European Food Safety Authority

EFSA Journal 2009; 7(10):1355

TECHNICAL REPORT OF EFSA

Development of web monitoring systems for the detection of emerging risks[1]

European Food Safety Authority[2, 3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

The Emerging Risks Unit (EMRISK) has the mandate to assess and develop IT tools to detect and monitor food and feed emerging risks. For media-scanning systems, it has received a request to evaluate the Medical Information System (MedISys) developed by the Joint Research Centre (JRC). This report describes the MedISys system (i.e. history, terminologies, and ways of processing information) and presents data on the evaluation of the system as a monitoring and early detection tool for food and feed hazards. The performance of MedISys was assessed through its timeliness to report specific hazards (i.e. *Salmonella*, mycotoxins, heavy metals, aluminium, dioxins, melamine, DDT, *Vibrio cholera*, and radioactivity) selected from the Rapid Alert System for Food and Feed (RASFF). MedISys timeliness was also compared to ProMED-mail, another internet-scanning system. This initial evaluation suggests MedISys is an efficient tool for the early detection and monitoring of hazards. It tended to detect food and feed hazards earlier, a few days to several weeks in advance, and more frequently than ProMED-mail, but the information retrieved on each system was usually complementary. It is recommended that MedISys is further developed by adding more specific and multi-lingual food and feed categories and new sources (blogs in general and sources specifically from southern hemisphere countries) to increase its sensitivity to food and feed hazards. To add new feed and food categories, existing food databases such as the RASFF should be considered. To refine and set up the categories, linguistic and technical support should be provided.

KEY WORDS

MedISys, ProMED-mail, RASFF, media monitoring, internet-scanning, timeliness, emerging risk

1 On request of EFSA, Question No EFSA-Q-2009-0490, issued on 15 October 2009.

2 Correspondence: emrisk@efsa.europa.eu

3 Acknowledgement: EFSA wishes to thank: Jenya Belyaeva, Monica Gemo, and Jens Linge (EC-JRC, IPSC) for the support provided to this EFSA scientific output.

For citation purposes: European Food Safety Authority; Development of web monitoring systems for the detection of emerging risks. EFSA Journal 2009; 7(10):1355. [50 pp.]. doi:10.2903/j.efsa.2009.1355. Available online: www.efsa.europa.eu

SUMMARY

The Emerging Risks Unit is mandated to assess and develop IT tools to detect food and feed emerging risks. In particular, it is requested to assess the efficiency of web-scanning systems such as the Medical Information System (MedISys), an application of Europe Media Monitor (EMM) developed by the Joint Research Centre (JRC).

This report describes the MedISys system (history, terminologies and ways of processing information), and presents data on the evaluation of MedISys as a monitoring and early detection tool for food and feed hazards. First, a retrospective approach based on several case studies selected from the Rapid Alert System for Food and Feed hazards (RASFF) over a two-month period (between January 2007 and March 2009) was used. These hazards (i.e. salmonella, mycotoxins, heavy metals, aluminium, dioxins, melamine, DDT, *Vibrio cholera* and radioactivity) were traced back on MedISys and compared to ProMED-mail, another web-based system which disseminates information on outbreaks of infectious diseases and acute exposures to toxins that affect human health. Second, food and feed-borne hazards detected on MedISys, over a three months-period (January to March 2009), during weekly monitoring, were traced back on the RASFF and ProMED-mail.

MedISys proved to be an efficient tool for the monitoring and early detection of food and feed hazards. On several occasions, it reported hazards before ProMED-mail and RASFF. The time delay with ProMED-mail was usually small (a couple of days) and probably related to the time required by experts to assess the value of the information, whereas the time delay with RASFF was usually longer (a couple of days to several weeks) and possibly linked to the time for laboratories to confirm the data and/or to the time for hazards which emerged outside EU to reach EU borders and markets. However, although the web-scanning systems, MedISys and ProMED-mail, have different time delays in reporting hazards, the information they deliver is usually complementary.

The analysis of the news sources and categories processed by MedISys showed that the system needs further customization to increase its sensitivity to food and feed hazards. In particular, news sources and media coverage need to be expanded (i.e. more sources from southern hemisphere countries, and more sources and blogs related to food and feed areas). Existing food and feed databases such as the RASFF could be used to make a selection of hazard and product categories. However, technical support is still required to define these categories (i.e. proper list of words and patterns to reduce noise, and multi-lingual translations). Moreover, their access on the MedISys portal needs to be implemented to increase the efficiency of monitoring activities on MedISys. In conclusion, when the system is better customized with new sources and categories, more resources are required to conduct daily monitoring and store the information for later analysis.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

The mission and tasks of EFSA are described in Regulation (EC) No 178/2002 and include the responsibility to set up systems for identifying emerging risks. "The Authority shall establish monitoring procedures for systematically searching for, collecting, collating and analysing information and data with a view to the identification of emerging risks in the field of its mission" (art. 34.1).

Among the various systems (EMM, ProMED, PubMED, RASFF, EUROSTAT, TRADE databases, etc.) that have been identified as potential tools for monitoring emerging risks, the EFSA's Scientific Committee suggested to test the Europe Media Monitor system (EMM-MedISys) developed by the Joint Research Centre (JRC) (EFSA/EMRISK/029, 2009).
This system needs to be customized and automated, before it can be further used by EFSA. Preliminary assessments made by the Emerging Risks Unit show that the EMM system performs well as an information and early warning tool. Further study cases are currently under investigation to better determine the efficiency of the system.

TERMS OF REFERENCE AS PROVIDED BY EFSA

In the process of establishing automated monitoring devices and procedures to identify emerging risks in food and feed, the Emerging Risks Unit is requested to carry out the preparatory work on the EMM[4] as follow:

1. Set up a system for downloading and analyzing the EMM system for the purpose of detecting signals of emerging risks:

- Define a specific set of categories and keywords on the advanced search tool (Rapid News Service - RNS) of the MedISys portal to increase the specificity of the searches on emerging risks in the food and feed sectors.
- Screen the information on the EMM in search of signals of emerging risks.
- Report and assess the signals found on the EMM in task force meeting (primary filter) and disseminate the information to EFSA customers (appropriate Units and Panels) when relevant signals are detected.
- Assess the efficiency of the EMM to detect and monitor signals of emerging risks.
- Customize the EMM system to the needs of the EFSA.
- In urgent situations provide other Units and Panels with most recent available information from the EMM.

2. Other potential databases

[4] EMM contains four portals, the Medical information System (MedISys), The NewsBrief, the NewsExplorer and the EMM-Labs. The Emerging Risks Unit focused on MedISys, a real-time news alert system for medical and health-related topics, because it best covers the Unit's mandate which is on the development of tools to monitor food and feed hazards. Therefore, in the above terms of reference, the short terminology "EMM" should be read as "MedISys, an application of EMM".

- Characterize other information technology systems (e.g. ProMED-mail, PubMED, etc.) and determine their potential usefulness for the identification of emerging risks, and establish procedures for access by EFSA staff, as appropriate.

- To share expertise on emerging risks, collaboration between the Emerging Risks Unit and EFSA customers is essential. In addition, a tight collaboration between the Emerging Risks Unit and the JRC (the Institute for the Protection and Security of the Citizen) is required to adjust the EMM system to the needs of the EFSA. The ECDC, which is also currently defining new categories for searches on the EMM, should also be consulted to share its expertise with the Emerging Risks Unit.

Timeline and expected deliverables

The expected timeline for the assessment of the EMM system would be within the first half of the year 2009 and the customization by the end of 2009.

From January 2009 to the end of 2009

- Screen the information on the EMM; record and report in ad-hoc unit meetings the signals found on food and feed alerts on a weekly basis and whenever signals are identified, discuss them with the task force on a monthly basis.

- Report relevant signals identified by the primary filter to EFSA customers (i.e. appropriate Units and Panels for secondary filtering).

By May 2009

- Submit a technical report, with specific study cases that may have different media coverage in relation to their notification level in the RASFF system (rarely, moderately and highly notified) to assess the limits and outcomes of the EMM in identifying signals of emerging risks. Three categories of study-cases will be investigated: 3 food/feed incidents frequently reported (e.g. salmonella), 3 food/feed incidents rarely reported (e.g. radioactivity), and 3 food/feed incidents infrequently reported (e.g. melamine). A selection of commodities could be done based on signals coming from the RASFF, the media monitoring or EFSA units and panels and described in an internal report to be issued by the task force.

- Produce a draft handbook to present the use of the EMM in a concise and user-friendly way.

By the end of 2009

- If the outcome of the assessment made on the use of the EMM as a tool for monitoring emerging risks is positive, the system will be further developed to be routinely used by the Emerging Risks Unit, or alternatively,

- Other systems scanning scientific information (e.g. ProMED-mail, PubMED, etc.) will be investigated by the Emerging Risks Unit to detect signals from other IT systems processing web-based information.

This report covers the following preparatory work of the terms of references:

- Monitor MedISys to detect food and feed hazards (section 3.2.1)
- Report hazards found on MedISys (section 3.1.3)
- Assess MedISys efficiency (section 3.2)
- Customize MedISys (section 3.1)
- Provide support in urgent situations (sections 3.1.3 and 3.2.1)
- Assess other potential databases such as ProMED-mail (section 3.2)

1. Introduction

Currently, many public health agencies use web-based data streams to monitor infectious disease outbreaks (Grein et al., 2000; Heymann and Rodier, 2001). Such systems have proved to be efficient in detecting outbreaks early and with reduced cost and increased transparency (Wilson and Brownstein, 2009).

Several internet-scanning systems have been developed starting in 1997 with the Global Public Health Intelligence Network – GPHIN (Mykhalovskiy and Weir, 2006), and followed by many others such as the World Health Organization's of the Global Outbreak Alert and Response Network – WHO-GOARN (Heymann and Rodier, 2001), the Program for Monitoring Emerging Diseases from the International Society for Infectious Diseases - ProMED-mail (Woodall, 2001; Madoff, 2004), the HealthMap system (Friefeld et al., 2008), the Medical Information System MedISys (Steinberger et al., 2008), and the Global Health Monitor based on the BioCaster ontology (Collier et al., 2008). More recently, internet search-term technologies such as Google (Ginsberg et al., 2009) and Yahoo (Polgreen et al., 2008) have been shown to have much potential for disease surveillance, but the information they recover is usually not detailed or reliable enough to estimate relevant epidemiological parameters of incipient outbreaks (Linge et al., 2009).

With the exception of GPHIN and ProMED-mail which are human-curated, all these systems are automated and some receive notifications from ProMED-mail (e.g. WHO, GPHIN, EMM, and HealthMap). ProMED-mail, which has been extensively studied, has been demonstrated to have a very high level of accuracy (Friefeld et al., 2008) and to be efficient as an early-warning system (Mykhalovskiy and Weir, 2006). However, an important limitation of ProMED-mail, along with most other web-scanning systems, is the restricted language coverage which narrows the geographic coverage, and the risk of losing scientific information during translation (Cowen et al., 2006). The major features that distinguish the EMM from the other systems are that it covers many more languages, aggregates information across documents and languages, calculates notification levels, offers various ways to disseminate the information and in combination with the Rapid News System (RNS), it allows users to moderate the EMM input and disseminate it.

MedISys which is an application of the EMM screens information on internet to detect early warning of emerging risks in the health sector. If MedISys can be further developed to detect emerging food and feed risks, it would be highly relevant to the EFSA Emerging Risks Unit. The objectives of this report are to evaluate the work required to customize MedISys on food and feed hazard detection and to determine its performance as a monitoring and early-warning system.

2. Materials and methods

2.1. MedISys description

2.1.1. History and access

The EMM was initially developed by the JRC as an in-house application for the European Commission's Directorate General Communication (DG-COMM). To date, the EMM is the web-monitoring system that has the widest media and language coverage with an average of approximately 90,000 news articles from more than 2200 news sites in 50 languages per day (figures of July 2009). MedISys became online in August 2004 and has been continuously extended.

The EMM contains three portals, the MedISys, the NewsBrief and the NewsExplorer. MedISys (http://medusa.jrc.it) identifies in a timely manner potential threats on infectious diseases,

bioterrorism, and chemical, biological, radiological and nuclear (CBRN) in open-source media. It is fully automatic 24/7 and disseminates the information to the public health of the European Commission (Linge et al., 2009).

MedISys displays only those articles with interest to Public Health. It analyses the news and warns users with automatically generated alerts. Upon request, users will receive email or SMS messages notifying them of breaking news stories or of new articles about any of the diseases, etc. RSS feeds for all categories are available for integration in third-party environments.

On the public site http://medusa.jrc.it/, all users get access to MedISys. Specialist Public Health organisations can request access to the restricted site, which offers more functionality, more categories and more news sources. Users can select the languages and news sources they want to see. Registered users additionally get access to the EMM tool Rapid News Service (RNS), which allows them to select sources, compile newsletters and alert colleagues via email or SMS.

2.1.2. Terminologies

To better understand the way MedISys operates and processes information, the following terms need to be defined (**https://medirns.jrc.it/RNS/notificationeditor/help/help.html**):

- **News sources**: Sources comprise news portals, online editions of newspapers, official sites from competent authorities, scientific journals and specialist medical sites. MedISys processes approximately 90,000 news articles from more than 2200 news sources (news and medical sites) in 50 languages per day (figures of July 2009). New functionality to monitor blogs and tweets is under development; blog and twitter sites will then be added as sources.

- **Categories**: In MedISys, a category usually refers to a specific hazard, e.g. "AIDS-HIV" or "SwineFlu". However, one can define categories freely, e.g. a category only for "mercury" or "EFSA". A category contains keywords in several languages. These keywords are used in a pattern-matching approach. Furthermore, Boolean expressions and context-specific patterns are possible.

- **Notifications** define the characteristics that a news source must have in order to be identified and categorized.

- **Patterns** correspond to a list of simple or combined words, including a list of combinations with “must contain at least” and “but none of”, translated in several languages. Pattern matching is the act of detecting the presence of the constituents of a given pattern, e.g. in MedISys the pattern "legionell%" would match both "Legionellosis" and "Legionella pneumophila" ("%" stands for one or more non-whitespace charaters, whereas "_" stands for exactly one non-whitespace character).

- **Combinations** are a set of two or more lists of patterns. (“OR” and/or “NOT” lists).

- **Proximities** (only for combinations) describe a word context size within which the combination terms have to occur.

- **Thresholds** are the minimum sum of weights matched patterns required to trigger the notification.

- **A weight** (only word-weight list) of a pattern is the value summed when patterns are matched. The sum of the weights of the matched patterns is compared to the threshold value to determine whether or not to trigger the notification.

- **Alerts** correspond to specific hazards and diseases that are picked up by MedISys when the article matches a particular pattern. For example the pattern of the alert "food poisoning", contains the combinations "foodborne and illness, "food and infection", food and intoxication" (etc.), the "must contain at least" "food" and "poisoning%" (% for one or more non-whitespace), and the "but none of" "H1N1", "SIV", "swine and flu" (etc.). To increase the detection of relevant articles, alerts can be refined by weighting the words in the patterns list and defining a threshold value to determine whether or not to trigger the alert containing the words. In the lists of combinations, a proximity which describes a word context size within which the combination terms have to occur can be set up to increase the chance of getting more relevant articles.

