Background

• Tremendous threats of emerging and re-emerging epidemics

• Public awareness of infectious diseases plays an important role in disease control

• Need for epidemic preparedness

• Key tasks in dealing with infectious diseases
  → Surveillance – regional, national, and global
  → Investigation & early control
  → Prevention measures
  → Monitoring and evaluation

• Concerns for traditional passive systems
  → Lack of infrastructure and high management costs
  → Delays between events and notifications
  → Poor information flow esp. in remote areas
• Development of information communication technology methods in public health

• The Internet → efficient health-related communication and data collection
  • Online news stories, social network data, blog/microblog data
  • Integration in disease surveillance and control

• Rise of various ICT interventions and Internet-based systems as a vital tool in the battle against infectious diseases
**Objective I** → investigate web-based infectious disease surveillance systems that focus on disease occurrence and the early detection of outbreaks

**Objective II** → serve as an overview and reference for prospective infectious disease surveillance systems
Key words in search process including using newly coined terms:

**NEW PUBLIC HEALTH TERMS**

**BIOSURVEILLANCE, INFOVEILLANCE, INFODEMIOLOGY**
Indicator-based surveillance

- Credible information but slow reporting
- Mainly for known diseases
- Not well established in some areas

Event-based surveillance

- Rapid detection of public health events
- Complements traditional surveillance
- For threats with unknown origin

Data

- Risk assessment
- Collect, analyze, interpret
- Risk management
- Healthcare facilities
- Informal channels (media)

Events

- Early warning
- Capture, filter, verify
- Response

Capture, filter, verify
The 11 Surveillance Systems

Event-based surveillance systems

Moderated Systems
- ProMED-Mail
- Global Public Health Intelligence Network (GPHIN)
- Global Outbreak Alert and Response Network (GOARN)

Automated Systems
- MedISys
- EpiSimS
- BioCaster
- EpiSPIDER
- HealthMap
- GET WELL

News Aggregators
- Google Flu Trends
- Influenzanet
Moderated Systems

- Information is processed by human analysts or first processed automatically and analyzed by people → *Automatic + Manual*

- Screen for epidemiological relevance of the data found within the information before dissemination

| Moderated Systems | Origin (Year started) | Area of Service | Language | Data Source | Data access | Format               |
|-------------------|-----------------------|-----------------|----------|-------------|-------------|----------------------|
| ProMED-mail       | USA (1994)            | Worldwide       | 7 languages | News/Media reports | Public     | E-mail alert         |
| GPHIN             | Canada (1997)         | Worldwide       | 8 languages | News/Media reports | Restricted | Website/e-mail alert |
| GOARN             | WHO, UN (2000)        | Worldwide       | English, multilingual | News/media reports | Restricted | Network-based        |
Program for Monitoring Emerging Diseases (ProMED-Mail)

- A widely acknowledged email service to identify unusual health events related to emerging and re-emerging infectious diseases and toxins by public health experts
- “One Medicine” concept
- Disseminates information via its website and through social media channels (Twitter and Facebook) and RSS feeds.
- First to report on numerous major and minor disease outbreaks including SARS, MERS, Ebola and the early spread of Zika.

FLOW OF INFORMATION

PROCESS OF PROMED-MAIL

WHO

Media

Health worker

Mins of Health

Local health official

Lay public

Receipt of information

Internal review & verification

Dissemination
Global Public Health Intelligence Network (GPHIN)

- A secure, restricted access system for outbreak alert
- Gathers information by monitoring global media sources on a 24/7 basis
- Primary sources of outbreak information → Factiva and Al Bawaba
- Scans six key areas for news of outbreaks: infectious diseases, biologics, and chemical, environmental, radioactive and natural disasters
- Supplies a significant amount of the WHO's early warning outbreak information; WHO verification

**PROCESS OF GPHIN**

Data collection (National/local news & media report) → Data processing & analyst assessment (automated + manual) → Intelligence turned into alert notifications

Information on public health risk → Intelligence on public health risk
Automated Systems

