Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

http://www.epiwork.eu

Project no. 231807

D 8.4 Final Year Workshop

Period covered: Months 36th – 54th
Start date of project: February 1st, 2009
Due date of deliverable: Month 54th
Distribution: public

Date of preparation: Month 54
Duration:
Actual submission date:
Status:

Project Coordinator: Alessandro Vespignani
Project Coordinator Organisation Name: ISI Foundation
Lead contractor for this deliverable: ISI

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
- FCG-IGC
- TAU
- MPG
- AIBV
- SMI
- KULeuven
- CREATE-NET

Change log

Version	Date	Amended by	Changes

Epiwork Final International Workshop on "Digital Epidemiology"

Fondazione ISI

Via Alassio 11/c 10126 Torino

May 30,31 2013

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, citizens engagement and support to public health interventions at individual and population levels.

This workshop aims to explore how this revolution impacts all these aspects. The availability of real time Big Data from social networks, blogs, games, participatory web platforms provide a constant stream of situation-aware information which, in joint action with cutting edge Computer Science and ICT, allow for prompt data harvesting and analyses and response to public health threats. Existing systems illustrating the potential of analysis of online data include HealthMap, participatory systems for flu surveillance in three different continents (namely Influenzanet, Flu Near You and Flutracking), Medysis, Crowdbreaks, ProMed, biodiaspora, Google Flu Trends.

Moreover, social media and social networks have reshaped the routes on which risk communication and public information about health threats spread among the population.

The goal of this workshop is to bring together public health professionals working in public health and epidemic intelligence services in WHO, ECDC, CDC and researches in on line data mining, crowdsourcing and Social Media user engagement to raise awareness on the critical global applications of public health, user participation and risk communication.

The topics include:

- Web-based participatory systems for public health
- epidemic intelligence and early warning systems from social media data mining
- crowdsourced data collection, user engagement
- situational awareness for public health
- data analysis of disease prevalence
- visual analytics and exploratory data analysis of social media for public health
- developed of integrated platforms harvesting and analysing real time data for early warning and rapid response
- information diffusion models for public health
- Disease Surveillance, detection and reporting

The website of the Workshop can be found at the page: http://www.isi.it/workshop_digital_epidemiology

The slides of the various presentations can be found at the page: http://www.isi.it/workshop_digital_epidemiology/talk-slides

Program

	Thursday May 30
1.00- 2.10 pm	*Registration and Lunch*
2.10 - 2.20 pm	Opening
	Session 1 - Social Networks and Public Health **chair: Daniela Paolotti**
2.20-3.00 pm	**Marcel Salathé** *The dynamics of Health Behavior Sentiments on a Large Online Social Network*
3.00-3.40 pm	**Mark Dredze** *Why you can't predict influenza from Twitter data – and here's how*
3.40-4.00 pm	**Karuthan Chinna** *Using Google Dengue Trend to Track Dengue Outbreaks*
4.00-4.30 pm	***coffee break/posters***
	chair: Marcel Salaté
4.30-4.50 pm	**Paola Velardi** *ILI Surveillance through natural language processing on Twitter*
4,50-5.10 pm	**Patty Kostkova** *Risk Communication of Public Health Emergencies: is there a role for Twitter?*
	Session 2 - Contacts
5.10-5.50 pm	**Gabriel Leung** *Preliminary observations on validating SocioPatterns measurement in a Hong Kong primary school*
5.50-6.10 pm	**Alain Barrat** *An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices*
6.10-6.30 pm	**Mart Stein** *Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand*
8.00pm	**Social Dinner**

	Friday May 31
	Session 3 - Participatory Studies & Beyond **chair: Pasi Penttinen**
9.00-9.40 am	**John Edmunds** *The role of the internet in disease surveillance in the 21*st *century*
9.40-10.00 am	**Yannick Vandendjick** *Web-based monitoring of Influenza-like illness in Flanders: the Great Influenza Survey*
10.00-10.20 am	**Christophe Bayer** *Internet-based syndromic monitoring of acute respiratory illness in the general population of Germany during week 35 /2011 – 34/2012*
10.20-10.40 am	**Paolo Bajardi** *Effect of recruitment methods on attrition in internet-based studies*
10.40-11.10 am	**Coffee break**
	Chair: Patty Mabry
11.10-11.50 am	**Pasi Penttinen** *Epidemic Intelligence at ECDC – where are the gaps*
11.50 -12.30 pm	**Rumi Chunara** *Crowdsourcing Technologies for Improved Public Health Surveillance*
12.30-12.50 pm	**Vittoria Colizza** *Who’s your citizen scientist? The European participatory flu surveillance experience*
12.50-2.20 pm	**Lunch & posters**
	Session 4 - Needs, Tools and Methods for Digital Epidemiology **chair: Rumi Chunara**
2.20-3.00 pm	**Patty Mabry** *Behavioral Digital Epidemiology: Current Research and Future Directions*
3.00-3.20 pm	**Màrio Silva** *Epidemic Marketplace - An information platform for epidemiological data*
3.20-3.40 pm	**Duccio Medini** *Data challenges for the development of the Novel 4CMenB Vaccine*
3.40-4.00 pm	**coffee break**

	Chair: John Edmunds
4.00-4.20pm	**Adrienn Ecseki** *Healthcare databases in Hungary; present and future*
4.20-4.40 pm	**Etienne Sévin** *Developing opensource solutions for mobile data collection: example of a hospital-based cohort*
	Session 5 - Modeling
4.40-5.00 pm	**Caterina Scoglio** *Individual-based information Dissemination in Multi-layer Epidemic Modeling*
5.00-5.20 pm	**Eiko Yoneki** *Influential Neighbours Selection for Information Diffusion*

Posters
Paulo Graça *A Group-based Access Control Model for Protected Scientific Data*
Francisco Couto *Semi-automated Annotation of Epidemiological Resources*
Andrzej Jarynowski *Knowledge of the spread of sexually transmitted diseases through computer simulations*
Veronica Dorgali *Kicking against the pricks: vaccine skeptics have a different social orientation*
Daniela Paolotti *Determinants of participation in the Influenzanet Cohort*
Carl Koppeschaar *Disease radar:* Measuring and Forecasting the Spread of Infectious Diseases and Zoonoses

Program at a glance

Session – Social Networks and Public Health

1) M. Salathé, The Dynamics of Health Behavior Sentiments on a Large Online Social Network
2) M. Drezde, Why you can't predict influenza from Twitter data - and here's how.
3) K. Chinna, Using Google Dengue Trend to Track Dengue Outbreaks
4) P. Velardi, ILI surveillance through natural language processing on Twitter
5) Patty Kostkova, Risk Communication of Public Health Emergencies: is there a role for Twitter?

