



Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

<http://www.epiwork.eu>

Project no. 231807

D4.1 Static single layer visualization techniques

Period covered: months 1st - 12th
Start date of project: February 1st, 2009
Due date of deliverable: month 12th

Date of preparation:
Duration:
Actual submission date: February

Distribution: public

1st, 2010

Status:

Project Coordinator: Alessandro Vespignani

Project Coordinator Organization Name: ISI Foundation

Lead contractor for this deliverable: ISI Foundation

Work package participants

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

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

Change log

Version	Date	Amended by	Changes
1	04-02-2010		

Table of Contents

Visualization maps: data and simulation results	4
Simulation framework	5
Simulations and epidemiological data visualization	7
Simulations	7
Confirmed H1N1 cases and simulations	9
Animations	10

Surveillance data	12
-------------------------	----

List of figures

FIGURE 1: VORONOI DECOMPOSITION OF THE WORLD POPULATION INTO GEOGRAPHICAL CENSUS AREAS AROUND MAIN TRANSPORTATION HUBS.	6
FIGURE 2: MOBILITY NETWORKS IN EUROPE	6
FIGURE 3: NUMBER OF CASES, MAY 17. LEFT: WORST-CASE SCENARIO, WITH NO INTERVENTION; RIGHT: SCENARIO WITH ANTIVIRAL TREATMENT, 30% COVERAGE.	8
FIGURE 4: LIKELIHOOD OF OCCURRENCE OF CASES, MAY 17. LEFT: WORST CASE SCENARIO, WITH NO INTERVENTION; RIGHT: SCENARIO WITH ANTIVIRAL TREATMENT.	9
FIGURE 5: CONFIRMED NUMBER OF CASES WORLDWIDE FOR ALL INFECTED COUNTRIES, UPDATED TO MAY 19, 00:00 GMT	9
FIGURE 6: PROJECTIONS FOR WORST CASE SCENARIO, MAY 31. NUMBER OF CASES.	10

Visualization maps: data and simulation results

Describing in a comprehensive way the spatial spreading of infectious diseases critically relies on the possibility of visualizing data and results in relation to the underlying geography. Tools such as choropleth maps, dasymetric maps and mash-ups are widely used in public health to identify patterns, characterize behavior, predict outcomes and, most importantly, to inform and empower health related decision-making.

In the scope of **Work Package 4 – Epidemic Modeling Platform**, visualization techniques have been developed, specifically custom-designed for use in public health, leveraging mostly on Geographic Information Systems (GIS). The visualizations make use of many different data sources, ranging from stochastic simulations results, epidemiologic data, environmental and socio-economic data, long and short-range transportation networks, population distribution etc. The visualization of the dynamical behavior of large-scale systems represents both a theoretical and technical challenge that deals with producing easily readable and meaningful representations of massive amount of interlinked time-dependent information. The

GIS tools are merged with network visualization techniques to shape new methods to visualize the geo-temporal spread of infectious diseases.

WP 4 revolves around designing and implementing a platform for the computational modeling of infectious diseases spreading. The aim is to integrate real data and visualization techniques to perform simulations and provide access to state-of-the-art computational modeling to a wide audience of both experts and non-experts.

Simulation framework

The computational modeling platform is called GLEaM (Global Epidemic and Mobility model). GLEaM is a discrete stochastic epidemic computational model based on meta-population approach in which the world is defined in geographical census areas connected in a network of interactions by human travel fluxes corresponding to a transportation infrastructures and mobility patterns. The strength of this approach is the strong integration between high-resolution worldwide population data and commuting patterns in more than 30 countries in five continents. The highly detailed population database with demographic data allow for a Voronoi decomposition of the world surface in census cells of 15×15 minutes covering the entire Earth surface (source: [GPWv3–Gridded Population of the World v3](#), Columbia University) and centered on the International Air Transport Association (IATA) airports locations (source: [IATA](#), International Air Transport Association), each one corresponding to a subpopulation (see figure below). The air travel accounts for long-range mobility of the subpopulations, while the commuting patterns include the effect of short range mobility corresponding to ground movements among subpopulations.