2.1.3. Categorization and filtering

All news items are additionally categorized into hundreds of categories. Categories include geographic regions such as each country of the world, organizations, themes such as *diseases* or *bio-terrorism*, and more specific classes such as *swine flu*, *anthrax* or *thyroid cancer*. Articles fall into a given category if they satisfy the category definition, which consists of Boolean operators with optional vicinity operators and wild cards. Alternatively, cumulative positive or negative weights and a threshold can be used. Uppercase letters in the category definition only match uppercase words, while lowercase words in the definition match both uppercase and lowercase words. Many categories are defined with input from the institutional users themselves.

The system keeps statistics on the 14-day average number of articles falling into any given country-category combination (e.g. *Estonia-tuberculosis*). If the number of articles for this combination found in the last 24-hours (normalized by weekday fluctuations) is significantly higher than this average, a country-category-specific alert is triggered and users are notified using alert ranking graphs. Alert ranking graphs exist for the whole globe or for selected regions (e.g. European Union or South America).

As the categories are defined in many different languages (depending on the user interests, some are defined in all languages, others in only a few), the statistics are fed by all languages and are thus not language-dependent. This means that the sensitive alerting tool will detect a sudden rise in *any* of the languages, having the consequence that users may see an alert even before the event is reported in their own language. For humanitarian and public health institutions whose main concern is early warning and rapid reaction, this is a highly appreciated functionality.

Each article is geo-tagged, i.e. potential place names are identified and ambiguities are resolved. Resolving ambiguities is necessary as place names are often homographic (same spelling) with common words (e.g. there is a place called *And* in Iran and city called *Split* in Croatia), with people's names, and even with other locations. For instance, world-wide there are 15 places called *Paris*, 102 places called *San Francisco* and 195 places called *Victoria* (Pouliquen et al., 2006). An algorithm that considers the place hierarchy (city is part of a region which is part of a country) and the frequency of mentions determines the major location in a cluster. This is used to visualize the location of the current news items on a geographical map. Languages have a regional bias so that they differ significantly in the world regions they report about. The regional complementarity of the different reporting languages shows how important it is to monitor the media in different languages. Only the biggest stories will be translated into English and other languages.

Every ten minutes and in each of the languages, the application clusters the latest news items (considering each time a window of four hours or more, depending on the number of recent articles) and presents the largest cluster as the current top-ranking media theme (referred to as *Top Stories*). The title of the cluster's medoid (the article closest to the cluster centroid) is selected as the most representative title and thus as the title for the cluster. The *top stories* section of MedISys thus shows at any given moment which public health related themes are being discussed most across the large

variety of media sources world-wide. Unlike the news *categories*, which contain news about a specific disease or threat, the clusters are formed automatically and in an unbiased manner, thus covering any possible subject that may be discussed in the news. They may talk about a disease outbreak, about a health threat regarding a public event, about new highly discussed medical insights, or anything else.

2.2. MedISys evaluation

The evaluation of MedISys was conducted on the following aspects:

- Identification of ways to implement the system to increase its sensitivity to food and feed hazards (i.e. category definitions and identification of gaps in media coverage).
- Evaluation of the efficiency of MedISys as a monitoring tool by assessing the reporting time of the hazards found on MedISys and on other systems such as ProMED-mail and RASFF.
- Evaluation of the efficiency of MedISys as an early detection tool by using case studies from the RASFF database and by assessing their timeliness in MedISys and ProMED-mail.

For these evaluations, ProMED-mail, a web-scanning system, was used to be compared to MedISys. ProMEd-mail (http://www.promedmail.org) was formed in 1994 as an initiative of the Federation of American Scientists (FAS). In 1999, it became a programme of the Society for Infectious Diseases (SID) in collaboration with the Harvard School of Public Health and the Oracle Corporation. This system provides early warning of outbreaks of emerging infectious diseases and episodes of acute toxicity, and the spread of antibiotic and disease vector resistance worldwide, by e-mail, 7 days a week. It is moderated by infectious diseases specialists and comprises about 30,000 subscribers from 180 countries (Madoff, 2004; Madoff and Woodall, 2005). It focuses on rapid reporting of outbreaks rather than detailed epidemiological surveillance. Sources of information include media reports, official reports, online summaries, and local observers. Reports are available on the search archives menu, from the date of the creation of the system, in 1994, up to present time. They can be visualized with the EpiSPIDER program (Semantic Processing and integration of Distributed Electronic Resources for Epidemics [and Disasters]; www.epispider.net).

RASFF (http://ec.europa.eu/food/food/rapidnotification/index_en.htm) was created in 1979 and its legal basis is laid down by the Regulation EC/178/2002, establishing the EFSA and laying down several procedures in matters of food safety (O.J. No L 31 of 1 February 2002). RASFF scope and procedures are defined in articles 50, 51 and 52. This system is an effective tool to exchange information between Member States and the European Commission on measures taken to ensure food safety. The RASFF database includes market (notifications and information), border rejection and news notifications. Market notifications mainly come from official controls on the internal market and European border controls. They are triggered by the Member States and sent when a food or feed presents a direct or indirect risk for human health. Information notifications concern a food or feed that was placed on the market for which a risk has been identified, but for which the other Members do not have to take immediate action because the product is not yet or no longer present on their market. Border rejection notifications concern food and feed consignments that have been tested and rejected at the external borders of the EU. News notifications concern any type of information related to the safety of food and feed which has not been communicated as a notification, information or border rejection notification and which is usually based on information picked up from the media. RASFF notifications are available on the RASFF archive database (http://ec.europa.eu/food/food/rapidnotification/archive_en.htm), from 1979 up to present time. Recently, EFSA developed a system for the routine analysis of data from the RASFF to facilitate the identification of potentially relevant trends of emerging risks (EFSA, 2009c).

In the process of identifying emerging risks, indicators and signals as defined by EFSA (2006, 2007) need to be detected and followed. These indicators and signals are various and of diverse nature. For MedISys which processes information, these terms relate to specific definitions (Table 1).

Table 1: Indicators and signals of emerging risks on MedISys

	EFSA definitions	On MedISys
Indicator	It is a component of risk assessment and is comprised of a focused selection of parameters, directly or indirectly related to the food chain that can be measured/calculated qualitatively and/or quantitatively. Ideally, an indicator should be reliable, sensitive, quantifiable, and should provide the information on the nature of the hazard (agent/process involved) and the source of the risk	It is in the form of terms that refer to unique events that may relate to an emerging issue (e.g. contamination, poisoning, dead, intoxication, unknown, mystery, etc.) and that are directly or indirectly related to the food chain.
Signal	It is identified as a temporal or spatial trend in an indicator value	It is identified as a sudden occurrence and/or increase of articles matching the terms described above

To detect signals of emerging risks in food and feed, MedISys has been monitored continuously since January 2009 (weekly until June 2009 and daily since July 2009). In particular, top-ranking media themes or *top stories*, and categories related to food and water borne diseases in the *diseases*, *chemical* and *nuclear* themes were monitored. A new category *environmental food* containing information on *food poisoning*, *food* and *water contamination* which was recently added to MedISys was also monitored. When a particular signal of an emerging hazard was found on MedISys, its occurrence in ProMED-mail and the RASFF database was checked and noted by date of publication.

For the identification of ways to implement MedISys, a descriptive analysis was conducted on news sources by language and country. In addition, a preliminary linguistic analysis was conducted on three case studies (i.e. salmonella, melamine, and DDT) selected from each aforementioned level of notification. For these examples, word frequencies and proximities in "related" and "unrelated" articles were assessed to determine patterns for their definition. Articles found on MedISys were considered "related" when they described the searched hazard and contamination event, and alternatively, they were defined as "unrelated" when they described other events (e.g. articles on business, research, society, issues). These articles were picked randomly on MedISys with the advanced search option and by entering one single word (the hazard name).

A few studies which compared the performance of systems that process event-based outbreak information used **timeliness**, which is the number of days that elapses between the systems under evaluation to report the same event, as an evaluation parameter (Friefeld et al., 2008; Woodall, 2001; Cowen et al., 2006).

To determine the efficiency of MedISys as an early warning system, specific hazards were selected in the RASFF database and the time required by MedISys to report these hazards was assessed. In total, nine **case studies** of food and feed hazards were selected between January 2007 and February 2009, with high, medium and low levels of notifications (Figure 2).

- **High level**: hazards frequently reported in high numbers: e.g. salmonella, heavy metals and mycotoxins.
- **Medium level**: hazards infrequently reported in high numbers: e.g. melamine, aluminium and dioxins.

- **Low level**: hazards infrequently reported in low numbers: e.g. DDT, *Vibrio cholera*, and radioactivity.

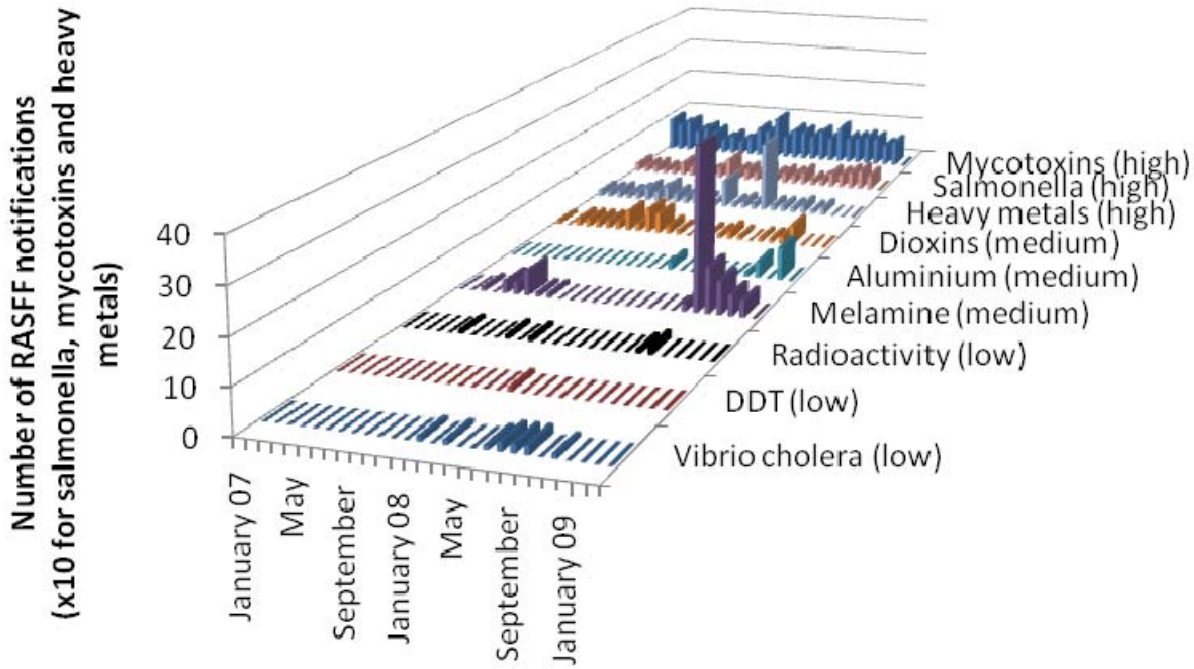


Figure 1: Number of RASFF notifications for each case study between January 2007 and February 2009.

The hazards found in RASFF were systematically searched in the media-internet scanning systems, MedISys and ProMED-mail. Searches were conducted with the hazard name, in English, and over a two-month period (i.e. the calendar month "n" having the highest number of RASFF notifications and the month "n-1" to increase search success). Then if these searches were unsuccessful, finer searches were conducted using combinations of words (e.g. hazard AND food product), several languages (e.g. the languages of the countries which notify the hazard and the countries where the hazard originates from), and longer time intervals.

3. Results

3.1. Customization

A preliminary assessment was made on the news sources and categories (CBRN) used by MedISys and on some categories definition (these terms are defined in 2.1.2). This evaluation was conducted to determine how MedISys can be tailored to EFSA needs.

3.1.1. News sources

MedISys covers 158 countries with various levels of media coverage (Figure 3). However, the majority of these countries (64%) are poorly represented and 34 countries are not covered by MedISys

(Figure 4). In particular, the Southern hemisphere is under-represented; the Caribbean area, Africa, and South-East Asia. There are also a few uncovered countries in Europe (Principality of Monaco, Principality of Liechtenstein, Andorra, San Marino). However, these locations are located within countries monitored by MedISys (France, Germany, Spain, and Italy), and therefore they should be represented by the press of these countries.

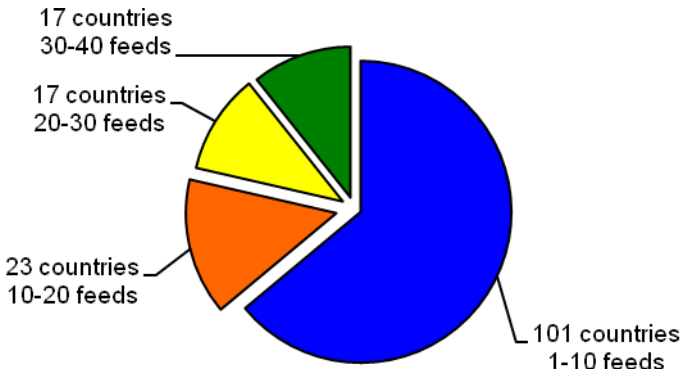


Figure 2: Media coverage (expressed in classes of number of feeds) among the 158 countries monitored by MedISys.