Session – Contacts

1) G. Leung, Preliminary observations on validating SocioPatterns measurement in a Hong Kong primary school
2) A. Barrat, An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices.
3) M. Stein, Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand

Session – Participatory Studies & Beyond

1) J. Edmunds, The role of the internet in disease surveillance in the 21st century
2) Y. Vandendjick, Web-based monitoring of influenza-like illness in Flanders: the Great Influenza Survey
3) C. Bayer, Internet-based syndromic monitoring of acute respiratory illness in the general population of Germany during week 35 /2011 – 34/2012.
4) P. Bajardi, Effect of recruitment methods on attrition in internet-based studies
5) P. Penttinen, Epidemic Intelligence at ECDC – where are the gaps
6) R. Chunara, Crowdsourcing Technologies for Improved Public Health Surveillance
7) V. Colizza, Who’s your citizen scientist? The European participatory flu surveillance experience

Session – Needs, Tools and Methods for Digital Epidemiology

1) P. Mabry, Behavioral Digital Epidemiology: Current Research and Future Directions
2) M. Silva, Epidemic Marketplace - An information platform for epidemiological data
3) D. Medini, Data challenges for the development of the Novel 4CMenB Vaccine
4) A. Ecseki, Healthcare databases in Hungary; present and future
5) E. Sévin, Developing opensource solutions for mobile data collection: example of a hospital-based cohort

Session – Modeling

1) C. Scoglio, Individual-based Information Dissemination in Multi-layer Epidemic Modeling
2) E. Yoneki, Influential Neighbours Selection for Information Diffusion

Posters

1) P. Graca, Access Control for Shared Epidemic Datasets
2) F. Couto, Semi-automated Annotation of Epidemiological Resources
3) A. Jarynowski, Knowledge of the spread of sexually transmitted diseases through computer simulations
4) V. Dorgali “Kicking against the pricks: vaccine skeptics have a different social orientation”
5) D. Paolotti “Determinants of participation in the Influenzanet Cohort”
6) C. Kopperschaar, Disease radar:Measuring and Forecasting the Spread of Infectious Diseases and Zoonoses

Invited Speakers - Abstracts

Marcel Salathé

Title: *The Dynamics of Health Behavior Sentiments on a Large Online Social Network*

Abstract:

Modifiable health behaviors, a leading cause of illness and death in many countries, are often driven by individual beliefs and sentiments about health and disease. Individual behaviors affecting health outcomes are increasingly modulated by social networks, for example through the associations of like-minded individuals - homophily - or through peer influence effects. Using a statistical approach to measure the individual temporal effects of a large number of variables pertaining to social network statistics, we investigate the spread of a health sentiment towards a new vaccine on Twitter, a large online social network. We find that the effects of neighborhood size and exposure intensity are qualitatively very different depending on the type of sentiment. Generally, we find that larger numbers of opinionated neighbors inhibit the expression of sentiments. We also find that exposure to negative sentiment is contagious - by which we merely mean predictive of future negative sentiment expression - while exposure to positive sentiments is generally not. In fact, exposure to positive sentiments can even predict increased negative sentiment expression. Our results suggest that the effects of peer influence and social contagion on the dynamics of behavioral spread on social networks are strongly content-dependent.

Michael Dredze

Title: *Why you can't predict influenza from Twitter data - and here's how.*

Abstract:

We're in the middle of an influenza epidemic. We need reliable information about the current flu rate, the sooner the better. Web and social media data, such as Twitter, have been shown to deliver real time information about the flu rate. However, a closer look reveals a more complex picture, and analyses on multiple flu seasons show current methods for Twitter lacking.

We present a new approach to tracking flu on Twitter that addresses these problems, including the ability to separate out reports of infection from general flu chatter. Judging from several flu seasons, including the current epidemic, our approach yields state of the art results for measuring influenza infection rates.

Gabriel Leung

Title: *"Preliminary observations on validating SocioPatterns measurement in a Hong Kong primary school"*

John Edmunds

Title: *The role of the internet in disease surveillance in the 21st century*

Pasi Penttinen

Title: *Epidemic Intelligence at ECDC – where are the gaps*

Abstract:

The mandate of ECDC mandate is to identify, assess and communicate current and emerging health threats to human health from communicable diseases. Within this mandate Epidemic intelligence (EI) is used at ECDC for threat detection and monitoring. EI is defined as the systematic collection, collation, validation and analysis of information from a variety of indicator-based and event-based surveillance (EBS) sources. The main focus of EI at ECDC is on EBS related components and follows a systematic process that starts with information screening and event filtering (risk detection). Filtered events are validated and analysed for their potential threat to human health (risk assessment). Several web-based news aggregators and early warning systems are used on a daily basis for this purpose, which leads to an over-reliance on media information. A comparison of EU threats detected at ECDC using these sources over a 5-year period 2008-2012 is presented, with a comparison with results from indicator-based surveillance and literature. Potential ways to increase coverage of threat detection are discussed.