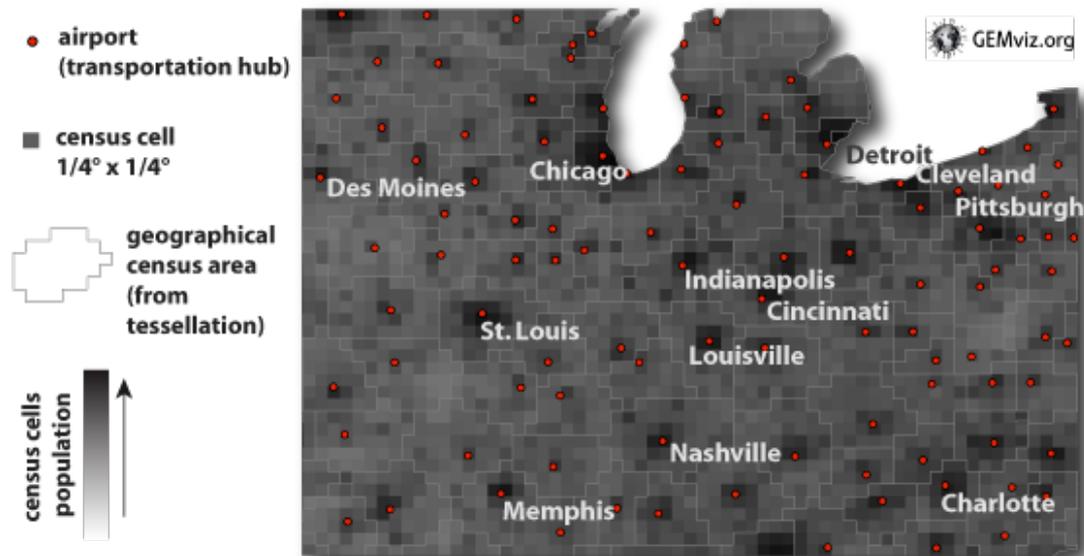


Figure 1: Voronoi decomposition of the world population into geographical census areas around main transportation hubs.

The multi-scale networks obtained by means of such data integration are the substrate on top of which the spreading of the infection is explored.

Below, the long range (left) and short range mobility patterns in Europe are shown.

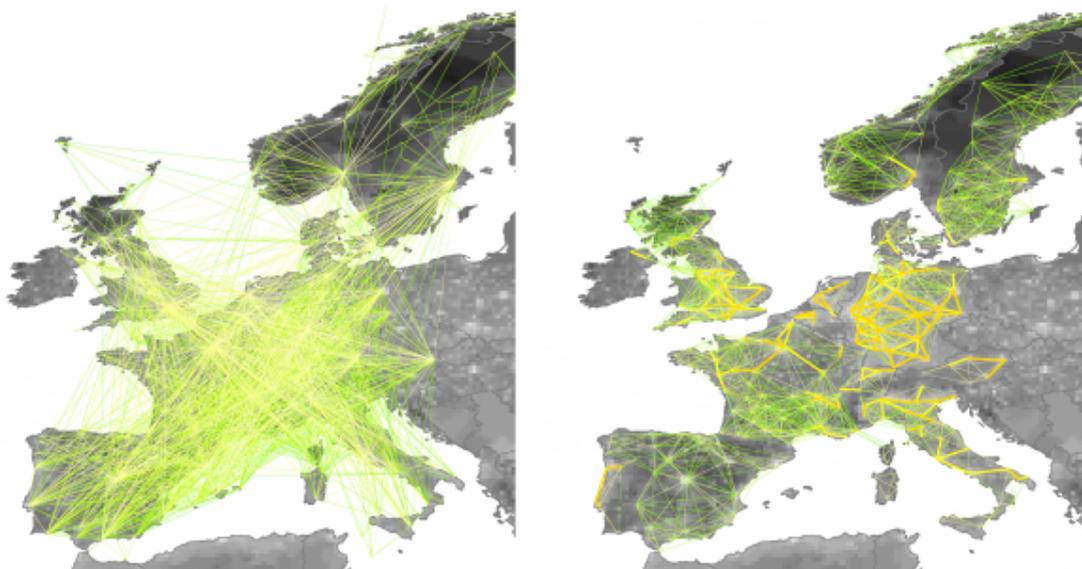


Figure 2: Mobility Networks in Europe

The infection dynamics takes place inside each geographical census area, and is described by compartmental schemes in which the discrete stochastic dynamics of the

individuals among the different compartments depend on the specific etiology of the disease and the containment interventions considered. Local outbreaks taking place in the geographical census areas are coupled by the mobility (short- and long-range) of individuals.

GLEaM simulation engine provides simulations data about the daily incidence, prevalence or cumulative incidence at different geographical levels (basin, region, country, continent, etc.). Data are stored in dbf (data base file) format.