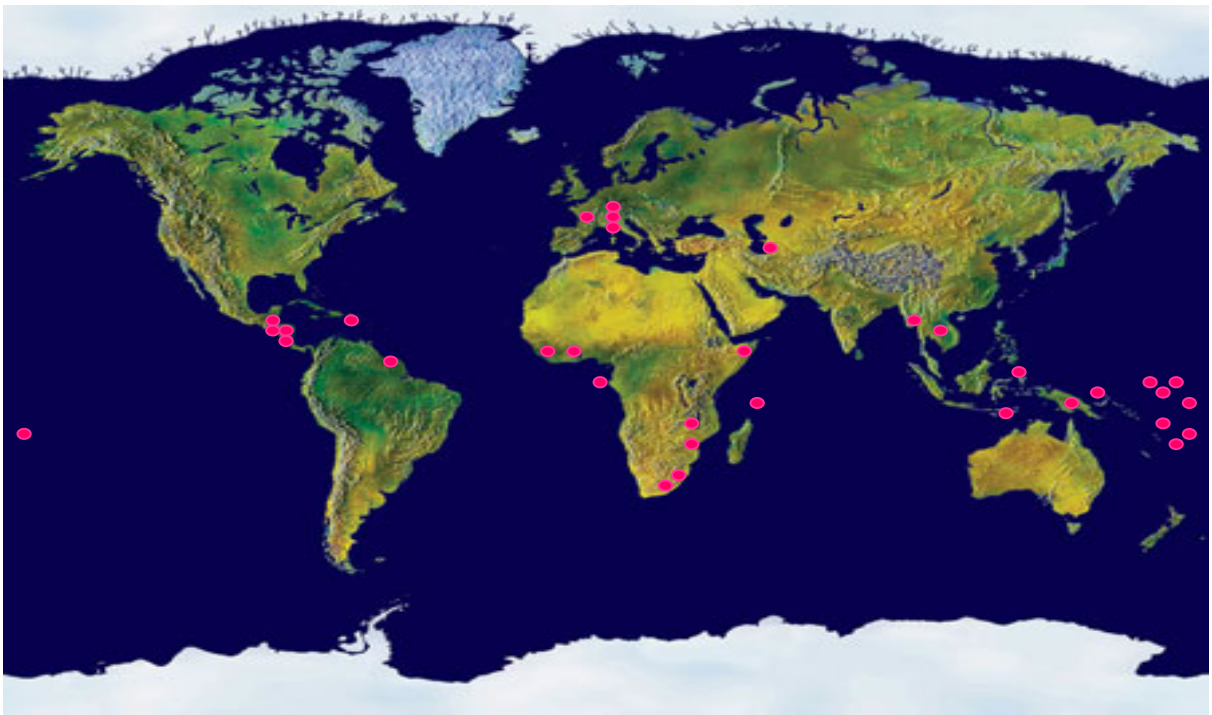


Figure 3: Gaps of media coverage in Europe and worldwide (pink dots represent countries whose media are not covered by MedISys).

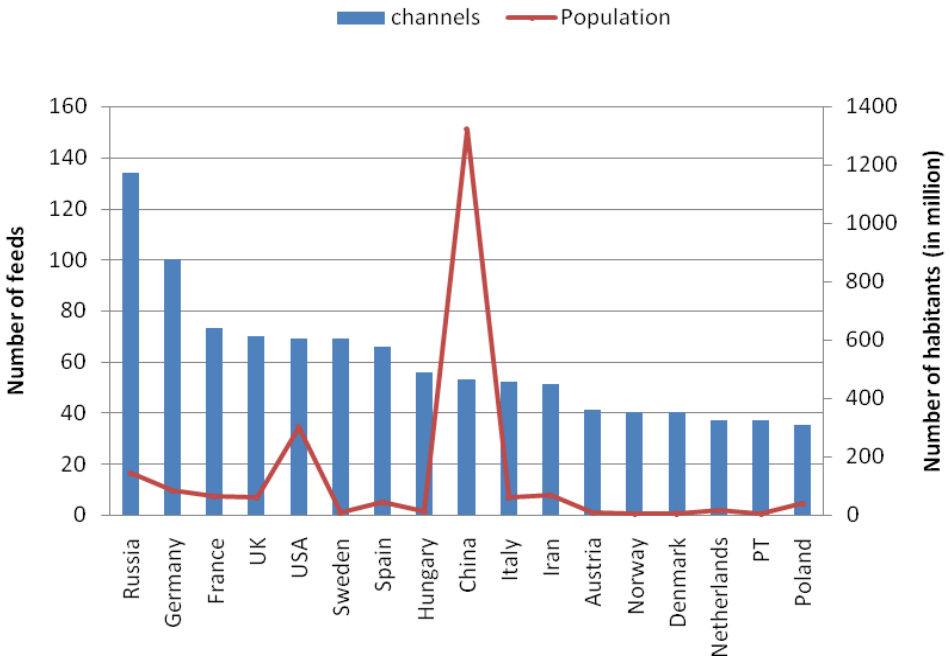


Figure 4: Number of feeds filtered by MedISys, worldwide, in the top 17 countries (PT: Palestinian Territories) and population size.

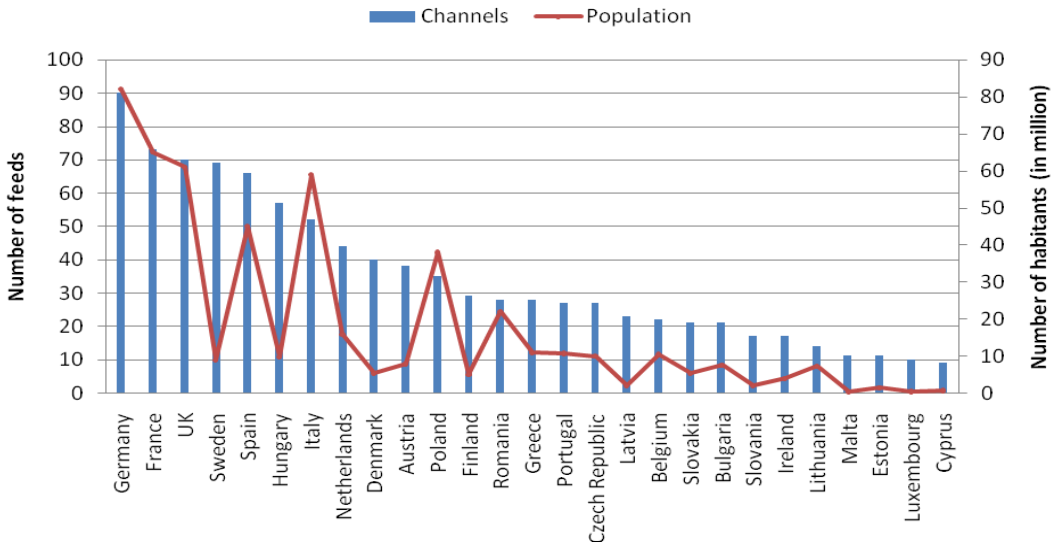


Figure 5: Number of feeds filtered by MedISys in the 27 Member States and population size.

Within the top 17 countries (i.e. the best covered by MedISys), Russia, China, the USA, Iran and the Palestinian Territories are the predominant ones at the international level and Germany, France, the UK, Sweden, Spain, Hungary, Italy, Denmark, the Netherlands, Austria, Poland and Finland (in decreasing order) are the best covered at the European level (Figure 5). However, media coverage is usually related to the country population size as shown on Figure 6. While media coverage in the EU is quite well represented on MedISys, it could be implemented for other countries like China and the USA, for example (Figure 5).

In addition, within the EU, countries that entered before 1996 tended to be better represented than the ones that joined more recently (Figure 5). However, within those well-represented European countries, some sources relevant for EFSA and food safety issues may still be missing as demonstrated by the monitoring of nicotine in wild mushrooms (see section 3.2.1).

New sources were suggested by the Emerging Risks Unit for Bulgaria, France, Germany, Greece, and Italy (Tables 1-5, Appendix A).

Within the 27 Member States, MedISys monitors a total of 969 sources which are mostly coming from the media (92.3%). While all general media are included in the system, the financial press, which is a reliable source, is not always well-represented. A few sources come from the medical (3.8%), TV/radio (3.7%), and the scientific (0.2%). The latter is not well-presented because MedISys needs to detect early outbreaks that would be described much later in the peer-reviewed journals. However, a few scientific sources pertaining to the field of emerging risks were suggested by EMRISK (Table 6, Appendix A).

The Communication Directorate of EFSA provided a list of 688 media sources coming from EU and outside EU (USA, Turkey, Finland, and China) to be added to MedISys. The JRC has recently added blog sources to be monitored by MedISys. The EMRISK proposed a small list of food and feed blogs found on internet with the search engine "technorati" at http://technorati.com. However, a more systematic search needs to be conducted to find relevant and reliable blog sources covering both food and feed and emerging risks areas.

3.1.2. Categories

MedISys focuses on health-related events. While such events are relevant to EMRISK, they need to be further developed to better cover food and feed-related issues. A preliminary evaluation of the group of categories "***Diseases***" and "***Chemicals***" allowed us to identify gaps which were further filled in by adding new categories within these two groups (Tables 1 and 2, Appendix B).

- The "***Diseases***" was implemented with the "common causes of food borne disease" list (http://www.cidrap.umn.edu/cidrap/content/biosecurity/food-biosec/causes/causes-foodborne.html) found on the Centre for Infectious Disease Research and Policy (CIDRAP). This proposed new list of food borne diseases is presented in Table 13 (Appendix B).

- The "***Chemicals***" was expanded with four RASFF hazard categories (i.e. "chemicals contamination", "metals", "industrial contaminants", "pesticide residues" and "residues of veterinary medicinal products"). This detailed list is provided in Table 2 (Appendix B).

The preliminary word analysis conducted on articles that were either related or unrelated to hazard events such as salmonella, melamine, and DDT gave similar trends for words combination and proximity. However, words weights varied among these three groups of hazards.

In related articles, the **combination of words** picked up for salmonella (Tables 1 and 2, Appendix C) was congruent with the one found for melamine (Tables 5 and 6, Appendix C) whereas, there was no consistency of words for both salmonella (Tables 3 and 4, Appendix C) and melamine (Tables 7 and 8, Appendix C) in unrelated articles. The DDT hazard, both related and unrelated articles showed no pattern of consistency in combination of words.

The analysis on **proximity**, which is the distance (calculated in number of words) at which searched words are located in the text, showed that it is higher (i.e. within a small number of words) in related articles than in unrelated articles (Appendix C). Words in articles dealing with the salmonella hazard occurred closer in the text than those for the melamine and DDT hazards. In addition, in unrelated

articles, searched words never occurred in the title (to the exception of DDT) whereas it appeared quite frequently in related articles (Appendix C). More hazards from each category need to be tested to determine whether these preliminary data correspond to any real trend.

The frequency at which the words were cited, which can be referred to the **weight** given to any specific words when setting a category in the notification editor, also varied from one group to another. These trends varied in the same way as above. In related articles, congruent words (Tables 2, 6, 10, Appendix C) appeared more frequently than in unrelated articles (Tables 4, 8, 12, Appendix C). Words in articles dealing the salmonella hazard occurred more frequently in the text than those from the melamine and DDT. Once again, to confirm these trends more hazards from each category need to be tested.

These results show the importance of developing a proper strategy to implement categories, reduce noise during the processing phase and therefore increase the chance of getting relevant articles during monitoring activities.

3.2. Monitoring

The monitoring of MedISys showed that all hazards picked on this system were not systematically reported by ProMED-mail and rarely by RASFF (Table 2). On the few occasions these systems notified a hazard, they usually did it within a short delay (between 1-3 days), except for the peanut butter in the USA which took six days to be notified by the RASFF and for the Ebola Reston virus which was reported by ProMED-mail 18 days after MedISys (Table 2).

Table 2: List of hazards retrieved on MedISys from January to April 2009, notification dates, and Time delay for ProMED-mail and RASFF to report these hazards (positive and negative results indicate early and late notice, respectively).

		Dates of released notifications			**Time delay**	
Hazards	Products	MedISys	ProMED-mail	RASFF	ProMED -mail	RASFF
Salmonella	**Peanut butter**	**10-01-09**	**10-01-09**	**16-01-09**	**0**	**+6**
Ebola Reston	**Swine**	**23-01-09**	**10-02-09**	**no notification**	**+18**	**-**
Additives	**Milk**	**13-02-09**	**no report**	**no notification**	**-**	**-**
E. sakazaki	**Milk**	**13-02-09**	**no report**	**16-02-09**	**-**	**+3**
Clenbuterol	**Pig organs**	**23-02-09**	**25-02-09**	**no notification**	**+2**	**-**
Rabies	**Dog meat**	**17-03-09**	**18-03-09**	**no notification**	**+1**	**-**
Rabies	**Dog meat**	**30-03-09**	**30-03-09**	**no notification**	**0**	**-**
Salmonella	**Pistachios**	**30-03-09**	**02-04-09**	**01-04-09**	**+3**	**+2**
Salmonella	**Pepper**	**30-03-09**	**02-04-09**	**no notification**	**+3**	**-**
Nicotine	**Mushroom**	**13-11-08**[(a)]	**no report**	**no notification**	**-**	**-**

(a) the nicotine hazard was first published in the media before January 2009, but the European Commission asked the EFSA to deliver an assessment in April 2009 so it was included in the table

Hazards that were not reported by the RASFF were either not relevant to the RASFF or the EU (e.g. clenbuterol on Hangzhou markets, milk additives in one Chinese company, traditional meat processing that lead to rabies disease) or they were not presenting any threat to the EU (e.g. pepper from the USA, pigs organs from China).

The request made by the European Commission to assess the risk of contaminated mushrooms by nicotine was notified in the media in November 2008. However, a search in the English language with the words “mushrooms, nicotine” led to no result, whereas a search in the German language with the words “nikotin, steinpilze” gathered a few articles (n=6). A further research on other internet-scanning

systems such as **Google** and **Yahoo** was more successful and more articles were picked up (n=7 and n=19, respectively).

3.3. Early detection

A summary of the results of the searches conducted for the nine case studies on the various monitoring systems (MedISys, ProMED-mail and RASFF) is given in Table 3.

Detailed information on searches and references (RASFF notifications, MedISys articles and ProMED-mail reports) is described in Appendices D and E (for the MedISys references, some http links may not be available; for copy right reasons, the full articles could not be published in this report).

Table 3: Period of search and number of articles, reports and notifications retrieved on MedISys, ProMED-mail and RASFF, respectively for the nine case studies.

Hazards	Period of search	RASFF notifications	MedISys articles	ProMED-mail reports
High level				
Salmonella	**01/10/07-30/11/07**	**81**	**148**	**18**
Mycotoxins	**01/12/07-31/01/08**	**188**	**18**	**0**
Heavy metals	**01/07/07-31/08/07**	**71**	**451**	**0**
Medium level				
Melamine	**01/09/08-31/10/08**	**75**	**4904**	**15**
Aluminium	**01/01/09-28/02/09**	**8**	**2236**	**0**
Dioxins	**01/09/07-31/10/07**	**8**	**88**	**0**
Low level				
Vibrio cholera	**01/07/08-31/08/08**	**4**	**24**	**1**
DDT	**01/02/08-31/03/08**	**1**	**80**	**0**
Radioactivity	**01/09/08-31/10/08**	**3**	**140**	**0**

3.3.1. Case study 1: Salmonella

Out of the 81 RASFF notifications found on salmonella contaminations in food and feed between October and November 2007 (Table 3), MedISys, detected only one event related to these notifications and it was on **alfalfa sprouts** "*Norway's Food Safety Authority has stopped the sale of alfalfa sprouts from the company… after ten people have been found to have the bacteria Salmonella Weltevreden*"[5] whereas ProMED-mail did not detect any of these notifications.