Rumi Chunara

Title: *Crowdsourcing Technologies for Improved Public Health Surveillance*

Abstract:

Global patterns of disease burden are constantly shifting. At the same time, traditional public health capacity has remained stagnant, suffering from limits due to latency, high cost, inherent contributor biases and imprecise resolution. A critical barrier in disease epidemiology and control has been a lack of rapid, quantitative intelligence from the point of care. New tools including Internet-based and point-of-care sensors using mobile phones, smart-messaging service (SMS) and smart phone applications can help us address these gaps by sourcing information directly from individuals. In this talk I will discuss how we use this crowdsourced data in a variety of settings to obtain better spatial and temporal insight into disease dynamics.

P. Mabry
Title: *Behavioral Digital Epidemiology: Current Research and Future Directions*

Abstract:

Behavioral digital epidemiology is the study of behavioral and social phenomena that spread through social networks or affect the spread of diseases and conditions that spread through social networks. This can include behaviors relevant to infectious diseases, such as panic behavior, obeying or disobeying quarantines and travel restrictions, handwashing, mask wearing and vaccination. It can also include behaviors relevant to chronic diseases such as physical activity, dietary behavior, drug and alcohol use, smoking, risky sexual behaviors, and stress relieving behavior, as well as attitudes, beliefs and knowledge. In this talk, I will review a few illustrative studies in behavioral digital epidemiology. I will also discuss important methodological considerations for the future of this field including measurement issues, and verification and validation issues. Finally, I will describe some of NIH's interests and opportunities in this field.

Oral Contributions - Abstracts

SESSION - Social Networks and Public Health

K. Chinna
Organization/Institution name: University Malaya, Kula Lumpur
Department: Social & Preventive Medicine

Title: *Using Twitter to Track Dengue Outbreaks*

Abstract:

As in other parts of the world, dengue ranks as one of the most significant mosquito-borne viral human diseases in Malaysia. Based on the recently published national health indicator, 18,500 dengue cases were reported in 2011 and the incidence rate was 64 per 100,000. For dengue disease control, timely detection and reporting of dengue cases is essential. Early identification can help minimize the spread of disease. Traditional disease surveillance activities are tedious and the lack resources have interfered with timely detection and reporting.

Dengue data from official sources are often not available until after some substantial delay.

Alternatively, surveillance efforts can be turned to "Big Data Analytics" data sources, such as social media search queries, which have been shown to be effective for monitoring disease outbreaks, like influenza. Having real-time information available in the public domain through social media, like Twitter, could be useful for watching out for the first clues to major dengue outbreaks and detecting developing disease hotspots. A new Web application, MappyHealth, uses Twitter to monitor and produce automatic hourly reports on what illnesses people are talking about the most.
In this on-going study, data are captured from Twitter, using MappyHealth. These data are then cross checked with the data obtained from the healthcare department in the vicinity of the Twitter data origin. If there are any spurious spikes, they are removed.
Univariate linear models are tested by fitting a time lagged fraction of Twitter data capture against the 'gold standard' time series of dengue cases reported.

Keywords: Social media, dengue, Twitter, MappyHealth

P. Velardi
Organization/Institution name: Bambino Gesù Children's Hospital
Department: Research Area of Multifactorial Diseases and Complex Phenotypes

Title: *ILI surveillance through natural language processing on Twitter*

Abstract:

Syndromic surveillance is aimed at identifying illness clusters early, before diagnoses are confirmed and reported to public health agencies, in order to mobilize a rapid response (MMWR 2004). In recent years, surveillance data became available directly from individuals through the digital traces they leave as a consequence of social networking and increased use of electronic devices1 (eg: Google FLuTrends, FluNearYou). Twitter has also been analysed as a source of syndromic surveillance data, through algorithms that identified occurrence of single prespecified terms, consisting either in the name of the clinical condition or its synonyms (eg: H1N1 or swine flu), or in prespecified terms related to the clinical syndrome itself.

Since people use everyday language rather than medical jargon, knowledge of patient's terminology in blog mining may improve the accuracy of twitter analysis for syndromic surveillance. We present a minimally trained algorithm to learn associations between technical and everyday language terms, based on pattern generalization and complete linkage clustering. Starting from a small set of medical conditions composed by 193 pairs of technical and common terms manually extracted from Freebase, the algorithm goes through the following 5 steps: web mining, pattern generalization, clustering, reinforcement, testing phase. Performance of learning processes was evaluated through the following parameters: precision, recall, F-measure, coverage e mean reciprocal rank (MRR). The best clustering result was selected on the basis of the balance between MRR and a cross-validation threshold and used as the final model for extracting naïve medical language.

We applied this approach to surveillance of influenza-like illness. First, we run the algorithm on the technical terms extracted from the ECDC case definition for ILI8, obtaining an additional set of ILI-related jargon terms. Secondly, we built a Boolean query based on the ECDC case definition for ILI, using both technical and related jargon terms as identified by the algorithm. Using the available APIs, we are collecting a dataset of Twitter messages including at least one of the singleton terms composing our query. On this dataset, we extracted a set of tweets matching the ILI query.

We will present results of tweet message monitoring as of the end of the flu season, and we will compare the Twitter trends with data from routine surveillance systems.

The main contribution vrs previous work on infodemiology (and other applications of social media mining) is i) relevant keywords are extracted automatically, rather than manually identified; ii) the keywords are used to obtain detailed information on observed diseases (e.g. clustering micro-blogs of patients according to available disease case descriptions), rather than solely predicting disease outbreaks using keyword frequency and query volumes. The originality of our approach is in the formulation of a case definition specific for blog and social network mining: first, we will not use single terms as keywords, but a complex Boolean query translating a case definition; secondly, jargon terms will be included in the query, allowing to trace both technical and naïve terminology and thus producing a much larger body of evidence.