Simulations and epidemiological data visualization

Simulations

The numerical output of the simulations needs to be analyzed, visualized and compared to further understand the spreading process. The use of visual geographical representation is key to understanding, therefore our work in this direction was to link simulation outcome and visualization tools.

ArcGIS™ is an integrated collection of GIS software products that provides a standard-based platform for spatial analysis, data management, and mapping. ArcGIS is able to handle geo-referenced files, called shapefiles (.shp). A shapefile contains the geographic description of a layer (e.g. a point layer, or a line, polygon layer), whose elements can be described in terms of a list of attributes (e.g. a country can be characterized by name, population, surface area, etc.). Attributes can belong to the original definition of a shapefile or can be added by the user through a dbf file. It is then straightforward to add the simulation data to the map, by simply linking the simulations output dbf file to the shapefile. By doing so, the shapefile can store the data of daily incidence, prevalence or cumulative incidence, at the geographical resolution of choice and at several different dates. The resulting integrated map can be exported in a graphic format, such as jpeg, png, and pdf, and also transformed in form of animation. The geographic spread of a disease can thus be explored in all its aspects by means of this powerful tool.

The GLEaM simulation engine has been used to characterize the H1N1 2009 pandemic influenza, forecast its future spread and identify strategies to mitigate it.

Simulations were calibrated by using the reported cases of H1N1 during the early stages of the outbreak.

The Figure 3, below, provides updated results for the comparison between a worst-case evolution scenario and a scenario in which some basic containment measures are considered. In this second scenario early detection of cases and their treatment with antiviral drugs are included. The simulations were calibrated considering the actual confirmed data as of May 06. The maps show the spatial distribution of the expected number of cases and the maximum number of cases (in parentheses) according to a 90% reference range in major urban areas. The color code corresponds to the maximum number of cases at a resolution level of $\frac{1}{4}^\circ$.

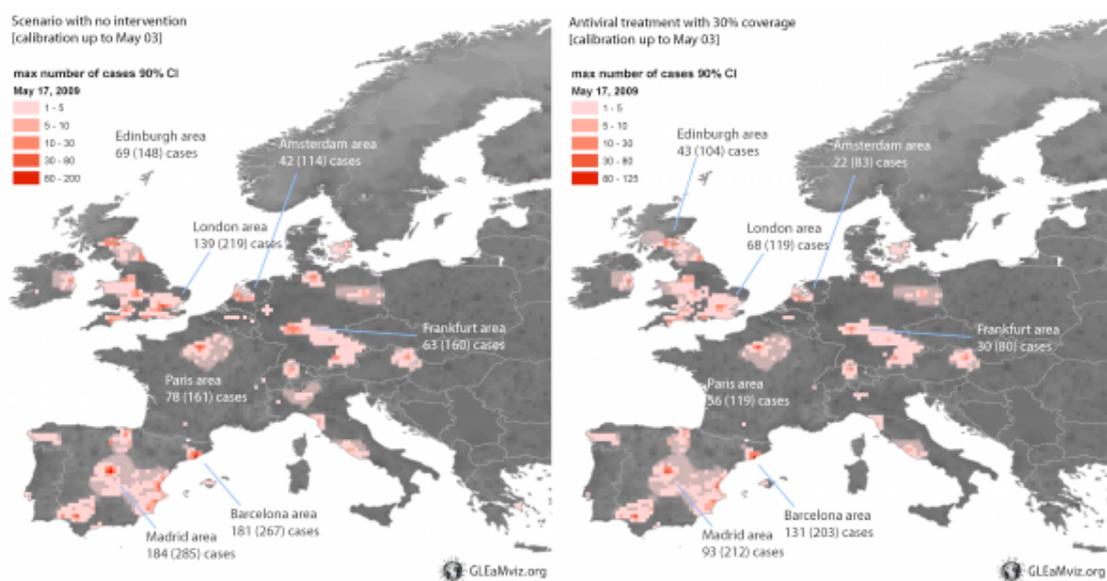


Figure 3: Number of cases, May 17. Left: worst-case scenario, with no intervention; right: scenario with antiviral treatment, 30% coverage.

It was also crucial to build risk maps (Figure 4) illustrating the likelihood of case generation in two different scenarios: the worst-case and the containment scenario. The two maps show how the country risk of being reached by the infection at that date was reduced by the introduction of antiviral drugs in the treatment of cases.