[5] World News (26/10/2007)

However, both MedISys and ProMED-mail reported salmonella contaminations that were not notified by RASFF. MedISys reported contaminations in **white chocolate** "*white chocolate baking squares may be contaminated with salmonella*"[6], **hot pie** "*XX voluntarily stopped production Tuesday at the Missouri plant that makes its Banquet pot pies after health officials said the pies may be linked to 139 cases of salmonella in 30 states*"[7], **hummus** "*More than 200 people were treated for food poisoning in Jordan after eating hummus - ground chickpeas - from a restaurant in Jerash south of the capital, a health official said on Saturday 26-10-2007*"[8], **brazil nuts** in chocolate "*The company believes the outbreak originated from a batch of contaminated Brazil nuts, which are used to make one of the brand's best known ranges "Just Brazils*"[9], **chocolate** "*Fox's Confectionery, makers of Fox's Glacier Mints and other sweets, said on Tuesday it had halted production at a plant after discovering salmonella but said no contaminated products had left its factories*"[10], and **meat and poultry** "*the Agriculture Department's Food Safety and Inspection Service will increase testing of Canadian meat for salmonella"*[11] (Figure 7). ProMED-mail also reported the salmonella contaminations in hot pies[12] and hummus[13].

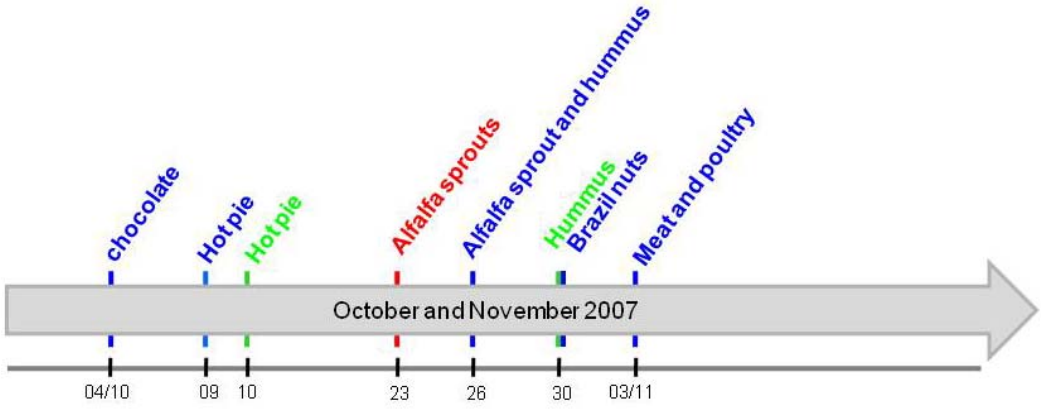


Figure 6: Chronology of the salmonella contamination events reported by RASFF (in red), MedISys (in blue), and ProMED-mail (in green).

In summary, the contamination of **alfalfa sprouts** was notified by RASFF three days earlier than MedISys and the contaminations of **hot pies** and **hummus** were notified by MedISys between one and four days earlier than ProMED-mail (Table 4).

Table 4: Time delay in MedISys and ProMED-mail to report salmonella contaminations (positive and negative time delays indicate early and late notices, respectively).

[6] CBC (4/10/2007)
[7] Msnbc (9/10/2007)
[8] Iol (27/10/2007)
[9] DailyMail (30/10/2007)
[10] Mirror (30/10/2007)
[11] USA Today (03/11/2007)
[12] ProMED (10-10-2007)
[13] ProMED (29-10-2007)

Products	Dates of released information/notifications			Time delay	
	RASFF	MedISys	ProMED-mail	MedISys	ProMED-mail
Hot pie	**-**	**09-10-2007**	**10-10-2007**	**-**	**-**
Alfalfa sprouts	**23-10-2007**[14]	**26-10-2007**	**-**	**-3**	**-**
Hummus	**-**	**26-10-2007**	**30-10-2007**	**-**	**-**

3.3.2. Case study 2: Mycotoxins

Among the different mycotoxins notified by RASFF between December 2007 and January 2008 (i.e. aflatoxin, deoxynivalenol, and ochratoxin), aflatoxin accounted for 96% of the notifications. However, no article was detected by MedISys with the word "mycotoxin". Therefore, another search was conducted on MedISys with the word "aflatoxin" and contaminations were found in **spices** coming from Pakistan "*the European Union has finally imposed a ban on export of chilies from Pakistan because the commodity has been found infected with a disease called Aflatoxin*"[15], in **spices** coming from India "*Indian spices exports to European Union are likely to come under closer and tougher scrutiny with the detection of higher than permissible level of aflatoxins… The European Union authorities propose to check 50% of all the export consignments of pepper (Black and White), ginger, nutmeg, chilli and turmeric powder from India*"[16], in **pistachios** "*The health services have blocked another consignment of pistachio nuts from Iran at Limassol port due to the high level of a carcinogenic substance found in the nuts…. The Health Services found levels of aflatoxin in the pistachios that went beyond the accepted limits set by EU regulations*"[17], and in **pet food** "*The pet food, which was contaminated with a mold called aflatoxin, was produced at Diamond Pet Foods' plant in South Carolina*"[18] (Figure 8).

On ProMED-mail, the words "mycotoxin", "aflatoxin", "deoxynivalenol", "ochratoxin", and "mould" gave no result and the word "toxin" brought many irrelevant notifications (e.g. undiagnosed deaths, botulism, etc.) to the exception of one report of *Penicillium citreonigrun* contamination in rice coming from Brazil[19], but it was not related to any RASFF notification. The word "mycotoxins" (plurial form) gave three notifications of undiagnosed deaths of camels from Saudi Arabia[20, 21, 22] and the word "fungus" gave many notifications of crop diseases (bud rot, blister blight, wheat stem rust, ergot) that may lead to mycotoxin contamination, but none of them was related to the RASFF notifications.

[14] RASFF 2007.0760
[15] Daily Times (15/12/2007)
[16] The Financial Express (30/01/2008)
[17] Cyprus-Mail (2009)
[18] USA-Today (04/01/2008)
[19] ProMED (08-01-2008)
[20] ProMED (08-01-2008)
[21] ProMED (10-01-2007)
[22] ProMED (21-01-2008)

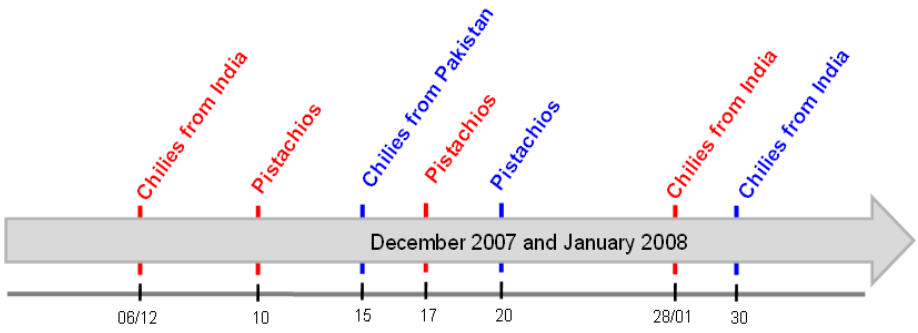


Figure 7: Chronology of the mycotoxin contamination events reported by RASFF (in red), MedISys (in blue), and ProMED-mail (in green).

The contamination of **chilies** from India was notified by RASFF between two and 55 days earlier than MedISys, and the contamination of **pistachios** between three and 10 days before MedISys (Table 5).

Table 5: Time delay in MedISys and ProMED-mail to report mycotoxins contaminations (positive and negative time delays indicate early and late notices, respectively).

Products	Dates of released information/notifications			Time delay	
	RASFF	MedISys	ProMED -mail	MedIsys	ProMED -mail
Chilies from India	**06-12-2007**[23] **& 28-01-2008**[24]	**30-01-2008**	-	**-55 & -2**	-
Pistachios	**10**[25] **& 17-12-2007**[26]	**20-12-2007**	-	**-10 & -3**	-

3.3.3. Case study 3: Heavy metals

In RASFF, the category "heavy metals" contains 16 hazard types, but between July 2007 and August 2007, only 7 hazard types were reported (i.e. mercury, cadmium, nickel, lead, arsenic, chromium, and zinc)[27, 28, 29, 30, 31, 32, 33].

A huge number of articles were found in MedISys with the word "heavy metal" (Table 3), but only three were related to food contamination events. They were on **water** contamination by uranium "*The water supply of a tiny community in the Okanagan Valley shows a high level of uranium, but regional health officials are saying the water is safe to drink*"[34], **pineapple** contamination by cadmium "*High*

[23] RASFF 2007.CVI
[24] RASFF 2008.AEL
[25] RASFF 2007.CWQ
[26] RASFF 2007.CXZ
[27] RASFF 2007.BPZ
[28] RASFF 2007.0439
[29] RASFF 2007.BQZ
[30] RASFF 2007.0517
[31] RASFF 2007.BTU
[32] RASFF 2007.BRX
[33] RASFF 2007.CAI
[34] CBC News (13/08/2007)

levels of the toxic heavy metal cadmium have been found in South African pineapples after farmers unwittingly used fertilizer contaminated with this trace element"[35] and about **general food** contaminated by kitchen metal knives "*Yet many other people who cook pay little attention to their kitchen knives*"[36]. However, none of these articles was related to any of the RASFF notifications.

ProMED-mail did not report any heavy metal contamination. When other words related to this category (i.e. arsenic, cadmium, etc.) were searched for, a lot of noise was created with the word "lead" and the verb "to lead". Finally, only one report was found (on arsenic), but it was not related to any RASFF notification.

3.3.4. Case study 4: Melamine

The first RASFF notifications on melamine contaminations were reported late September whereas ProMED-mail and MedISys reported on these events earlier. The first observations were made by ProMED-mail on 10-09-2008 with infants sick and hospitalized in China supposedly due to contaminated infant **milk powder**[37]. The day after, MedISys mentioned the possible contamination of children with **milk** tainted by melamine "*Chinese newspapers report that some infant formula has been linked to kidney problems and kidney stones in babies in China because the formula contains melamine*"[38]. Then, on 19-09-2008, WHO made an international recall on dairy products and the EFSA delivered an opinion on 25-09-2008 (Figure 8). Thereafter, RASFF reported melamine contaminations in several products that were reported earlier by MedISys and ProMED-mail (Figure 9).

MedISys mentioned the contamination of various milk products by melamine (e.g. **cookies**, **candies**, **milk drink**, **crackers**, **yoghurts**, **and chocolates**) "*Other milk products such as powdered milk, liquid milk, candies, cookies, milk tablets, toffees, cream crackers, egg rolls, yoghurt drinks and chocolates from China are affected by the ban*"[39], **pies** "*Hypermarkets and shops in the city have taken off the shelves all suspected milk and milk-based products from China. Health Ministry officials, who conducted random checks at the hypermarkets and shops, found one shop that had one product suspected of containing melamine*"[40], and **animal food** "*the Canadian Food Inspection Agency (CFIA) will require dairy ingredients and soybean meal originating from China intended for use as livestock feed to be tested for melamine and cyanuric acid....*"[41] (Figure 9).

ProMED-mail reported melamine contaminations in **cookies**[42], **candies**[43], **crackers**[44], **chocolate**[45], **yoghurt**[46], and **animal food**[47] (Figure 8), but not in **milk drink** and **pies**.

[35] Iol (06/07/2007)
[36] Nst (28/06/2007)
[37] ProMED (10-09-2008)
[38] USAToday (11-09-2008)
[39] Ghanaweb (23-09-2008)
[40] Nst (25-09-2008)
[41] News.gc.ca (17-10-2008)
[42] ProMED (02-10-2008)
[43] ProMED (24-09-2008)
[44] ProMED (25-09-2008)
[45] ProMED (04-10-2008)
[46] ProMED (19-09-2008)
[47] ProMED (20-10-2008)

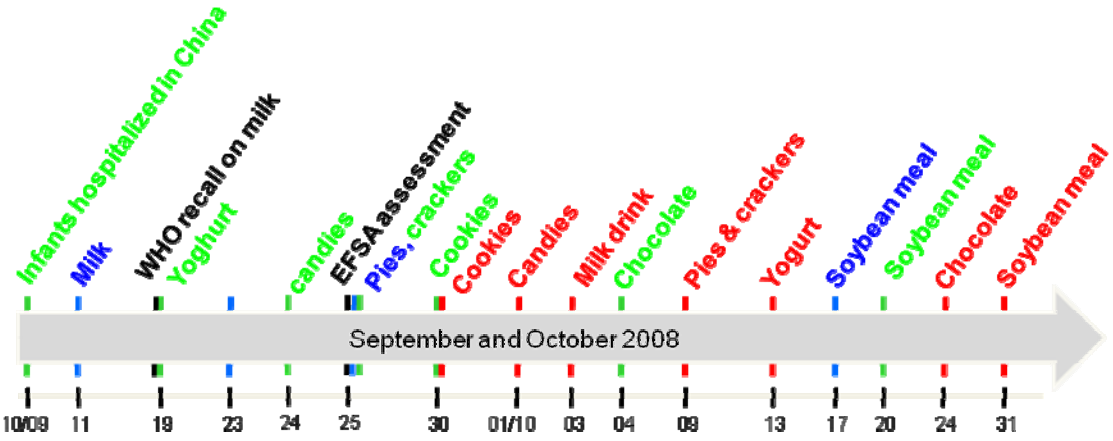


Figure 8: Chronology of the melamine contamination events reported by RASFF (in red), MedISys (in blue), and ProMED-mail (in green).

In summary, MedISys and ProMED reported melamine contaminations between 1 and 31 days, and 1 and 24 days, respectively, before RASFF (Table 6).

Table 6: Time delay in MedISys and ProMED-mail to report melamine contaminations (positive and negative time delays indicate early and late notices, respectively).

Products	Dates of released information/notifications			Time delay	
	RASFF	MedISys	ProMED-mail	MedISys	ProMED-mail
Milk drink	**03-10-2008**[48]	**23-09-2008**	**-**	**10**	**-**
Yogurt	**13-10-2008**[49]	**23-09-2008**	**19-09-2008**	**20**	**24**
Cookies	**30-09-2008**[50]	**23-09-2008**	**30-09-2008**	**7**	**0**
Candies	**25-09-2008**[51]	**23-09-2008**	**24-09-2008**	**2**	**1**
Chocolate	**24-10-2008**[52]	**23-09-2008**	**04-10-2008**	**31**	**20**
Crackers	**09-10-2008**[53]	**23-09-2008**	**25-09-2008**	**16**	**14**
Pies	**09-10-2008**[54]	**25-09-2008**	**-**	**14**	**-**
Soybean meal (feed)	**21-10-2008**[55]	**17-10-2008**	**20-10-2008**	**4**	**1**

[48] RASFF 2008.1184
[49] RASFF 2008.1246
[50] RASFF 2008.1163
[51] RASFF news 08.475
[52] RASFF 2008.1325
[53] RASFF 2008.1221
[54] RASFF (2008.1223)
[55] RASFF news 08.497

3.3.5. Case study 5: Aluminium

All the eight RASFF notifications between January and February 2009 were on **noodles** contamination, with the first notification published on 04/02/2009[56]. On MedISys, numerous articles were found with the word "aluminium", but only one was related to food, "*Study shows levels of metals and other elements in food stable or lower since 2000*"[57], but it was not related to any RASFF notification. No aluminium contamination was reported by ProMED-mail.