P. Kostkova
Organization/Institution name: UCL
Department: Computer Science

Title: *Risk Communication of Public Health Emergencies: is there a role for Twitter?*

Abstract:

Digital epidemiology is a fast emerging research discipline covering many aspects of public health, however, risk communication is often neglected. The communication of risk in any public health emergencies is a complex task traditionally managed by governments and healthcare agencies. This task is made more challenging in the current situation when citizens are confronted with a wide range of online resources, ranging from traditional news outlets to information posted on blogs and social networks. Inevitably, some have greater scientific veracity than others. Twitter stands out as a resource with a multiple function: it is an information source but is also a central hub for the publishing, dissemination, and finding out about online media.
In our research we investigated the relationship between social network propagation of media coverage of the WHO declaration of swine flu 2009 a "pandemic" on 11th June 2009 and the spread of this dramatic piece of information the following 24 hours. Our results demonstrated that more 'reputable' and 'trusted' media such as BBC are more successful in communicating the risk of pandemic through Twitter than less reputable outlets, while communications from the public heath agencies themselves (WHO, CDC, ECDC) were largely ignored interest by Twitter users. While this was a pilot study, this illustrates an important lesson for health authorities planning future risk communications and stresses the need to a close collaboration with media to ensure accurate, scientific and timely coverage.

SESSION - Contacts

A. Barrat
Organization/Institution name: CNRS & ISI Foundation

Title: *An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices.*

Abstract:

Background
The integration of empirical data in computational frameworks designed to model the spread of infectious diseases poses a number of challenges that are becoming more pressing with the increasing availability of high-resolution information on human mobility and contacts. This deluge of data has the potential to revolutionize the computational efforts aimed at simulating scenarios, designing containment strategies, and evaluating outcomes.
However, the integration of highly detailed data sources yields models that are less transparent and general in their applicability. Hence, given a specific disease model, it is crucial to assess which representations of the raw data work best to inform the model, striking a balance between simplicity and detail.

Methods
We consider high-resolution data on the face-to-face interactions of individuals in a pediatric hospital ward, obtained by using wearable proximity sensors. We simulate the spread of a disease in this community by using an SEIR model on top of different mathematical representations of the empirical contact patterns. At the most detailed level, we take into account all contacts between individuals and their exact timing and order. Then, we build a hierarchy of coarse-grained representations of the contact patterns that preserve only partially the temporal and structural information available in the data. We compare the dynamics of the SEIR model across these representations.

Results
We show that a contact matrix that only contains average contact durations between role classes fails to reproduce the size of the epidemic obtained using the high-resolution contact data and also fails to identify the most at-risk classes. We introduce a contact matrix of probability distributions that takes into account the heterogeneity of contact durations between (and within) classes of individuals, and we show that, in the case study presented, this representation yields a good approximation of the epidemic spreading properties obtained by using the high-resolution data.

Conclusions
Our results mark a first step towards the definition of synopses of high-resolution dynamic contact networks, providing a compact representation of contact patterns that can correctly inform computational models designed to discover risk groups and evaluate containment policies. We show in a typical case of a structured population that this novel kind of representation can preserve in simulation quantitative features of the epidemics that are crucial for their study and management.

M. Stein
Organization/Institution name: University Medical Center Utrecht
Department: Utrecht Center for Infection Dynamics (UCID)

Title:
Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand

Abstract:

Background:
Information on social interactions is needed to understand the spread of respiratory infections through a population. However, previous studies mostly collected egocentric information hereby ignoring large-scale network structures within populations which are important for disease dynamics. Respondent-driven sampling (RDS) is a sampling technique allowing respondents to recruit contacts from their social network. We explored the feasibility of online RDS for studying contact patterns relevant for the spread of respiratory pathogens.

Methods: We developed a web-based RDS system for tracking and facilitating recruitment by Facebook and email. One-day diary surveys were conducted by applying webRDS among a convenience sample of Thai students. Students were asked to record numbers of social encounters at different locations and self-reported influenza-like-illness symptoms, and to recruit online four contacts whom they had met in the previous week. These contacts were asked to do the same to create a network tree of socially connected individuals.

Results: 44 (23.0%) of the initially approached students recruited one or more contacts. In total 257 persons participated, of which 168 (65.4%) were recruited by others. Facebook was the most popular recruitment option (45.1%). We reached up to 6 waves of contacts of initial respondents, using non-material incentives. Strong assortative mixing was seen for age between pairs of linked nodes, indicating a tendency of participants to connect to contacts of similar age. By contrast, (weak) disassortative mixing was seen for daily encounters or degree, implying a social connectedness between individuals with different levels of contacts.

Conclusions: Despite methodological challenges (e.g. clustering among participants and their contacts), applying RDS provides new insights in mixing patterns relevant for close-contact infections in real-world networks. Such information increases our knowledge of the transmission of respiratory infections within populations and can be used to improve existing modeling approaches.

SESSION - Participatory Studies & Beyond

Y. Vandendijck
Organization/Institution name: Hasselt University
Department: Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

Title:
Web-based monitoring of influenza-like illness in Flanders: the Great Influenza Survey

Abstract:

In 2003, an internet-based monitoring system of influenza-like illness (ILI), the Great Influenza Survey (GIS), was initiated in Belgium. For the Flemish part of Belgium, we investigate the representativeness of the GIS population and assess the validity of the survey in terms of ILI incidence during eight influenza seasons (from 2003 through 2011). The validity is investigated by comparing estimated ILI incidences from the GIS with recorded incidences from two other monitoring systems, (i) the Belgian Sentinel Network and (ii) the Google Flu Trends, and by performing a risk factor analysis to investigate whether the risks on acquiring ILI in the GIS population are comparable with results in the literature. A random walk model of first order is used to estimate ILI incidence trends based on the GIS. Good to excellent correspondence is observed between the estimated ILI trends in the GIS and the recorded trends in the Sentinel Network and the Google Flu Trends. The results of the risk factor analysis are in line with the literature. In conclusion, the GIS is a useful additional surveillance network for ILI monitoring in Flanders. The advantages are the speed at which information is available and the fact that data is gathered directly in the community at an individual level.

Author (s): Yannick Vandendijck, Christel Faes en Niel Hens.

C. Bayer
Organization/Institution name: Robert Koch Institute
Department: Department for Infectious Disease Epidemiology

Title:
Internet-based syndromic monitoring of acute respiratory illness in the general population of Germany during week 35 /2011 – 34/2012.

