rigor and easiness of use required in the context of the policy making process. The number of agencies and researchers that are using our tools are constantly growing, as detailed in the next section.

Finally let us also remark that this project has paved the way to further development and a more general approach that includes also social contagion phenomena, collective social phenomena, socio-technical systems analysis. It is possible then to envision a wealth of opportunities to leverage the results obtained by Epiwork to tackle new scientific questions and the development of new tools. A few among those are:

- Enlarge the portfolio of contagion phenomena and health problems that can be addressed in the modeling, data and monitoring infrastructure.
- Adapt the current infrastructure to allow a scalable cloud computational architecture for health crisis management.
- Integrate systematically in the computational infarstructure of the social and behavioral adaptation component.
- Extend data integration to high quality economic determinants in order to develop predictive capabilities in problems such as Global Trade and Emerging infections.
- Develop a platform for the simulation of social contagion phenomena and social collective behavior that integrates data and methodologies implemented in Epiwork.
- Extend the participatory platforms to consider other diseases and social behavior.

We believe that the project has put the seed to important future scientific and technological achievements. We are looking forward to lend our expertise and tools to any future effort in this direction.