Another search was conducted on MedISys with the word "noodle". A total of 162 articles were picked up, but only one was about food contamination - Thai noodle salads by *Salmonella*[58], though not on contamination by aluminium. Finally, since the contaminated noodles came from China, a finer search with the combination of words "noodle AND aluminium" in the Chinese language, gave many irrelevant articles.

3.3.6. Case study 6: Dioxins

RASFF notifications were found on **supplementary feed for ruminants**[59], **premixture**[60], **cod liver**[61] and **fish meal**[62]. MedISys and ProMED-mail did not detect any of these contaminations events. When a search was conducted in the Russian language with the combination of words "dioxin" and "food", four articles were retrieved in September and October 2008, but none of these articles were related to any RASFF notifications.

3.3.7. Case study 7: *Vibrio cholera*

RASFF notifications were found on frozen **black tiger shrimps**[63, 64, 65] and **pangasius fillets**[66]. On MedISys, food contaminations were found in **shellfish** "*The New Jersey Department of Environmental Protection's Bureau of Marine Water Monitoring announced a suspension of shellfish harvesting -- clams, mussels and oysters -- from state waters"* [67] and **water** in Kenya [68] and Pakistan[69]. However, none of these contaminations was related to the above RASFF notifications.

On ProMED-mail, the word "*Vibrio cholera*" led to only one report and the word "*Vibrio*" gave three alerts related to food contamination; one in **frozen fish**[70] from Norway ex-Vietnam, one in **oysters**[71], and one in **shellfish**[72] from the USA. However, none of them was related to the RASFF notifications.

3.3.8. Case study 8: DDT

The RASFF notification on DDT was found in **complementary feed**. On MedISys, only one article was related to food contamination by DDT but it was not related to the RASFF notification; it was on

[56] RASFF 2009.AGO
[57] FlexNews (29/01/2009)
[58] Earth Times
[59] RASFF 2007.0633
[60] RASFF 2007.CEY
[61] RASFF 2007.0697, 2007.0693, 2007.0694, 2007.0695
[62] RASFF 2007.CHF
[63] RASFF 2008.BDH
[64] RASFF 2008.BDQ
[65] RASFF 2008.BFP
[66] RASFF 2008.0942
[67] GoogleNewsHealth
[68] IriNews (29/08/2008)
[69] The News (01/07/2008)
[70] ProMED (27-07-2008)
[71] ProMED (21-08-2008)
[72] ProMED (21-08-2008)

vegetables "*Greenpeace found DDT, BHC, dieldrin and heptachlor openly sold in the vegetable markets of Karachi*"[73].

On ProMED-mail, there was no information on DDT. Other words such as pesticide and chemical gave many notifications, but none of them were related to the RASFF notification.

3.3.9. Case study 9: Radioactivity

RASFF notifications reported contamination by radioactivity in **hedgehog mushrooms** from Bulgaria [74] and **frozen bilberries** from Belarus [75]. On MedISys, 140 articles were retrieved with the word "radioactivity", but none of them were related to any food contamination. However, when a search was conducted with the combination of words "bilberry AND radioactive" in the Russian language, articles were picked up in the local press on 72 day before the RASFF notification[76] (Table 7).

On ProMED-mail, there was no information on contamination by radioactivity.

Table 7: Time delay for MedISys and ProMED-mail to report radioactivity contaminations (positive and negative results indicate early and late notice, respectively).

	Dates of released information/notifications			**Time delay**	
Products	RASFF	MedISys	ProMED-mail	MedIsys	ProMED-mail
Mushroom	**22-09**	**-**	**-**	**-**	**-**
Bilberries	**09-10**	**29-07-2008**	**-**	**72**	**-**

3.3.10. Summary on the evaluation of the case studies

Time delays for MedISys and ProMED-mail to report RASFF notifications for the nine case studies - *Salmonella*, mycotoxins, heavy metals, melamine, aluminium, dioxins, *V. cholera*, and radioactivity - are summarized in Table 8.

RASFF hazards were reported by MedISys in four cases out of nine and by ProMED-mail in one case out of nine. MedISys was earlier than ProMED-mail. In two cases out of nine, MedISys reported RASFF hazards early, between one and 2.5 months in advance whereas in two other cases out of nine, it reported RASFF hazards late, between 3 days and 1.5 month later than RASFF (Table 8).

Table 8: Time delay for MedISys and ProMED-mail to report RASFF hazards (positive and negative results indicate early and late notice, respectively).

Hazard (period)	***MedISys (in days)***	***ProMED-mail (in days)***
Salmonella	-3	nd

[73] Daily Times (25/02/2008)
[73] RASFF 2008.0270
[74] RASFF 2008.1113
[75] RASFF 2008. BOZ
[76] Novayagazeta

Mycotoxins	-55 to -3	nd
Heavy metals	nd	nd
Melamine	31	24
Aluminium	nd	nd
Dioxins	nd	nd
Vibrio cholera	nd	nd
DDT	nd	nd
Radioactivity	72	nd

nd: no relevant data found

The hazards that were never notified by RASFF may be due to an absence of legislation or because the hazards were too local and never reached the EU market. Conversely, some hazards that were notified by RASFF were never found in the media probably because they were not of public interest or because the press never got the information (e.g. border rejections). To anticipate risks and respond in a proactive way, the notifications delivered by the media that were not picked up by the RASFF may be valuable in the light of import data.

The small time delay between MedISys and ProMED-mail (a couple of days) may be related to the time required by experts to assess the value of the information. The bigger time delay between MedISys and RASFF (a couple of days to several weeks) may be linked to the time for laboratories to confirm the data and/or to the time for hazards which emerged outside EU to reach EU borders and markets.

Although the web-scanning systems, MedISys and ProMED-mail, have different time delays in reporting hazards, the information they deliver is usually complementary. ProMED-mail did not perform as well as MedISys and RASFF in signalling food and feed-borne hazards because this system focuses on human and animal outbreaks rather than on local food and feed contaminations. This is well-demonstrated by its low-reporting of hazards like aluminium, heavy metals, and radioactivity. However, ProMED-mail is an important additional source of information, in particular for hazards that are more frequently reported outside EU (mycotoxins and *Salmonella*) and which may pose a risk to EU through trade. In addition, as illustrated by the melamine case-study, ProMED-mail subscribers cover a wide geographic range which allows early detection, through observations at a local level, of emerging risks.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The results of this study lead to the following conclusions:

- MedISys is an efficient monitoring and early detection tool. At many occasions, it reported hazards before ProMED-mail (in the range of a couple of days earlier) and RASFF (in the range of a couple of days to a couple of weeks earlier). Whilst an advance of a couple of days does not allow much anticipation, it can still be useful in cases of crisis preparedness.

- The monitoring of other sources of information such as ProMED-mail and RASFF proved to be complementary to the monitoring of MedISys. While RASFF reports hazards that may present a risk in EU, ProMED-mail reports early hazards that may present a risk to EU through trade.

- MedISys sensitivity to food and feed hazards needs to be increased with further customization of the system. Preliminary analyses indicate that new news sources should be added (e.g. sources from countries in the Southern Hemisphere and blogs) and that multi-lingual categories related to food and feed need to be developed and appropriately defined.

RECOMMENDATIONS

- It is proposed to use MedISys for regular media monitoring for the identification of emerging risks.

To make a better use of MedISys for EFSA's needs, it is recommended to:

- Add more sources coming from emerging countries and countries located in the southern hemisphere. Expand blog sources related to food and feed.

- Use an existing food and feed database or classification to select a range of food and feed hazards and products to be added to MedISys. The RASFF database may be a good starting point to select such items.

- Develop multi-lingual food and feed categories and add Asian languages from representative countries in terms of import volumes to EU, hazard notifications in EU, emerging risks to EU (etc.).

- Get technical support to refine category definitions in order to reduce noise during monitoring activities (i.e. to decrease the probability of getting irrelevant articles).

- In agreement with the JRC, implement the access to food and feed categories on MedISys (e.g. group food and feed categories in a specific allocated web space to speed up monitoring activities; add automatic English translations for selected articles, etc.).

- Store relevant information found in MedISys on food and feed hazards in an Access database to be further analysed (e.g. trend analysis).

REFERENCES

Collier N, Doan S, Kawazoe A, Goodwin R, Conway M, Tateno Y, Ngo Q-H, Dien D, Kawtrakul A, Takeuchi K, Shigematsu M, Taniguchi K, 2008. BioCaster: detecting public health rumors with a web-based text mining system. Bioinformatics Applications Note 24 (24), 2940-2941.

Cowen P, Garland T, Hugh-Jones M, Shimshony A, Handysides S, Kaye D, Madoff L, Pollack P and Woodall J, 2006. Evaluation of ProMED-mail as an electronic early warning system for emerging diseases: 1996 to 2004. JAVMA 229 (7), 1090-1099.

EC (European Commission), 2002. Regulation (EC) No 178/2002 of the European Parliament and of the Council. Official Journal of the European Communities 1.2.2002 L 31/1.

EFSA (European Food Safety Authority), 2009a. The Emerging Risks Unit: Mission, Mandate, Tasks and Definition. (EFSA/EMRISK/029).

EFSA (European Food Safety Authority), 2009b. Mandate proposed by the EFSA to the Emerging Risks Unit for a self-tasking Working Group on the development of web monitoring for the detection of emerging risks. Mandate M-2009-0075.

EFSA (European Food Safety Authority), 2009c. Mandate proposed by the EFSA to the Emerging Risks Unit for a self-tasking Working Group on the establishment and maintenance of routine analysis of data from the Rapid Notification System on Food and Feed. Interim Technical Report. (EFSA/EMRISK/073).

EFSA (European Food Safety Authority), 2007. Definition and description of “emerging risks” within the EFSA’s mandate. (EFSA/SC/415 Final).

EFSA (European Food Safety Authority), 2006. Opinion of the Scientific Committee on a request from EFSA related to the early identification of emerging risks. The EFSA Journal (2006) 375, 1-14.

EFSA (European Food Safety Authority), 2003. Background document prepared by the Working Group on Emerging Risks (EFSA/SC/22), 13.11.2003.

ESCO WG on Emerging Risks, 2009. Technical report of EFSA prepared by the ESCO WG on Emerging Risks. EFSA Technical Report (2009) 224, 1-34.

Friefeld C, Mandl K, Reis B and Brownstein J, 2008. HealthMap: global infectious disease monitoring classification and visualization of internet media reports. JAMIA 15, 150-157.

Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M and Brilliant L, 2009. Detecting influenza using search engine query data. Nature Letters 457, 1012-1015.

Grein T, Kamara K-B, Rodier G, Plant A, Bovier P, Ryan M, Ohyama T, Heymann D, 2000. Emmerging Infectious Diseases 6 (2), 97-102.

Heymann D and Rodier G, 2001. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. Lancet Infect. Dis. 1, 345-353.

Linge J, Steinberger R, Weber T, Yangarber R, van der Goot E, Al Khudhairy D, Stilianakis N (2009). Internet surveillance systems for early notificationing of health threats. Eurosurveillance 14(13), 1-2.

Madoff L, Woodall J, 2005. The internet and global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. Arch. Med. Res. 36, 724-730.

Madoff L, 2004. ProMED-mail: an early warning system for emerging diseases. CID 39, 227-232.

Mykhalovskiy E and Weir L, 2006. The global public health intelligence network and early warning outbreak detection. A Canadian contribution to global public health. Revue Canadienne de Santé Publique 97 (1), 42-44.

Pouliquen B., Kimler M., Steinberger R., Ignat C., Oellinger T., Blackler K., Fuart F., Zaghouani W., Widiger A., Forslund A.C., Best C. (2006) Geocoding multilingual texts: Recognition, Disambiguation and Visualisation. Proceedings of the 5th International Conference on Language Resources and Evaluation LREC2006, pp. 53-58. Genoa, Italy, 24-26 May 2006.

Steinberger R, Fuart F, van der Goot E, Best C, von Etter P and Yangarber R 2008. Text Mining from the Web for Medical Intelligence. In: Fogelman-Soulié Françoise, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.): Mining Massive Data Sets for Security. pp. 295-310. IOS Press, Amsterdam, The Netherlands. Available from: http://langtech.jrc.it/Documents/2009_MMDSS_Medical-Intelligence.pdf

VWA (Food and Consumer Product Safety), 2006. Forming a global system for identifying food-related emerging risks – EMRISK. Final report (contract no. EFSA/SC/Tender/01/2004).

Wilson K and Brownstein J, 2009. Early detection of disease outbreaks using the internet. CMAJ online 1-3. http://www.cmaj.ca/cgi/rapidpdf/cmaj.1090215

Williams J, Phillips T, Jolly P, Stiles J, Jolly C, Aggarwal D, 2004. Human aflatoxicosis in developing countries: a review of toxicology, exposure, potential health consequences, and interventions. Am. J. Clin. Nutr., 80: 1106-22.

Woodall JP, 2001. Global surveillance of emerging diseases: ProMED-mail perspective. Cad. Saude Publica, Rio de Janeiro, 17(Suplemento): 147-154.