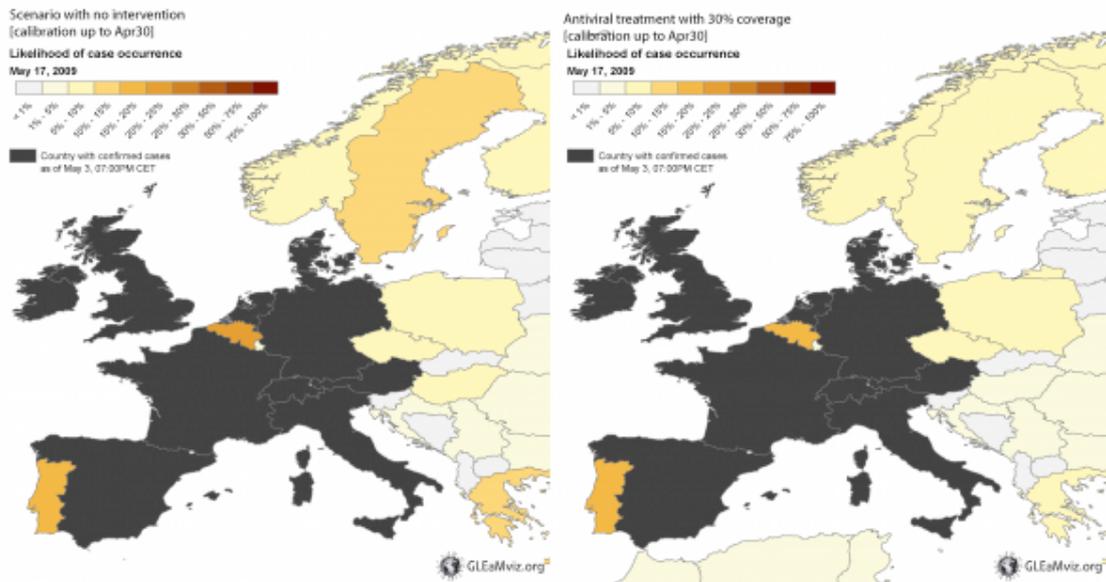


Figure 4: Likelihood of occurrence of cases, May 17. Left: worst case scenario, with no intervention; right: scenario with antiviral treatment.

Confirmed H1N1 cases and simulations

During the early phase of the pandemic, a huge amount of epidemiologic data was collected daily. The evolution of the spreading was readily visualized on choropleth maps (below figure) where every reported number of cases was assigned to each country reached by the infection, and –where possible- to a more localized area within the country, according to the information available.

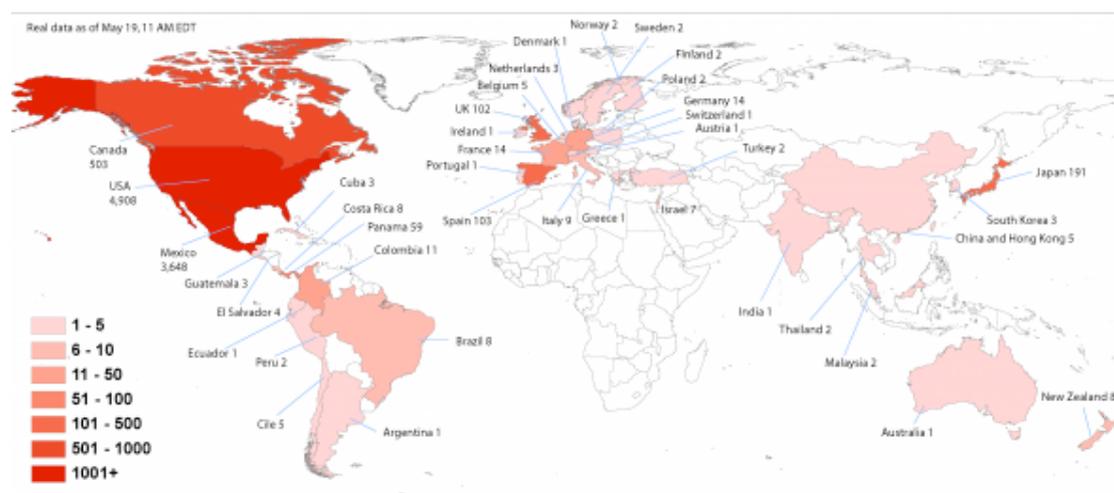


Figure 5: confirmed number of cases worldwide for all infected countries, updated to May 19, 00:00 GMT

The official data coming from the US were just a fraction of the actual cases. As of May 22 the [CDC was stating](#) “...more than 100,000 people probably have this virus now in the United States”. This number was obtained by applying a multiplying factor of approximately 10-20 to the number of confirmed cases, in order to take into account the sampling and under reporting of cases.