Abstract:

In 2011, the German sentinel surveillance system for influenza (“Arbeitsgemeinschaft Influenza” (AGI)) was complemented by an internet based syndromic monitoring system (GrippeWeb) to measure acute respiratory infections (ARI) and influenza-like-illness (ILI) directly within the population.

To assess representativeness of GrippeWeb participants, key demographic variables and life-time prevalence of asthma and diabetes were compared to data from the German general population. To “validate” GrippeWeb, we (1) compared weekly ARI and medically attended ARI (MAARI) rates, generated between week 35/2011 and week 34/2012, with AGI MAARI rates and (2) overlaid GrippeWeb ILI rates with the number of positive influenza samples obtained by the AGI. GrippeWeb experienced very high proportion of adherence (62 % of participants reported in at least 90 % of possible weeks since registration). Age distribution, including children, but with the exception of the elderly, and background chronic conditions of GrippeWeb participants compared reasonably well with the German general population.
Varying by age group participants reported between 1.4 and 6.2 ARI and between 0.1 and 2.3 ILI during the year.

Both, the GrippeWeb ARI incidence and the GrippeWeb MAARI incidence correlate significantly with the AGI MAARI incidence (Pearson correlation coefficient = 0.80 $p< 0.001$, 95% confidence interval (CI): 0.68, 0.88 and Pearson correlation coefficient = 0.89, $p< 0.001$, 95% CI: 0.82, 0.94). The correlation could be improved up to 0.89 by using a lag of two weeks for the correlation of GrippeWeb ARI incidence and AGI MAARI incidence.

Superimposing GrippeWeb ILI rates with the number of influenza virus confirmations in the AGI demonstrates that the influenza wave is reflected in the ILI rates of both adults and children.

The presented results underline that GrippeWeb already became a reliable monitoring system for ARI and ILI shortly after implementation. The high degree of agreement between GrippeWeb’s and AGI MAARI data lends support to the validity of both systems.

Authors:
C. Bayer (1, 2), C. Remschmidt (1), M. an der Heiden (1), K. Tolksdorf (1), M. Herzhoff (1), S. Kaersten (1), S. Buda (1), W. Haas (1), U. Buchholz (1) Affiliations:

1. Robert Koch Institute, Department for Infectious Disease Epidemiology, Berlin, Germany
2. Postgraduate Training for Applied Epidemiology (PAE, German FETP) and European Programme for Intervention Epidemiology Training (EPIET)

P. Bajardi
Organization/Institution name: Università degli Studi di Torino
Department: Dept. Veterinaria

Title: *Effect of recruitment methods on attrition in internet-based studies*

Abstract:

Internet-based systems for epidemiological studies have advantages over traditional approaches as they can potentially recruit and monitor a wider range of individuals in a relatively inexpensive fashion. We studied the association between communication strategies used for recruitment (off-line, on-line, face-to-face) and follow-up participation in nine internet-based cohorts: the Influenzanet network of platforms for influenza surveillance which includes seven cohorts in seven different European countries, the Italian birth cohort Ninfea and the New Zealander birth cohort ELF. Follow-up participation varied from 43% to 89% depending on the cohort.

Participants who became aware of the study through an on-line communication campaign compared with those through traditional off-line media had a lower follow-up participation in 8 out of 9 cohorts (overall odds ratio: 0.74, 95% confidence interval: 0.54, 1.01). We did not find differences in participation proportion between participants enrolled face-to-face and those enrolled through other off-line strategies.

An internet-based campaign for internet-based epidemiological studies seems to be less effective than an off-line one in enrolling volunteers who keep participating to follow-up questionnaires. This suggests that, also for internet-based epidemiological studies, an offline enrollment campaign would be helpful in order to achieve a higher participation proportion limiting the number of people lost to follow up.

Author(s):

Paolo Bajardi [1,2], Daniela Paolotti [3], Alessandro Vespignani [4,5,3], Ken Eames [6], Sebastian Funk [6], John Edmunds [6], Clement Turbelin [5], Marion Debin [7,8], Vittoria Colizza [7,8,3,] Ronald Smallenburg [9], Carl Koppeschaar [9], Ana O. Franco [10], Vitor Faustino [10], AnnaSara Carnahan [11], Moa Rehn [11], Franco Merletti [12], Jeroen Douwes [13], Ridvan Firestone [13], Lorenzo Richiardi [12]

[1] GECO- Computational Epidemiology Group, Department of Veterinary Sciences, University of Torino, Italy,
[2] Molecular Biotechnology Center CSU - Complex Systems Unit, University of Torino, Italy,
[3] ISI Foundation, Turin, Italy
[4] Laboratory for the Modeling of Biological and Socio-technical Systems Northeastern University, Boston, USA,
[5] Institute for Quantitative Social Sciences at Harvard University, Cambridge, USA,
[6] London School of Hygiene and Tropical Medicine, Great Britain,
[7] INSERM UMR 707,
[8] UPMC Université Paris 06, Faculté de Médecine Pierre et Marie Curie, UMR S 707, Paris, France,
[9] Aquisto-Inter BV, Amsterdam, The Netherlands,
[10] Instituto Gulbenkian de Ciencia, Oeiras, Portugal,
[11] Smittskyddsinstituet, Stockholm, Sweden,
[12] Department of Medical Sciences, University of Turin and CPO-Piemonte,
[13] Centre for Public Health Research, Massey University.