APPENDICES

A. LIST OF NEW SOURCES SUGGESTED BY EMRISK TO THE JRC

Table 9: Bulgarian sources

Source name	URL
BgFactor	**www.bgfactor.org**
Bgradio	**www.bgradio.net**
BNR	**www.bnr.bg**
Btv	**www.btv.bg**
Class	**www.class.bg**
Duma	**www.duma.bg**
Darikradio	**www.darikradio.bg**
Dnevnik	**www.dnevnik.bg**
Ekipnews	**www.ekipnews.com**
eurocom	**www.eurocom.bg**
Monitor	**www.monitor.bg**
m-sat	**www.m-sat.bg**
Ntv	**www.ntv.bg**
Radio express	**www.radioexpress.bg**
StandartNews	**www.standartnews.com**
Trud	**www.trud.bg**
24hours	**www.24hours.bg**
7dni	**www.7dni.tv**

Table 10: French sources

Source name	URL
(la) Depêche du Midi	**www.ladepeche.fr/**
(le) Dauphiné libéré	**www.ledauphine.com/**
France Soir	**www.francesoir.fr/**
Marianne	**www.marianne2.fr/**
(L')expansion	**www.lexpansion.com/**
(la) Montagne	**www.centrefrance.com/**
La Nouvelle République	**www.lanouvellerepublique.fr/**
(la) Provence	**www.laprovence.com/**
(le) Républicain Lorrain	**www.republicain-lorrain.fr/**
(les) dernières Nouvelles d'Alsace	**www.sdv.fr/dna**
(le) Télégramme	**www.letelegramme.com/**
Nice Matin	**www.nicematin.fr/**
Radio France Outremer	**www.rfo.fr/**
Sud Ouest	**www.sudouest.com/**

Table 11: German sources

Source name	URL
abendblatt	**www.abendblatt.de**
agrarheute	**www.agrarheute.com**
Ärzte Zeitung	**www.aerztezeitung.de**
chinanetz.info	**http://chinanetz.info**
china-observer	**www.china-observer.de**
die-topnews	**www.die-topnews.de**
Edizin	**www.edizin.de**
Food-monitor	**www.food-monitor.de**
Gabot	**www.gabot.de**
krankenkasseninfo	**www.krankenkasseninfo.de**
Leben Ohne Diät	**www.leben-ohne-diaet.de**
paradisi	**www.paradisi.de**
proplanta	**www.proplanta.de**
netdoktor	**www.netdoktor.de**
Short News	**www.shortnews.de**
stimme	**http://stimme.de**
szon	**www.szon.de**
talking food	**www.talkingfood.de**
Taz	**www.taz.de**
TVinfo	**www.tvinfo.de**
UA-BW	**www.untersuchungsämter-bw.de**
verbraucherNews	**www.verbrauchernews.de**
webnews.de	**www.webnews.de**
Wdr	**www.wdr.de**
Yigg	**www.yigg.de**

Table 12: Greek sources

Source name	URL
Eleftherotipia	**www.enet.gr**
Eleftheros tipos	**www.et.gr**
ERT tv channels	**www.ert.gr**
Ethnos	**www.ethnos.gr**
Kathimerini	**www.kathimerini.gr**
Proto Thema	**www.protothema.gr**
Skai TV tv channel	**www.skai.gr/**
Vima	**www.tovima.gr**

Table 13: Italian sources

Source name	URL
l'Espresso	**http://espresso.repubblica.it/**

Table 14: Scientific sources

Source name	URL
BMJ	**http://group.bmj.com/products/journals/**
Bulgarian Science	**http://Bgnauka.com**
Emerging Infectious Diseases	**http://www.cdc.gov/ncidod/EID/index.htm**
Greek Chemist Society	**www.eex.gr**

B. LIST OF NEW CATEGORIES (FOR DISEASES AND CHEMICALS) SUGGESTED TO THE JRC

Table 15: Diseases category

Category	Diseases
	***Bacillus cereus*, *Clostridium perfringens*, *Plesiomonas shigelloides*, *Scombroid*, ciguatera, paralytic shellfish poisoning, neurotoxic shellfish poisoning, amnesic shellfish poisoning, *euV. parahemolyticus*, *V. vulnificus*,*Yersinia enterocolitica*, *Cryptosporidium parvum*, *Cyclospora cayetanensis*, *Entamoeba histolytica*, *Giardia lamblia*, *Isospora belli*, *Toxoplasma gondii*, *Trichinella spiralis*, other viruses (eg, astroviruses, adenoviruses), Monosodium glutamate, Mushroom poisoning**

Table 16: Chemicals category

Category	Diseases
Metals	**Cadmium, chromium, cobalt, heavy metals, iron, lead, lead oxide, manganese, nickel, thorium, tin, uranium, zinc**
Chemical contamination	**Activated carbon; aluminium; aromatic hydrocabons; benzene; benzethonium chloride; bleach; chlorine; chlorophenol; cristalsulphate; detergent; diethylene glycol (DEG); ethylcarbamate; formol (formaldehyde); guaicol; kerosene oil; methanol; methyl isocyanate; monoethylene glycol (MEG); oil-contaminated; oily substance; paraffin; para-toluene sulphonamide; polybutadiene; potassium ferrocyanide; PVC; soda lye; sodium azide; sodium hydroxide; styrene; sulphur; tetrachloroethylene; toluene; tribromoanidole; trichloroethylene; volatile organic compounds**
Industrial contaminants	**1,3-dichloropropanol (1,3-DCP); 2,4-dinitrophenol (DNP); 3-monochlor-1,2-propanediol (3-MCPD); aliphatic hydrocarbons; benzo(a)anthracene; benzo(a)pyrene; bromate; chrysene; cyanuric acid; dioxin-like polychlorobifenyls; dioxins; desinfectants, snitiser; ethyl tertiary buthyl ether (ETBE); heavy polycyclic aromatic hydrocarbons; hydrogen peroxide; indeno(1,2,3-cd)pyrene; melamine; mineral oil; peroxide; petroleum hydrocarbons; polychlorobifenyls; polycyclic aromatic hydrocarbons (PAHs); semicarbazide (SEM)**
Pesticide residues	**1,4-dichlorobenzene; 2-chloroethanol; 4-CPA; acephate; acetamiprid; aldicarb; alpha-HCH; amitraz; azinphos-methyl; bifenthrin; bromide; bromopropylate; bupirimate; buprofezin; captan; carbaryl; carbendazim; carbofuran; carbosulfan; chlormequat; chlorothalonil; chlorpropham; chlorpyriphos; chlorpyriphos-ethyl; chlorpyriphos-methyl; clothianidin; cyfluthrin; cymoxanil; cypermetrin; cyprodinil; DDT; deltamethrin; diazinon; dichlofluanide; dichlorvos; dicloran; dicophol; dicrotophos; dieldron; dimethoate; dimethomorph; diniconazole; diphenylamine; dithiocarbamates; endosulfan; endosulphate; endrin; EPN; esfenvalerate; ethephon; ethion; ethofumesate; ethoprofos; ethylene; dibromide; ethylene oxide; famoxadone; fenarimol; fenhexamid; fenethrothion; fenpropathrin; fenthion; fenvalerate; fluazifop-P; fhidioxonil; flufenoxuron; flusilazoli; folpet; formetanate; HCH; HCH beta; heptachlor; heptenophos; hexachlorobenzene; hexachlorohexane; hexaconazole; hydrogene phosphide; hymexazol; imazalit; indoxacarb; iprodione; isofenphos-methyl; isoprocarb; lambda-cyhalothrin; lindane; lufenuron; magnesium phosphide; malathion; methiocarb; methomyl; methoxyfenozide; methylbromide; monocrotophos; oxamyl; oxydemeton-methyl; paraoxon-methyl; parathion; parathion-methyl; penconazole; pentachlorophenol; permethrin; pesticide resides; phentachloroaniline; phenthoate; phorate; phosalone; phosmet; phosphamidone; piperonylbutoxide; piridaben; pirimicarb; pirimiphos-methyl; prochloraz; procymidone; propamocarb; prophenophos; propiconazole; prothiopos; pyrazophos; pyridaben; pyri-methanil; quinalphos; quintozene; raticide; roxymidone; slug pellets; spinosad; syprodiinil; tebuconazole; tebufen pyrade; tecnazen; terbucarb; tetraconazole methyl; triadimefon; triadimenol; triazophos; triforine; vinclozolin**
Residues of veterinary medicinal products	**Aminoglycosides; antibiotics; amoxicillin; bacterial inhibitor; chloramphenicol; chlortetracycline; ciprofloxacin; clenbuterol; crystal violet; dexamethasone; difloxacin; dihydrostreptomycin; doxycycline; enroflaxin; erythromycin; hormones 19-nortestosterone; androstenedione; dehydroepiandrosterone (DHEA); dexamethasone; medroxy progesterone acetate (MPA); nandrolone; progesterone; testosterone; leucomalachite green; lincomycin; macrolides; malachite green; metronidazole; nitrofuran (metabolites) furaltadone (AMOZ); nitrofuran (metabolites) furazolidone (AOZ); nitrofuran (metabolites) nitrofurantoin (AHD); nitrofuran (metabolites)nitrofurazone**

	(SEM); norfloxacin; oxolinic acid; oxyphenylbutazone; oxytetracycline; penicillin; phenylbutazone; prednizolone; quinolones; ronidazole; streptomycin; sulphachlorpyridazine; sulphadiazine; sulphadimethoxine; sulphamidine; sulphamerazine; sulphametazin; sulphamethoxazole; sulphametoxine

C. CATEGORIES PATTERNS (COMBINATION AND PROXIMITY) FOR SALMONELLA, MELAMINE AND DDT

C.1. Salmonella

C.1.1. Combination of words in related articles

Table 17: List of words contained in three articles related to a Salmonella contamination event occurring between 01/02/2008-31/08/2008. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

Articles	Words	N
CBC News[77]		
	Cereals	**6**
	USA	**5**
	***Salmonella* / salmonellosis**	**4**
	Officials	**3**
	Food; Malt-O-Malt; Minnesota; Recalled; Suspect / Suspected; Tainted	**2**
Hon News[78]		
	Farm (-s); Outbreak	**7**
	Reported / reporters / reporting / reports	**6**
	CDC; Florida; Mexico; *Salmonella* (-*saintpaul*) / salmonellosis	**5**
	Cases; Health; Officials; USA; State (-s)	**4**
	Bacteria (-l); Cases; Contaminant / contamination; Food (-borne); Identification / identified / identify; Ill / illness / illnesses; Increase	**3**
	Infected / infection; Count; Investigation; Source; Strain (-s); Tainted; Team (-s); Teleconference; Texas; Tomato (-es); Victim (-s)	**2**
RedOrbit[79]		
	Outbreak; Mexican	**6**
	Farm; FDA; *Salmonella*; Tomatoes;	**4**
	Pepper (-s); Strain; Official (-s); Nationwide	**3**
	Center (-s); Contaminated; Discovery; Federal; Health; Irrigation; Jalapeno; Produce; Water	**2**

Table 18: List of words cited by three articles at least (N: Number of times the words were cited).

Words	N
Salmonella	**13**
Officials	**10**
Food	**8**
Health	**7**
Case	**6**
FDA; Tainted	**5**
Ill; Investigation	**4**
Sickened	**3**

[77] CBC News (14/04/2008): http://www.cbc.ca/consumer/story/2008/04/14/cereal-recall.html?ref=rss
[78] Hon News (20/06/2008): http://www.hon.ch//News/HSN/616744.html
[79] RedOrbit (31/07/2008): http://www.redorbit.com/news/health/1503785/salmonella_found_in_second_mexican_farm/index.html?source=r_health

C.1.2. Combination of words in related articles in unrelated articles

Table 19: List of words contained in three articles not directly related to a Salmonella contamination event occurring between 01/02/2008-31/08/2008. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

Guardian[80]	
People	**11**
Obese (-looking)/ Obesity / Obesogenic	**10**
Excuse (-s); Obesity; Tory (-MPs) / Tories; Fat / anti-	**9**
Week (-s)	**8**
Party	**6**
Chair /chairman; Diet (-ary) /dietician; Health (-y)	**5**
Goal; Public	**4**
Behaviour; Easy / easily; Politics / political (-par ties); Social / society; Light; Moral;	**3**
Advertised; Away; Back; Biology; Body; Children; Conference; Enjoy; Exercise; Food; Fun; Good; Ice cream; Influence; Labelling; Line; Long / Longer; Mention / mentioned; New; Noodle / Noodles; Offered; Perfect / Perfectly; Popular / Popularity; Pot; Range; Responsibility; Seats; Something; Spot / Spots; Stigmatise / Stigmatising; Support / Supporters; Top (-down); traffic; UK; Vote; Weight (weighing);	**2**
Abcnews-health[81]	
Irradiate / Irradiated / irradiating / Irradiates / Irradiation	**24**
Food (-Drug Administration, -irradiation, -safety) / Foodmakers / Foods	**20**
Say / Saying / Says	**18**
Produce / Producers / Produced / Product / Products	**12**
Kill / killed / Kills	**9**
Bacteria; Beef; Process	**7**
Companies / Company; E. coli; Spinach	**6**
Dose / Doses; Ground; Research; Technology; Test / tested / Testing	**5**
Add / Added; Cost; Facilities / Facility; Green / Greens; Lettuce; Shown / Showed	**4**
Build / Building / Built; Consumers; Fresh; Leafy; Meat; Radiation; Treat / Treating / Treats; Year / years; Considered;	**3**
Allowing / Allows; Bagged (-salad); Chlorine; Eaten; Employee; Handful; Including; Industry; Label / Labels; Leader / Leading; Medical; Outbreak; Plant / Plants; Poultry; President; *Salmonella*; Sell / Sells; Service; Ship / Ships; Spices; Started; Steaks; Step; US; Unit / United; USDA (-ARS); Vegetables; Washes	**2**
Los Angeles Times[82]	
Said	**9**
Minister	**8**
Nulera (-sub, -propelled)	**6**
Official / Officials; Prime	**5**
Centrifuges; Iran; Protesters	**4**
Brazil; Cheese; Dead / Death / Deaths; Defence; Food / Food Inc.; Government; Operate / Operating; Outbreak; Plant; President / Presidential; Program; Year / Years	**3**
Campaign; Canada / Canadian; Country; Democracy; End; Enrichment; Fighter / Fighters; Gas: Hope; Included; Increased / Increases; Massive; Meats; Militant; Moved / Moves; Number; Pakistan; People; Plan; Poisoning; Police; Preparing; Produce / produced; Preparing; Quebec; Reporters / Reports; *Salmonella*; Sanctions; Secured / Security; Spend; Submarine; Tainted; Tear; Thai / Thailand; UN; Use / Used; Week	**2**

[80] Guardian (31/08/2008): http://www.guardian.co.uk/commentisfree/2008/aug/31/conservatives.davidcameron?gusrc=rss&feed=networkfront

[81] Abcnews-health (30/08/2008) : http://abcnews.go.com/Business/story?id=5673165&page=1

[82] Los Angeles Times (30/08/2008) : http://feeds.latimes.com/~r/latimes/news/nationworld/world/~3/378743047/la-fg-briefs30-2008aug30,0,4414025.story

Table 20: List of words cited by three articles at least (N: Number of times the words were cited).

Words	N
Food (-Drug Admin., -irridiation, -makers, -safety) / Foods (-Inc)	**25**
Produce / Producers / Produced / products	**15**
Week	**11**
Health / Healthy Year / Years	**7**
Outbreak	**6**
***Salmonella*; Use / Used**	**5**
Allowance / Allowed / Allowing; Time	**4**
Back (-benchers); Develop / Developing / Development; Helped / Helping; Right / Rights / Right-wing	**3**

C.1.3. Proximity in related and unrelated articles

In related articles, the word "*Salmonella*" appeared in the title most of the times (7/10 cases) and on average within the first 75 words of the text. In unrelated articles, it never appeared in the title and on average within the first 242 words of the text.

C.2. Melamine

C.2.1. Combination of words in related articles

Table 21: List of words contained in three articles related to a melamine contamination event occurring between 01/09/2008-31/10/2008. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

Articles	Words	N
Voanews[83]		
	Melamine	**8**
	China/Chinese, Chinese-made, Northern China	**7**
	Milk -powder, -products, water-down-	**6**
	Chemical	**4**
	Authority (-ies); Cookies; Officials	**3**
	Scandal Products; Netherlands; Levels; Food (-s); Cracker (-s)	**2**
Financial Express[84]		
	Milk (-composition, -constituents, constitution, contamination, protein genes, solids, yield, synthetic-, cow's-)	**22**
	Food (-chain, -industry, -source)	**12**
	Product (-s)/production; China/Chinese	**9**
	Contaminated/contamination	**7**
	Protein (-s, -content); Safe/safety	**6**
	India; Casein (-s)	**4**
	Adulterated/adulteration; market; melamine; dairy (-cattle, -improvement	**3**
	Standard (-s); Level (-s); Norms; Gene coding/genetic make up; Government/Governing; Incident (-s); Trade	**2**
CNN Asia		

[83]Voanews (30/09/2008): http://www.voanews.com/english/2008-09-30-voa51.cfm?rss=asia
[84] Financial Express (01/10/2008): URL not available

[85]	**Melamine (-tainted)**	**11**
	Milk (-contaminated, Lipton - tea, - tea, tainted -); China/Chinese (-dairy companies, -made, -products)	**10**
	Authorities; Product (-s)	**5**
	Company (-ies)	**4**
	Candy (-products, -maker, -white rabbit cream); Scandal; Hong-Kong	**3**
	Trace (-s)/amounts; Taiwan; Killed/killing; FDA; Exports/exported; Chemical	**2**

Table 22: List of words cited by three articles at least (N: Number of times the words were cited).