In order to assess the qualitative agreement of our simulations with the empirical data collected, the above map was compared to the projections obtained (maps below, projections for the date May 31 published on <http://www.gleamviz.org> by means of GLEaM simulations). The color code corresponds to the maximum of the 95% reference range obtained by the stochastic runs of GLEaM. Numbers were also reported for the countries expected to experience the largest outbreak: the first value corresponds to the expected number of cases, and the value in brackets corresponds to the maximum of the 95% reference range.

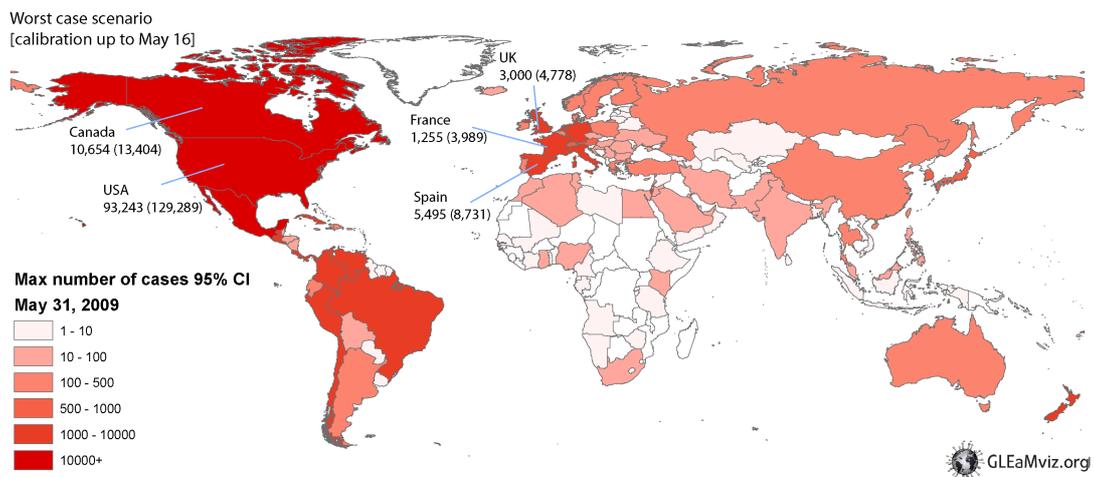
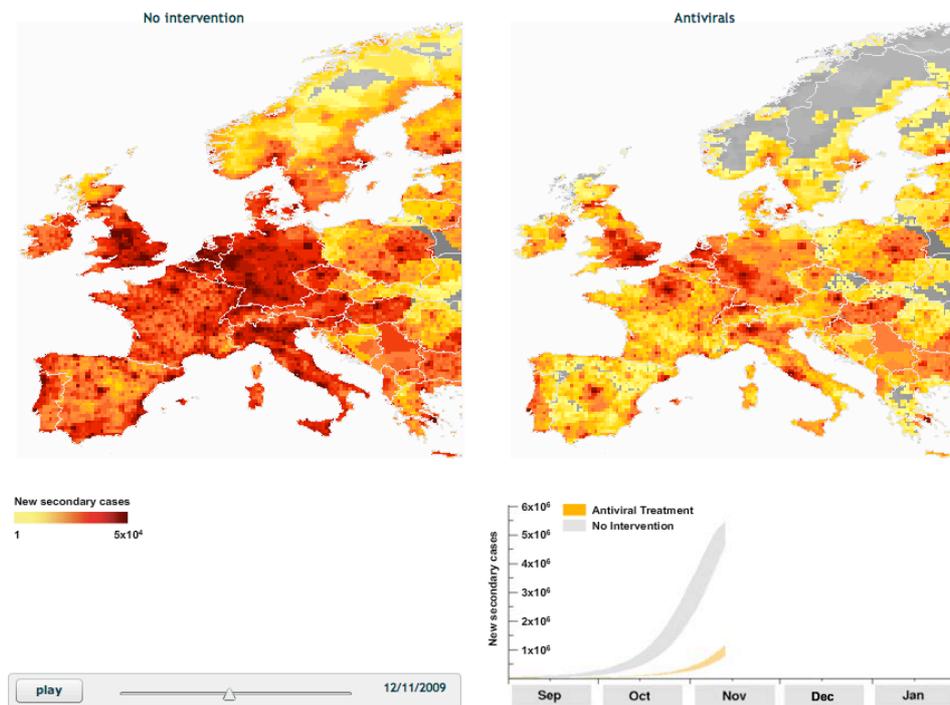


Figure 6: projections for worst case scenario, May 31. Number of cases.