V. Colizza
Organization/Institution name: INSERM, Université Pierre et Marie Curie, Paris, France & ISI Foundation
Title: Who’s your citizen scientist? The European participatory flu surveillance experience

SESSION - Needs, Tools and Methods for Digital Epidemiology

M. da Silva
Organization/Institution name: Lasige, FCUL
Department: Informatics Department

Title: *Epidemic Marketplace - An information platform for epidemiological data*

Abstract:

Epidemiology is a data-driven science, requiring access to large amounts of data. Such data is abundant:there is now a "data tsunami" generated by social networks, participatory web platforms, blogs, together with more specific data from emerging internet based surveillance systems and epidemics simulations. All of these resources are of potential use to epidemiological research, but to be used they must be maintained, locatable and obtainable.
The Epidemic Marketplace (EM) - available at http://www.epimarketplace.net (1) - is an information platform which tackles these challenges by enabling storage, management and sharing, while protecting epidemiological resources. The EM acts as a repository of epidemiological related resources, which enables users to search available data and upload their own resources.
To support resource discovery and data reuse, the EM makes use of semantic web technologies for the characterization of epidemic resources using metadata terms from a network of epidemiological ontologies. To cope with the privacy of

sensitive resources and promote trust in the repository, resource owners can autonomously manage their own resources and assign discretionary permissions to them. Permissions can be assigned both at the metadata and data levels, enabling users to make their resources visible while protecting sensitive data. Additionally, the repository fosters collaboration enabling users to share or request additional resources, and also comment on previously shared resources.

D. Medini
Novartis Vaccines and Diagnostics, Siena (I)

Title:
Data Challenges for the Novel 4CMenB Vaccine.

Abstract:

The prevention of invasive disease caused by capsular group B meningococci (MenB) is a long-recognized unmet medical need, with low incidence – requiring the use of correlates of protection in clinical studies – but high case fatality rate, particularly in infants. A novel protein-based vaccine named 4CMenB was recently approved for use by the European Medicines Agency.

Meningococcal surface proteins have inherent variability that derives from many mechanisms, including complex and dynamic population structure of the pathogen, recombination and phase variation. The potential effect of a protein-based vaccine on circulating MenB bacteria, therefore, depends on the extent to which the MenB antigens cross react with those present in the vaccine, and the degree of expression of each antigen on the cell surface.

MATS, a meningococcal antigen typing system dependent on both the extent of immunologic recognition and the level of expression of the antigen, was developed to surrogate the established correlate of protection, predicting whether a given MenB isolate would be susceptible to killing by post-vaccination antibodies induced by 4CMenB.
An independent study against representative MenB disease isolates from England and Wales showed that MATS underestimates 4CMenB strain coverage in relevant epidemiological settings.

MATS was transferred and standardized in 7 reference laboratories worldwide. Epidemiological collections of MenB isolates from 12 countries were typed, predicting 66% to 91% strain coverage for the 4CMenB vaccine on a country basis. MATS and genetic data were also compared to correlate vaccine typing with the pathogen population structure.
The potential use of MATS as a relevant tool for post-implementation surveillance is briefly discussed.

A. Ecseki
Organization/Institution name: Mediconcept Ltd.
Department: Research and Data Analysis

Title*: Healthcare databases in Hungary; present and future*

Abstract:

Due to the global economic crisis, most of the actions of the Hungarian government are focused on cost reductions, including in the health care sector. According to the current changes the most important objective is to create a well-monitored and centralized health care system.

The Secretary of State for Health Care of the Hungarian Ministry of National Resources has prepared a reform plan under the title „Revived health care, recovering Hungary – Semmelweis Plan to save health care". The health policy measure aims to establish a sustainable rearrangement of the health care system, while maintaining the uniform social insurance model.

According to this Plan and based on the current changes the government devotes special attention to create a well-monitored and centralized health care system.

The aim of our research is to present the current situation in Hungary (main databases, like national database, register data) and the main changes and developments in our information system. Based on the current plans a sectoral portal would be created, which would serve as legally authentic source of sectoral master data and would be a good bases to establish a uniform system of sectoral reports.

These changes are crucial for our health care system. The health monitoring and needs-based capacity planning system will integrate the aggregated data that are generated in the sector, and would like to make them analyzable through data-mining and GIS applications.

E. Sévin
Organization/Institution name: EpiConcept
Department: Web Information Systems

Title:
Developing opensource solutions for mobile data collection: example of a hospital-based cohort

Abstract:

Background
****
In 2011 a cohort study named ELFE (http://www.elfe-france.fr/index.php/en/) was started by the National Institute for Demographic Studies (INED), the National Institute for Health and Medical Research (INSERM) and the French National Blood Service (EFS) to analyze children life over 20 years covering health, social sciences and environmental health aspects. Four questionnaires were designed to collect a total of 300 variables.

EpiConcept developed the SICOMAT (Information System for Maternity data Collection) to manage the data collection over 20 000 newborns directly at mothers bedside.

Objectives
****
The SICOMAT had to include: 1/ Offline mode to avoid newborns exposure to 3G or WIFI and lack of internet connectivity in some maternity 2/ High security issues during data upload because of confidential personal health data 3/ User friendly system to be acceptable for maternity staff 4/ High availability characteristics to ensure availability during data collection periods. 5/ Auto-update of the questionnaires in case of structure changes over the collection period.

Material & methods
****
We used open source software for the whole project from the operating system (Ubuntu) to the data collection tool (Voozanoo) for electronic questionnaires.
We installed Voozanoo on each netbook to keep the system offline and on a central server to collect uploaded data from netbooks.
To insure both data collection system update and data upload to the central server, the Ubuntu built-in update system was exploited.
Each netbook was identified and we developed a unique token exchange system to increase the security.
We used A flash card local backup system to increase data integrity.
The scientist team developed questionnaires in Voozanoo.
Each newborn had a unique serie of four anonymous identifiers associated to each questionnaire to prevent obvious questionnaire matching.

Results
****
In 2011, 700 netbooks were distributed to 344 maternity units. Data were collected in periods of six days in April, July, September, December.
Up to date, 18 316 newborns were included in the cohort without any data loss. The recovery system had to be called for two netbooks and worked finely.
The scientific team revised some questionnaires adjustments between data collection periods. The ubuntu/voozanoo update system worked seamlessly for the operators.

Conclusion
****
No data were lost during the period indicating that the association of ubuntu/voozanoo was efficient and may be reproduced for other data collection campaigns.
Including a "ready to install" image containing both voozanoo and ubuntu system could reduce the implication of the scientific team in the technical aspects of questionnaire design and data collection. This will allow them to focus on the study design aspects.