Words	N
Milk	**38**
China	**26**
Melamine	**22**
Product/production	**16**
Food	**15**
Authorities	**9**
Scandal; Chemical; Companies	**6**
Level	**5**
Consumed/consumers; Ban/banned; Sickened	**3**

C.2.2. Combination of words in unrelated articles

Table 23: List of words contained in three articles not directly related to a melamine contamination event occurring between 01/06/2008-31/08/2008. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

[85] CNN Asia (30/09/2008): http://edition.cnn.com/2008/WORLD/asiapcf/09/30/china.tainted.milk/index.html?eref=rss_topstories

Articles	Words	N
Khaleej Times[86]		
	IPIC	**4**
	Borealis Fertilizer/fertilizer; Uzbek / Uzbekistan	**3**
	Complex; Marketing / markets; MOU; Partnering; Plant (-nutrients); Study; Uzkimyosaneat	**2**
Kuensel Online[87]		
	Bhutan (-Eastern, -experts) / Buthanese	**13**
	Bangladesh; Export (-s) / Exporters; Trade (-meeting, -relations, -route)	**8**
	Country (-ies)	**5**
	FTA Free Trade Agreement; Agreement /agreed	**4**
	Goods; ort; AARC; amabil	**3**
	Samdrupjongkhar; Products; Nations; Imported; India; Items	**2**
Austrian Times[88]		
	Producing /production / Products	**7**
	Melamine	**5**
	AMI Agrolinz Melamine International	**4**
	Plant; Material (-s)	**3**
	Raw; Italian; Company; Borealis	**2**

Table 24: List of words cited by three articles at least (N: Number of times the words were cited).

Words	N
Melamine	**7**

C.2.3. Proximity in related and unrelated articles

In related articles, only half of the articles (5/10) had the word “melamine” in the title and it was cited in the text on average within the first 37 words. In unrelated articles, the word “melamine” never appeared in the title and on average within the first 5296 words of the text.

C.3. DDT

C.3.1. Combination of words in related articles

Table 25: List of words contained in three articles related to a DDT contamination event occurring between 01/01/2007-31/03/2009. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

[86] Khaleej Times (22/08/2009): http://www.khaleejtimes.com/DisplayArticle.asp?xfile=/data/business/2008/August/business_August707.xml&section=business

[87] Kuensel Online (18/08/2009): http://www.kuenselonline.com/modules.php?name=News&file=article&sid=10992

[88] Asutrian Times (09/07/2008): http://www.austriantimes.at//index.php?id=7422

Articles	Words	N
ChinaPost[89]		
	Tea /teahouses / teabag	**15**
	Drink (-s, -ers); Lead	**6**
	Sample (-s); Council	**5**
	Long; test (-ed)	**4**
	Consumer; Hong Kong; leaves;pesticide	**3**
	Chinese; food; higher / higest; levels, limit	**2**
Denvers Post[90]		
	Food (-s)	**6**
	Chemical (-s)	**4**
	Asia (-n); Hanoi; noodle (-s)	**3**
	Bowl; export (–ed); formaldehyde; ítems; pho; problem; producers / product; tainted; Vientnam	**2**
Ghanaweb[91]		
	Hospital	**5**
	Banku;	**3**
	Akorabuokrom; family; maize; medical; Nyinahin; wife	**2**

Table 26: List of words cited by three articles at least (N: Number of times the words were cited).

Words	N
DDT, Food	**9**

C.3.2. Combination of words in unrelated articles

Table 27: List of words contained in three articles not directly related to a DDT contamination event occurring between 01/01/2007-31/03/2009. Composed words are listed into brackets and derived words are cited after a slash symbol (N: Number of times the words were cited).

[89] ChinaPost (16/01/2009): http://medisys.jrc.it/medisys/dynamic?language=all&page=1&edition=searchresults&option=&querytype=advanced&all=tea+teahouses+teabag+test+chinese+limit+highest+DDT&atLeast=&exact=&not=&dateFrom=01%2F01%2F2007&dateTo=09%2F05%2F2009&searchcategory=&cal_mon0=3&cal_year0=_&cal_mon1=4&cal_year1=_&restrict=lang&lang=en

[90] Denvers Post (17/06/2007): http://www.denverpost.com/nationworld/ci_6159762

[91] Ghanaweb (15/08/2008): http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=148486

	Words	N
Ghanaweb[92]		
	Ants	**5**
	Farmer (-s)	**4**
	Area; vegetable (-s)	**3**
	Army; attack; crops; holes; lawer; okro; sub-chief; tomato	**2**
CBC[93]		
	Mercury	**9**
	Contamination; levels;	**5**
	Researcher (-s)	**4**
	Food (-s)	**3**
	Arctic; increase; ocean (-s); PCBs; pollutants; problems; studies; traditional; work (-shop)	**2**
News com Au[94]		
	Reef	**6**
	River (-s)	**5**
	Barrier; crabs; Great; insecticides; park; pesticides; pollutants ; study	**3**
	Authority; Barron; Burnett; Cancer; Contaminated; DDT; dieldrin; executive; fat (-soluble); Fitzroy; Major; Marine; Mud; Organochlorine; Pioneer; Quality; Sediment; Water	**2**

Table 28: List of words cited by three articles at least (N: Number of times the words were cited).

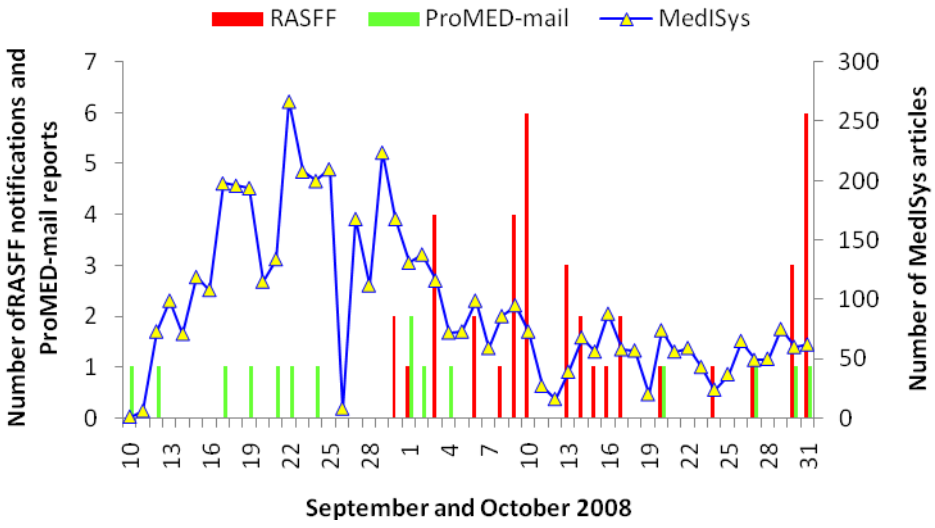
Words	N
Food	**5**
DDT	**4**

C.3.3. Proximity in related and unrelated articles

The proximity level of the word “DDT” could not be analyzed thoroughly because of the low number of related articles available (n=4). In the related articles, the word “DDT” appeared either in the title (2/4 cases) or in the text (2/4 cases) within the first 162 words. In the unrelated articles, the same pattern applied. Some articles had the word “DDT” in the title (3/10) or in the text (7/10) within, on average, the first 252 words.

[92] Ghanaweb (26/07/2008): http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=147467
[93] CBC (05/10/2007): http://www.cbc.ca/canada/north/story/2007/10/05/northern-food.html?ref=rss
[94] News com Aus (11/08/2007): http://www.news.com.au/story/0,23599,22226193-29277,00.html

D. NUMBER OF RASFF NOTIFICATIONS, MEDISYS ARTICLES AND PROMED-MAIL REPORTS RETREIVED OVER THE PERIODS OF SEARCH FOR EACH CASE STUDY AND LIST OF REFERENCES



m

Figure 9: Melamine in food and feed on RASFF, MedISys, and ProMED-mail in September and October 2008.

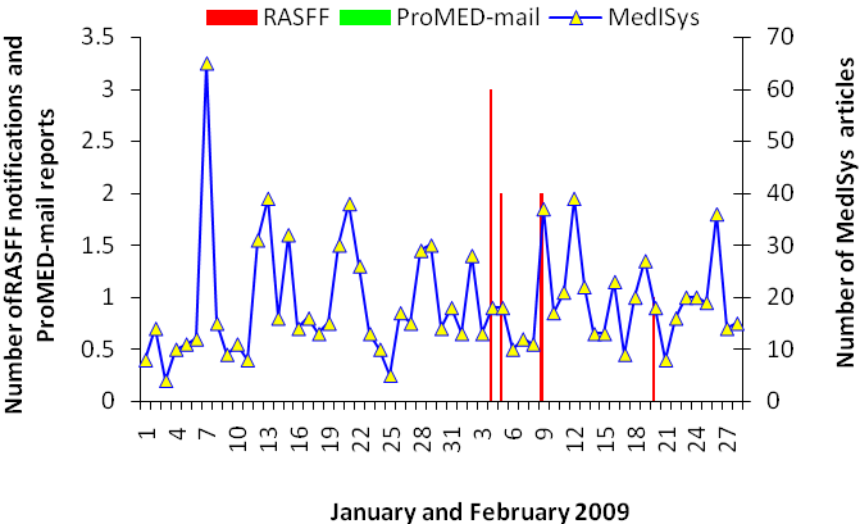


Figure 10: Aluminium in food and feed on RASFF, MedISys, and ProMED-mail in January and February 2009.

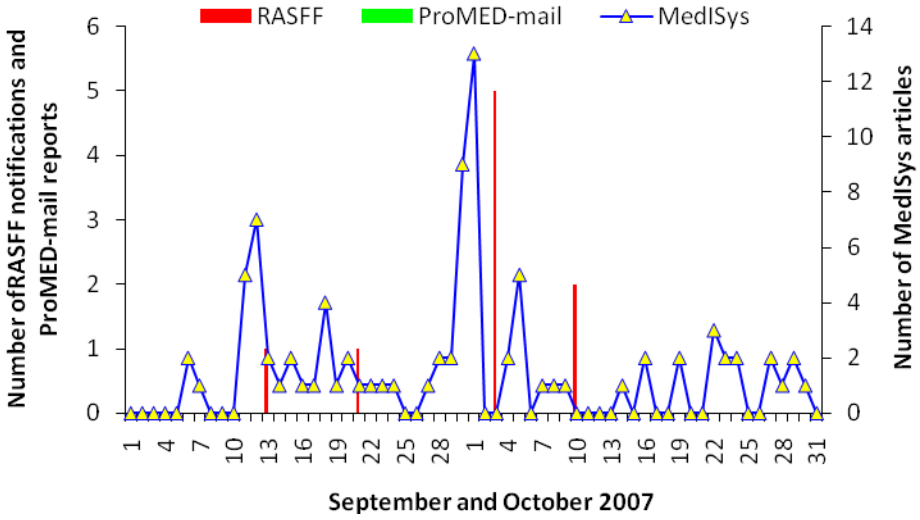


Figure 11: Dioxins in food and feed on RASFF, MedISys, and ProMED-mail in September and October 2007.

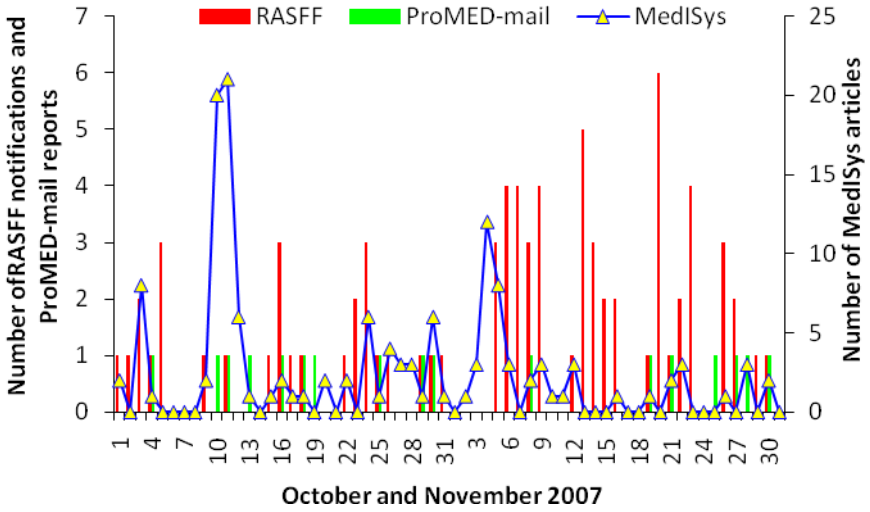


Figure 12: *Salmonella* in food and feed on RASFF, MedISys, and ProMED-mail in October and November 2007.

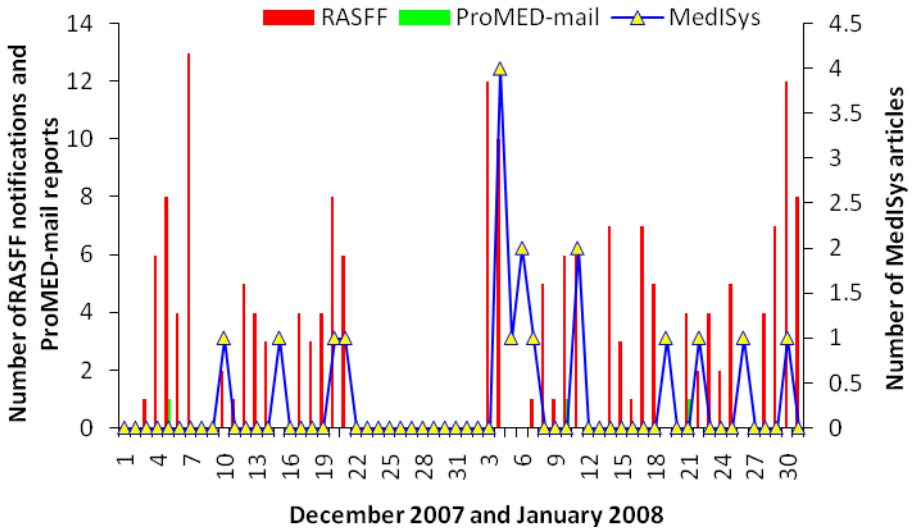


Figure 13: Mycotoxins in food and feed on RASFF, MedISys, and ProMED-mail in December 2007 and January 2008.