- More complex data collection than moderated systems
- Differs in the level of analysis performed, scope of information sources, language coverage, speed of information delivery and visualization methods

| Automated Systems | Origin (Year started) | Area of Service | Language | Data Source | Data access | Format |
|-------------------|-----------------------|-----------------|----------|-------------|-------------|--------|
| MedISys           | EU (2004)             | EU member states| 43 languages | News/media report, Europe Media Monitor | Limited to EU states | RSS feeds/e-mail, SMS alerts |
| EpiSimS           | USA (2005)            | United States   | English | US Census data Transformation infrastructure data | Restricted | Geographical interface/XML-based format |
| BioCaster         | Japan (2006)          | Priority to Asia-Pacific regions | 8 languages | Query | Public | RSS feed |
| EpiSPIDER         | USA (2006)            | N. America, Europe, Asia, Australia | English | News/media report, social media, CIA | Public | Timeline visualization/RSS feeds |
| HealthMap         | USA (2006)            | Worldwide       | 9 languages | Query, News/media report, ProMED, WHO, EuroSurveillance | Public | Mapping, RSS feed |
| GET WELL          | Sweden (2010)         | Sweden          | Swedish | Query | Restricted | Time-series graph, HTML |
**BioCaster**

- Ontology-based text mining system for detecting and tracking the distribution of infectious disease outbreaks from linguistic signals on the web

- Analyzes documents reported from RSS feeds, classifies them for topical relevance and plots them onto a Google Map using geocoded information

**Specific advantages:**
- Text mining techniques within a single system
- Text-level recognition of severity indicators
- Ontology-based inference \( \rightarrow \) fills in the gaps between symptoms and diseases
- Direct knowledge of term equivalence within and across languages
HealthMap

• Information and alert disseminated on a geographical map

• Key objective is to maximize flexibility in
  → user interface
  → collection of the underlying data

• Data organized across different dimensions (date, location and disease); view customized according to the geographic location, disease and type of outbreak.

• Primary focus on human disease surveillance

• Mainly relies on Google News and WHO news feeds
## News Aggregators

- Collect articles several from sources – commonly filtered by language or country
- Easy access to sources through a common portal – each article examined individually

| News Aggregators/ Others | Origin (Year started) | Area of Service | Language | Data Source | Data access | Format |
|--------------------------|-----------------------|-----------------|----------|-------------|-------------|--------|
| **Google Flu Trends**    | USA (2008)            | 28 countries    | 39 languages | Query, CDC | Public      | Mapping features, RSS feed |
| **Influenzanet**         | Europe (2008)         | 11 European countries | 10 languages |Self-report from volunteer (online questionnaire response) | Public | RSS feed |

![Google Flu Trends](image1.png)

![Influenzanet](image2.png)
Google Flu Trends

- Specific focus on seasonal influenza epidemics

- Use of internet search queries to detect the occurrence of influenza like illness (ILI)
  - Intersection among frequently occurring search queries in all US regions

- Anonymized and aggregated data

- Corresponding ILI rates from the CDC initially but forecasting capabilities in doubt
  - Estimates two times larger than CDC

→ Closure of the system
Discussion

**STRENGTHS AND CHALLENGES OF WEB-BASED SURVEILLANCE SYSTEMS**

- Intuitive, adaptable and cost-effective
- Automated and rapid collection of “big data” from abundant sources
- Provision of “real-time” detection and improved early notification of localized outbreaks
- System based on web queries have similar underlying mechanisms, applied to various infectious diseases (once established).
- Discrepancies between internet availability and seeking of health information – unequal use and limited access
- Unstructured nature of data sources
  - Crowd-source data
- Accuracy issues and biased reporting
- Privacy concern
• Expansion of more web and mobile-based disease surveillance systems in universities and research setting – likely outcome
  → the first systems initiated as pilot trials at the exploratory level and often based at or in corporation with universities or institutions, non-governmental organizations

• Trend → a general improvement in language coverage, service area and scope of data sources

• Merge of epidemiologic intelligence and ICT → newly coined terms reflect core functions of web-based surveillance systems and characteristics of data used

• Complementary aid to traditional surveillance rather than alternative
Considerations for Prospective Disease Surveillance Systems

1. Primary language in operation – compatibility issue
   - Language-related intricacies

2. Standardization
   - Universal access
   - Efficiency at low cost

3. Priority on timeliness
   - Enhanced preparedness and quick response to emergency

4. Better use and control of data in large quantity
Key Concept

What kind of raw data is used

How that data is used
→ Maximized use of various data sources

How to reuse data from surveillance systems for alert and further control of epidemics

DATA is the main ingredient in disease surveillance
THANK YOU