SESSION - Modeling

C. Scoglio

Organization/Institution name: Kansas State University
Department: Electrical and Computer Engineering

Title: *Individual-based Information Dissemination in Multi-layer Epidemic Modeling*

Abstract:

In the Susceptible-Alert-Infected-Susceptible (SAIS) model, individuals observe the health status of their neighbors in the contact network. Upon observing infection, susceptible individuals might adopt a cautious behavior to reduce their infection rate. The alertness generates an increment of the threshold and consequently an increase of the range of infection strengths for which the epidemics die out. Built upon the SAIS model, we investigate how information dissemination can help boosting the public health of the population against the spreading.

The information dissemination can be realized through an additional network among individuals, such as an online social network, which has the same nodes (individuals) but different links with respect to the contact network. Information links provide the health status of one individual to all the individuals she/he is connected to. We introduce an information dissemination index which is a quadratic form of the adjacency matrix of the information dissemination network and the dominant eigenvector of the adjacency matrix of the contact graph. In this paper, we show that the spectral centrality of the nodes and edges determines the optimal information dissemination network.

Our results suggest that monitoring the health status of a small subgroup of individuals and circulating the information about their infection status can greatly enhance the resilience of the society against infectious diseases, by increasing further the die-out threshold. We have extensively tested our results on a survey–based contact network for a rural community in Kansas.

Author(s):
Faryad Darabi Sahneh, Fahmida N. Chowdhury, Caterina M. Scoglio

E. Yoneki

Organization/Institution name: University of Cambridge
Department: Computer Laboratory

Title: *Influential Neighbours Selection for Information Diffusion*

Abstract:

The problem of maximizing information diffusion through a network is a topic of considerable recent interest. A conventional problem is to select a set of any arbitrary k nodes as the initial influenced nodes so that they can effectively disseminate the information to the rest of the network.

However, this model is usually unrealistic in online social networks since we cannot typically choose arbitrary nodes in the network as the initial influenced nodes.

From the point of view of an individual user who wants to spread information as much as possible, a more reasonable model is to try to initially share the information with only some of its neighbours rather than a set of any arbitrary nodes;but how can these neighbours be effectively chosen?

We empirically study how to design more effective neighbour’s selection strategies to maximize information diffusion. Our experimental results through intensive simulation on several real-world network topologies show that an effective neighbour’s selection strategy is to use node degree information for short-term propagation while a naive random selection is also adequate for long-term propagation to cover more than half of a network. We also discuss the effects of the number of initial activated neighbours. If we particularly select th highest degree nodes as initial activated neighbours, the number of initial activated neighbours is not an important factor at least for long-term propagation of information.

This work has many issues to share with a vaccination strategy to prevent the spread of disease in efficient manner.

POSTERS

P. Graça
Organization/Institution name: LaSIGE/ Faculty of Sciences, University of Lisbon
Department: Informatics

Title*: A Group-based Access Control Model for Protected Scientific Data*

Abstract:
Scientists only share sensitive information with access restrictions if they trust the data privacy and protection mechanisms available in the repository and if they can easily control who can access their information.

Scientific repositories typically use Role-Based Access Control Models, with which authorizations are centrally defined without user intervention. Some of them use more decentralized approaches with group-based discretionary access control (DAC) models, where resource owners can define who can access their resources, using centrally-defined groups. However, highly dynamic collaboration settings cannot be satisfied with predefined groups.

To address collaborative environments, we propose a decentralized group-based DAC model where users can define the authorizations for their resources as well as the groups who use them. To simplify group definitions, avoiding enumerating all user groups, our application supports dynamic groups, where user membership to a group is defined by a rule. In addition, taking into account the increased use of social networks in collaborative systems, we also support the definition of dynamic groups with social network information. Open Social is a public specification that presents a set of APIs for web-based applications for trading attributes with the Social Networks, which we explore for simplifying the specification of access restrictions. We are leveraging on the observation that the new generation of epidemiology researchers tends to form interest groups in social networks.

We outline this model and its implementation in the EM, a platform for integrating and sharing epidemiological data. The system relies on a Fedora Commons server as the data repository, a LDAP server to store user information and user-created groups, and XACML to define authorizations, which are stored in the Fedora repository. LDAP groups can be extended to include Open Social access control attributes.

F. Couto
Organization/Institution name: University of Lisbon
Department: LASIGE - Large-Scale Informatics Systems Laboratory

Title*: Semi-automated Annotation of Epidemiological Resources*

Abstract

Epidemiology is by nature a data-intensive research field, combining results from empirical, analytical and simulation studies. These would benefit from the new scientific methodology designated as the fourth paradigm, which addresses the challenges raised from our need to validate, analyze, visualize, store, and curate the large amounts of generated data [1]. In this context, we developed the Epidemic Marketplace (EM) [2] (available at www.epimarketplace.net) , a platform that enables the sharing of resources and knowledge within the Epidemiology community with a strong focus on the semantic annotation of resources.

To support the browsing and management of the resources stored in the EM, each one is described with a set of metadata elements providing biological information (e.g.: disease, symptom, host, vaccine, vector), geographical information, environment and socio-economic conditions, demographics and the associated time frame. To promote a precise and consistent characterisation, these metadata elements are described with well-defined terms from NERO (Network of Epidemiology Related Ontologies), a collection of existing ontologies and vocabularies with the purpose of covering the epidemiological domain [3, 4]. On uploading their resources to the EM, users can provide an accurate semantic annotation based on the metadata and NERO ontologies. However, this process can be time consuming, since a single epidemiological resource can refer to several diseases, symptoms, locations, etc.