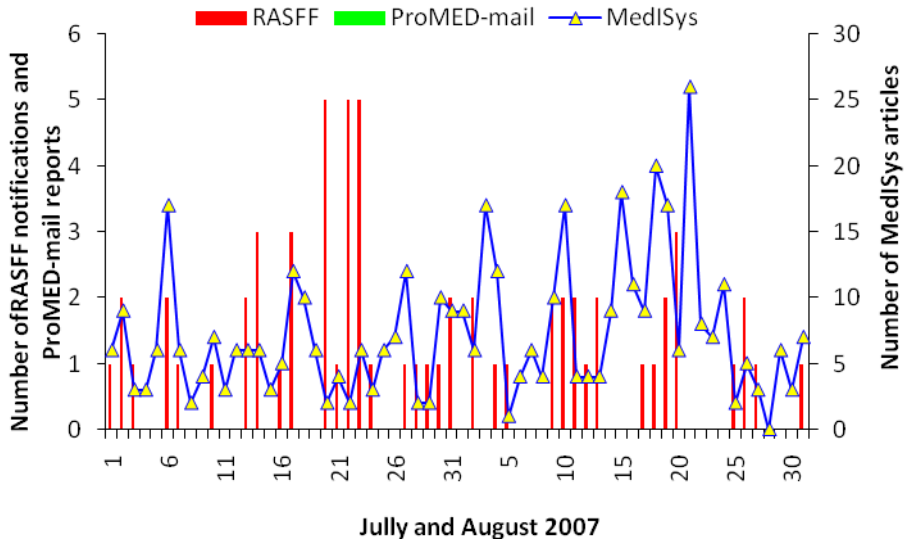


Figure 14: Heavy metals in food and feed on RASFF, MedISys, and ProMED-mail in July 2007 and August 2007.

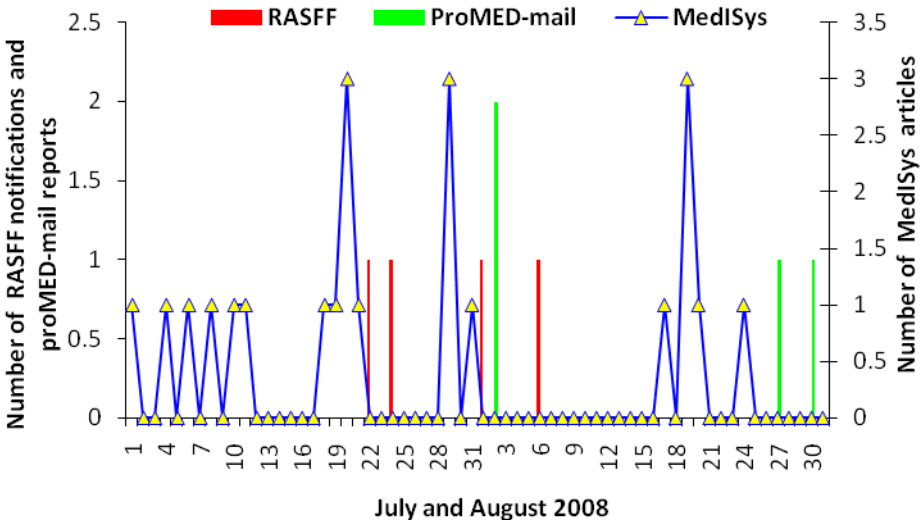


Figure 15: *Vibrio cholera* in food and feed on RASFF, MedISys, and ProMED-mail in July and August 2008.

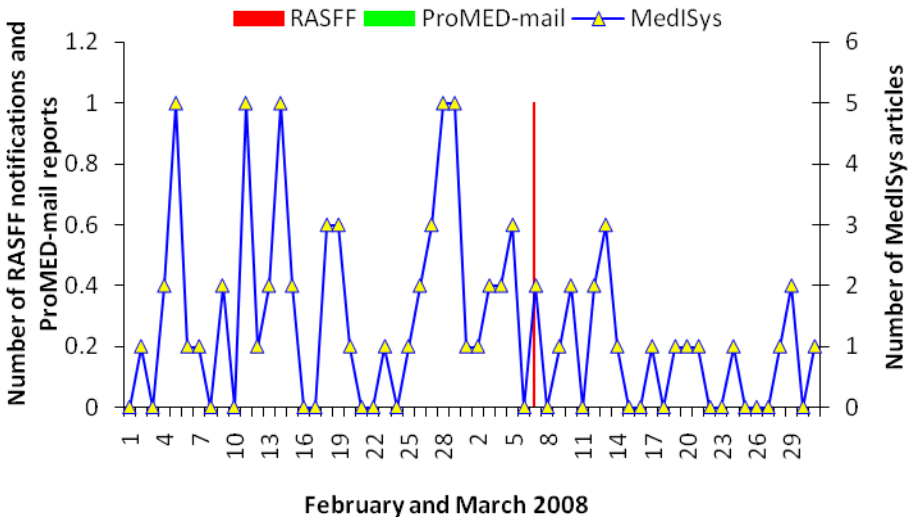


Figure 16: DDT in food and feed on RASFF, MedISys, and ProMED-mail in February and March 2008.

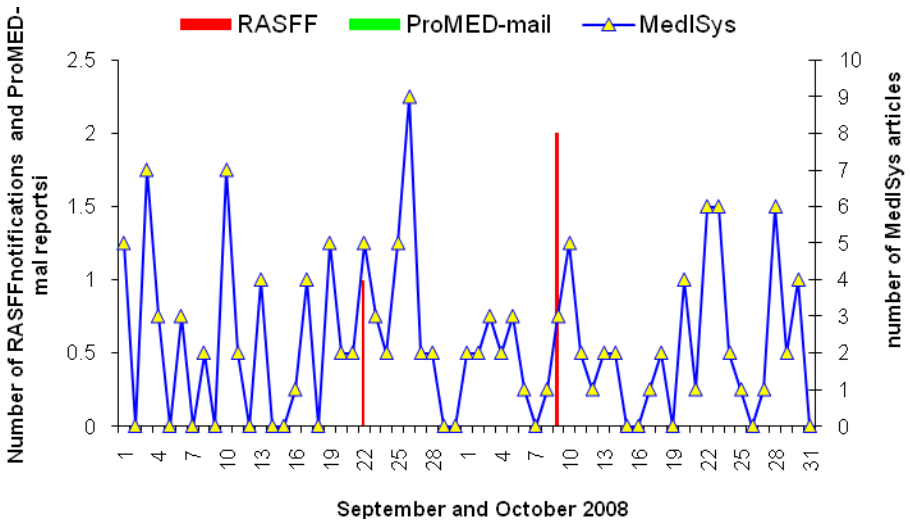


Figure 17: Radioactivity in food on RASFF, MedISys, and ProMED-mail in September and October 2008.

E. MEDIA REFERENCES

MedISys

5. http://www.cbc.ca/consumer/story/2007/10/04/squares-salmonella.html?ref=rss
6. http://www.msnbc.msn.com/id/21209566/
7. http://www.iol.co.za/index.php?from=rss_World&set_id=1&click_id=3&art_id=nw20071027145346701C165611http://www.iol.co.za/index.php?from=rss_World&set_id=1&click_id=3&art_id=nw20071027145346701C165611
8. http://article.wn.com/view/2007/10/26/Salmonella_sprouts
9. http://www.dailymail.co.uk/news/article-490801/Chocolates-recalled-nuts-contaminated-salmonella.html
10. http://www.mirror.co.uk/news/latest/2007/10/30/salmonella-shuts-fox-s-chocolate-plant-89520-20035096/
11. http://www.usatoday.com/news/washington/2007-11-03-usda-meat_N.htm?csp=34
16. http://www.theglobeandmail.com/servlet/Page/document/v5/content/subscribe?user_URL=http://www.theglobeandmail.com%2Fservlet%2Fstory%2FLAC.20071121.SICK21TOR%2FTPStory%2FTPNational%2F%3Fpage%3Drss%26id%3DGAM.20071121.SICK21TOR&ord=54349978&brand=theglobeandmail&force_login=true
17. http://www.financialexpress.com/news/Spices-exports-to-EU-may-come-under-check-for-aflatoxins/266936
18. http://www.cyprus-mail.com/news/main.php?id=36645&cat_id=1
19. http://www.usatoday.com/money/industries/2008-01-04-diamond-dog-settlement_N.htm?csp=34
35. http://www.cbc.ca/canada/british-columbia/story/2007/08/13/bc-uranium.html?ref=rss
36. http://www.int.iol.co.za/index.php?from=rss_South%20Africa&set_id=1&click_id=13&art_id=vn20070706042703356C499181
37. http://www.nst.com.my/Current_News/NST/Friday/Advanced/20070628153115/Article/indexpull_html
39. http://www.usatoday.com/news/health/2008-09-11-tainted-formula_N.htm?csp=34
40. http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=150524
41. http://www.nst.com.my/Current_News/NST/Thursday/Frontpage/2359191/Article/index_html
42. http://news.gc.ca/web/view/en/index.jsp?articleid=421779
57. http://www.flex-news-food.com/pages/21720/Dietary/UK/uk-study-shows-levels-metals-elements-food-stable-lower-2000.html
59. http://www.earthtimes.org/articles/show/dominickrsquos-recalls-thai-noodle-salad-sold-at-store-salad-bars,732952.shtml
71. http://www.irinnews.org/report.aspx?ReportID=79493
72. http://www.thenews.com.pk/daily_detail.asp?id=121591
76. http://www.dailytimes.com.pk/default.asp?page=2008/02/25/story_25-2-2008_pg12_11
79. http://www.novayagazeta.com.ru/news/300178.html

ProMED-mail

12. ProMED. PRO/AH/EDR> Salmonellosis, frozen poultry pie - USA (multistate): notification. 2007, 10-10: 20071010.3319. http://www.promed.org. Accessed 10-10-2007.
14. ProMED. PRO/AH/EDR> Foodborne illness, hummus - Jordan: (Jerash). 2007: 29-10: 20071029.3510. http://www.promed.org. Accessed 29-10-2007.
15. ProMED. PRO/AH/EDR> Salmonellosis, university - Canada: (ON). 2007, 21-11: 20071121.3770. http://www.promed.org. Accessed 21-11-2007.
20. ProMED. PRO/EDR> Toxic fungus, rice - Brazil: (Maranhao). 2008, 08-01: 20080108.0101. http://www.promed.org. Accessed 27-12-2007.
21. ProMED. PRO/EDR> Toxic fungus, rice - Brazil: (Maranhao). 2008, 08-01: 20080108.0101. http://www.promed.org. Accessed 27-12-2007.

22. ProMED. PRO/AH> Undiagnosed deaths, camels - Saudi Arabia (10): Salinomycin conf. 2008, 10-01: 20080110.0138. http://www.promed.org. Accessed 10-01-2007.
23. ProMED. PRO/AH> Undiagnosed deaths, camel - Saudi Arabia (02): mycotoxins. 2008, 21-01: 20080121.0262. http://www.promed.org. Accessed 20-01-2008.
38. ProMED. PRO/EDR> Infant kidney stones – China: Gansu, milk powder susp., RFI. 2008, 10-09: 20080910.2828. http://www.promed.org. Accessed 10-09-2008.
45. ProMED. PRO/EDR> Infant kidney stones - China (04): WHO, international recall. 2008, 19-09: 20080919.2951. http://www.promed.org. Accessed 18-09-2008.
47. ProMED.PRO/EDR > Melamine contaminated food products – Worldwide ex China. 2008, 02-10: 20081022.3107. http://www.promed.org. Accessed 30-10-2008.
49. ProMED. PRO/EDR> Infant kidney stones - China (07): HK, Macau. 2008, 24-09: 20080924.3024. http://www.promed.org. Accessed 24-09-2008.
51. ProMED. PRO/EDR> Melamine contaminated food products (02): Worldwide ex China. 2008, 04-10: 200881004.3129. http://www.promed.org. Accessed 02-10-2008.
53. ProMED. PRO/AH/EDR> Contaminated rice – Japan. 2008, 25-09: 20080925.3033. http://www.promed.org. Accessed 24-09-2008.
56. ProMED. PRO/EDR > Melamine contaminated food producst (03): Worldwide ex China. 2008, 20-10: 20081020.3324. http://www.promed.org. Accessed 08-10-2008.
73. ProMED. PRO/EDR> Food poisoning, frozen fish - Norway ex Viet Nam: RFI. 2008, 27-07. 20080727.2298. http://www.promed.org. Accessed 27-07-2008.
74. ProMED. PRO/AH/EDR> *Vibrio parahaemolyticus*, oysters - USA: (MD ex NJ). 2008, 21-08. 20080821.2608. http://www.promed.org. Accessed 19-08-2008.
75. ProMED. PRO/EDR> Azaspiracid toxins, shellfish - USA ex Ireland: notification. 2008, 21-08. 20080826.2661. http://www.promed.org. Accessed 15-08-2008.

GLOSSARY AND ABBREVIATIONS

AF	Advisory Forum
Notification	An notification defines the characteristics that a news source must have in order to be identified and categorized.
BioCaster	Global health Monitor, a research project at the National Institute of Informatics based on intelligence text mining for infectious disease
Category	A category defines the characteristics that a news source must contain in order to be identified and categorized, e.g. "Malaria" is a category.
Combination	A combination is a set of two or more lists of patterns. (“OR” and/or “NOT” lists).
CIDRAP	Center for Infectious Disease Research and Policy
ECDC	European Centre for Disease Prevention and Control
DG-COMM	Directorate General Communication
EFSA	European Food Safety Authority
EMRISK	Emerging Risk project
ESCO-WG	EFSA Scientific Cooperation working group
EMM	Europe Media Monitor
EpiSPIDER	an open source application in public health that processes and visualizes free text information
FAS	Federation of American Scientists
GPHIN	Global Public Health Intelligence Network Global Health Monitor
HealthMap	Global Disease Notification Map
IPSC	Institute for the Protection and Security of the Citizen
IT	Information Technologies
JRC	Joint Research Center
MedISys	Medical Information System
OIE	Office International des Epizooties - World Organization for Animal Health
Pattern	A pattern can contain alpha numeric characters as well as “%” which denote any number of

	characters and "_" which means only one character.
ProMEd	Program for Monitoring Emerging Diseases from the International Society for Infectious Diseases
Proximity	(only combinations) describes a word context size within which the combination terms have to occur.
PubMED	a service of the U.S. National Library of Medicine and the National Institutes of Health
RASFF	Rapid Notification System for Food and Feed
RNS	Rapid News Service
SC	Scientific Committee
SID	Society for Infectious Diseases
Sources (media Sources)	News Sources e.g. news websites, rss feeds… from providers like BBC, CNN etc.
Threshold	The threshold is the minimum sum of weights matched patterns required to trigger the notification.
VWA	Food and Consumer Product Safety Authority
Weight	The weight (only word-weight list) of a pattern is the value summed when patterns are matched. The sum of the weights of the matched patterns is compared to the threshold value to determine whether or not to trigger the notification.
WHO-GOARN	World Health Organization of the Global Outbreak Alert and Response Network