Animations

The visualization of the dynamical behavior of such large-scale systems can also be performed by means of map layers vertically overlaid and integrated to form an animation that shows the time and space evolution of the epidemic in different conditions.

Below is a snapshot of an animation showing the predicted geo-temporal evolution of the H1N1 influenza epidemic in Europe for the fall of 2009, comparing the no intervention scenario (left map) with a scenario in which mitigation strategies were considered (right map). Simulations for the no-intervention scenario were obtained from the maximum likelihood analysis used to estimate the transmission potential of the new influenza A(H1N1) (Balcan D et al, *Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility*. BMC Medicine 2009, 7:45).

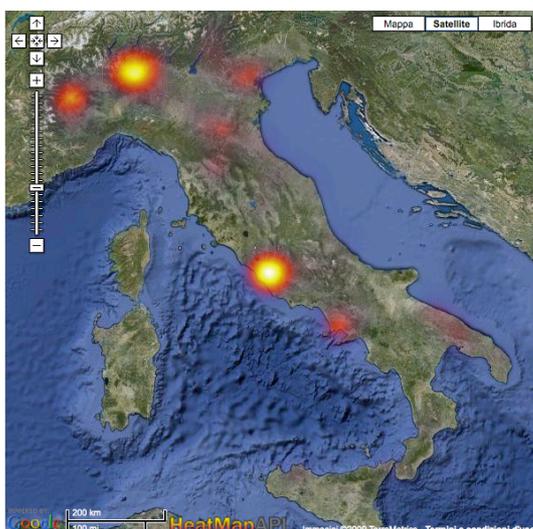


This kind of animation has been posted on the website reporting all our research results related to GLEaM and H1N1 (<http://www.gleamviz.org/2009/09/eu-winter-projections-mitigation-effect-of-antiviral-treatment/>). The animation is built by means of static maps integrated in a mapfile and used by Mapserver (an OpenSource platform for publishing spatial data and interactive mapping applications to the web) to generate a sequence of snapshots, each corresponding to one day in the timeline of the epidemic evolution. The snapshots were then assembled in a Flash™ animation along with the quantitative information by means of a plot.

The maps display the daily new number of cases at a resolution level of $\frac{1}{4}^\circ$ with a color code ranging from yellow (low activity) to dark red (peak activity). The data mapped were obtained from the average value of the daily new number of cases calculated on 2,000 stochastic realizations of the model, for each of the two scenarios. The timeline shown is from September 1, 2009 to January 31, 2010. The maps have been developed at the beginning of 2009 fall, before the fall wave of H1N1 pandemic would strike. The plot at the bottom of the page reports the corresponding country profiles in the same time window. The shaded area corresponds to the 95% CI obtained from the simulations. The two animations readily show the time delay in reaching the epidemic peak achieved with the implementation of antiviral treatment.

Surveillance data

Visualization techniques are crucial also for surveillance purposes, to display epidemiologic surveillance data allowing the examination of the disease incidence at a given level of spatial resolution and in real time, as the data are collected. For example, Web surveillance data collected by the WP5 influenza surveillance platforms (Internet Monitoring Systems or IMS), in the several countries participating to the project, can be conveniently displayed in a multi-layer fashion with zoom-in and zoom-out features on top of Google API maps (figures below).



The left map shows the ILI cases reported by the participants of the Italian IMS Inluweb (www.influweb.it). ILI cases are shown by means of a heat map, i.e. a

graphical representation of three-dimensional data (X, Y and Value) on a two dimensional surface by using colors. On a geographical map, a heat map is a representation of the density of latitude and longitude points, each corresponding to the geographical location of a reported case of ILI. Heat maps allow users to promptly visualize the cumulative density of the reported cases in the country. The Italian heat map is generated by means of Googlemaps API integrated with external API's (<http://www.heatmapapi.com/>) that allow users to build a customized heat map on top of a Googlemap.

The right map is a mash-up combining surveillance data collected with the contribution of the volunteers of the UK IMS Flusurvey (flusurvey.org.uk) and Googlemaps API. Green (healthy users), blue (users with cold) and red (users with ILI) markers are placed at the latitude and longitude of the postal code where each user lives. This allows to promptly visualize in real time the general health status reported by the users throughout the season. This approach was first developed for the Italian IMS and then readily integrated into the UK platform for the surveillance of H1N1 influenza pandemic.