In order to require minimal human intervention, we will develop a semi-automated annotation module for the EM, which automatically identifies terms of NERO in a given text-based resource and suggests them as a default characterization of the resource. This module will be based in the recognition of epidemiologically relevant terms in the text and their resolution, i.e., mapping, to NERO terms. Our proposal is to adapt the machine learning approach and semantic similarity techniques described by Grego et al. [5] to perform the recognition and resolution tasks. The machine learning approach requires a training corpus which we will create from the over 100 fully annotated resources already available in the EM. Semantic similarity techniques are normally used with single domain ontologies, but the multidisciplinary nature of NERO will require novel techniques which we intend to adapt from previous work [6].

The integration of this module in the EM will be able to automatically generate candidate semantic annotations to be later validated by the users. This will cut the time and effort needed to provide a complete semantic annotation for research papers and other text-based epidemiological resources, effectively encouraging users to contribute with more

resources and provide richer annotations.

A. Jarynowski
Organization/Institution name: Jagiellonian University
Department: Department of Theory of Complex System

Abstract:

The aim of this work is to supplement the knowledge of the spread of sexually transmitted diseases through computer simulations. The model has aggregated the most important paths of HPV infections and development of cervical cancer in demographic changing Polish society. The main goal is the authoritative analysis of the potential epidemiological control strategies and their impact on future situation during 25 years. Constructed model indicates an effective tool against cancer (vaccination with screening). It also alarm authorities, that some danger processes (taking place in Poland) like aging of society and increase of sexual activity could recall epidemic if prevention would not act properly.

V. Dorgali
Organization/Institution name: Università di Firenze

Title: *Kicking against the pricks: vaccine skeptics have a different social orientation*

"Background: In any country, part of the population is skeptical about the utility of vaccination. In order to develop successful vaccination programmes, it is important to study and understand the defining characteristics of vaccine skeptics. Research till now mainly focused either on the underlying motives of vaccine refusal, or on socio-demographic differences between vaccine skeptics and non-skeptics. It remained till now unexplored whether both groups differ in terms of basic psychological dispositions.
Methods: We held a population survey in a representative sample of the population in Flanders, Belgium (N=1050) in which we investigated whether respondents' attitude to vaccination wasassociated with their basic disposition towards other community members or society in general, as measured by the Triandis and Gelfand social orientation scale.
Results: We found that skeptics and non-skeptics have a different social orientation, even when several variables are controlled for. More specifically, vaccine skeptics scored significantly lower on both horizontal individualism and horizontal collectivism, indicating a lower disposition to see others as equals.
Conclusion: These findings need confirmation in the context of different countries. Such insights can be valuable to optimize the design of effective communication strategies on vaccination programmes. "

Author(s):
1)Jeroen Luyten
Centre for Health Economics Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
Institute of Philosophy, Catholic University of Leuven, Kardinaal Mercierplein 2 bus 3200, 3000 Leuven, Belgium
Jeroen.luyten@ua.ac.be

2) Pieter Desmet Erasmus School of Law, Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands desmet@law.eur.nl

3) Maria Veronica Dorgali Department of Statistics, Informatics, Applications (DISIA), University of Florence, 'G. Parenti', viale Morgagni 59, 50134 Firenze, Italy & Department of Economics and Management, University of Pisa, Via Cosimo Ridolfi 10, 56124 Pisa, Italy Veronica.Dorgali@gmail.com

4)Niel Hens Centre for Health Economics Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium Niel.hens@uhasselt.be
Center for Statistics, Hasselt University, 3590 Diepenbeek, Belgium

5) Philippe Beutels Centre for Health Economics Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium & School of Public Health and Community Medicine, The University of New South Wales Sydney, Australia Philippe.beutels@ua.ac.be

D. Paolotti
Organization/Institution name: Fondazione ISI, Italy

Title: *Determinants of participation in the Influenzanet Cohort*

Influenzanet is a network of internet-based platforms aimed at influenza surveillance collecting real-time data in seven European countries: Sweden, United Kingdom, the Netherlands, Belgium, France, Italy and Portugal. More than 30000 European volunteers participate every year in the study, representing one of the largest existing internet-based multicenter cohorts. Focusing on the influenza season 2011-12 as the first season where the Influenzanet system is completely harmonized within a common framework among all the seven countries, we investigated the propensity of users to regularly come back to the platform to provide information about their health status. The purpose of this work is to investigate socio-demographic factors associated with the active participation at follow-up. By means of a multilevel logistic regression we found a lower participation at follow-up associated with lower educational status (OR=0.80, 95 %C.I.=0.75-0.85), smoking behavior (OR=0.64 95%C.I.=0.59-0.70), younger age (OR ranging from 0.3 to 0.7), not being vaccinated against seasonal influenza (OR=0.77 95%C.I.=0.72-84) and living in a household with no children (OR=0.69 95%C.I.=0.65-0.74). The crude and adjusted odds ratios (OR) show that most of these results are still valid when single countries are analyzed separately.

Author(s):
Paolo Bajardi [1,2], Lorenzo Richiardi[3], Alessandro Vespignani 1,4, Sebastian Funk 5, Ken Eames 5, John Edmunds 5, Clement Turbelin 6, Marion Debin 6, Vittoria Colizza 6, Ronald Smallenburg 7, Carl Koppeschaar 7, Ana Franco 8, Vitor F austino 8, Annasara Carnahan 9, Moa Rehn 9 , Daniela Paolotti 1

[1] ISI Foundation, Turin, Italy, Cancer Epidemiology Unit,
[2] GECO- Computational Epidemiology Group, Department of Veterinary Sciences, University of Torino, Italy
[3] CPO-Piemonte and University of Torino, Italy
[4] Northeastern University, Boston, USA,
[5] London School of Hygiene and Tropical Medicine, Great Britain,
[6] IU707, INSERM UPMC Univ Paris 06, Paris, France,
[7] Aquisto-Inter BV, Amsterdam, The Netherlands,
[8] Instituto Gulbenkian de Ciencia, Oeiras, Portugal,
[9] Smittskyddsinstituet, Stockholm, Sweden.

C. Kopperschaar

Organization/Institution name: Science in Action, The Netherlands

Title: *Determinants of participation in the Influenzanet Cohort*