Far from the Maddening Crowd: Integrating Collaborative Translation Technologies into Healthcare Services in the Developing World

Erin Lyons
University of Maryland

ABSTRACT

Crowdsourced and collaborative translation technologies have been at the centre of a heated debate in the translation industry in recent years, as questions have been raised regarding labour practices, the widespread integration of machine translation (MT) as well as concerns regarding quality and professional practices. However, despite the criticism of this emergent technology, the union of collaborative translation platforms and mobile communication technology has bridged a knowledge, resource and communication gap in the developing world, allowing healthcare and medical services to be re-imagined to reach a previously unimaginable community – often instantaneously. The rich data network supplied by mobile phones, when combined with automated data integration, can now be merged with translation services to contribute to initiatives, such as slowing the spread of malaria or stopping stock-outs of life-saving drugs at local clinics. We will take a closer look of the role of translation (machine translation, human translation and controlled language) in some of the leading crowdsourced translation applications, how translators bridge the gap between algorithm and on-the-ground communication and the implications for the development of "lite", mobile-ready versions of CAT tools and TMs.

1. Introduction

Crowdsourced and collaborative translation technologies have been at the centre of a heated debate in the translation industry in recent years, as questions have been raised regarding labour practices, the widespread integration of machine translation (MT), as well as concerns regarding quality and professional practices. However, despite the criticism of this emergent technology, the union of collaborative translation platforms and mobile communication technology has bridged a knowledge, resource and communication gap in the developing world, allowing healthcare and medical services to be re-imagined to reach a previously unimaginable community – often instantaneously. The rich data network supplied by mobile phones, when combined with automated data integration, can now be merged with translation services to contribute to initiatives, such as slowing the spread of malaria or stopping

stock-outs of life-saving drugs at local clinics. We will take a closer look of the role of translation (machine translation, human translation and controlled language) in some of the leading crowdsourced translation applications, how translators bridge the gap between algorithm and on-the-ground communication and the implications for the development of "lite", mobile-ready versions of CAT tools and TMs.

2. Computers are Ineffective, People are Inefficient

Crowdsourced and other collaborative translation methods occupy a unique position, regardless of the computational model used, populating the spectrum between machine translation (MT), professional human translation (HT) and open-source innovation, as illustrated in Figure 1. Scalable crowdsourced translation and information processing is still a fairly nascent technological application undergoing continual development, as the cutting edge of the data science industry trickles down and mobile services and capabilities continue to expand – in even the most remote regions of the world. This continual proliferation of mobile and Web-based technology means that speakers of more than 5,000 languages now have real-time access to data and voice communication (Crowley and Chan 2011).

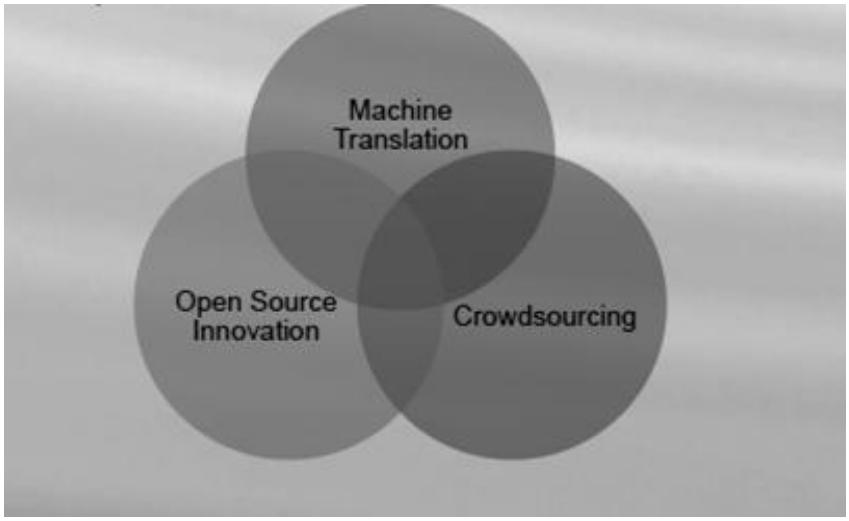

Figure 1: Computation models rely on a combination of machine translation, open source innovation and crowdsourcing to populate the spectrum between professional human translation and machine translation

Yet, despite the myriad of revolutions in information technology, there has been a gap between technological innovation and the ability to process and understand large amounts of data to support linguistically diverse populations. The volume and velocity of data has overwhelmed the same technology that had been the driver behind this dynamic. Yet, with about 85.5% of the world subscribing to mobile services (versus about 33% using the Internet) (Lorentz 2004), the penetration of mobile phone services and Short Messaging Service (SMS), combined with crowdsourcing and micro-tasking have emerged as forms of collective intelligence in the field of translation, adding value through shared data communication, and transforming technology into a means of including the world's many under-represented languages. Consequently, mobile technology provides robust architecture that is virtually ubiquitous, allowing crowdsourced translation to become the missing link in a globalised world where

linguistic variation is the norm, enabling global collaboration, humanitarian development and action.

3. On-the-Ground Applications of a Nascent Technology

The "crowd" has emerged as a buzzword in recent years, as virtually all problems – from start-up financing, to searching for missing persons, to voting for Coca-Cola World Cup promotions (Joseph 2014) – seem to be easily resolved via the input of voices from the collective crowd. Crowdsourced translation and the micro-outsourcing of translation- and language-related tasks has served as the momentum behind a wave of inclusion, wherein translation has made our global, multilingual and multicultural society just that much smaller and more closely intertwined.

The role of technology, when combined with the power of translation – whether outsourced to local crowdsourcing translators or combined with professional human translation – has the power to open the flood gates to new voices and strengthen humanitarian action and manage big data. Harnessing digital communication platforms through crowdsourcing translation is the missing link that will become the driver behind optimising and extending humanitarian and social aid in years to come. An examination of four ground-breaking applications of the technology provides insight into current capabilities and achievements when these innovations are applied to cross-lingual communication, as well as the areas of weakness requiring further development and integration. As better synthesised by Robert Munro, the computational linguist and data scientist behind Mission 4636, the first application of crowdsourced translation for humanitarian relief during the 2010 Haiti earthquake, "The future of how we talk to each other is changing and even crowdsourced translation looks nothing now like it did 12 months ago. Who knows how it will look in just a few more years?" (Junglelightspeed 2014).

4. Real-Time Intelligence for Healthcare Services in the Developing World

In recent crowdsourced translation campaigns for healthcare initiatives (Stop Stock-Outs, ProMED-mail, MalariaSpot, etc.), human-centric multilingual systems have been designed around major and minor language processing systems to engage users and non-professional translators to bridge socio-linguistic gaps in framing and contextualising reports from otherwise geographically and/or linguistically excluded populations. Collaborative human translation and interaction add precision and quality control that are otherwise lacking in machine translation, particularly in the context of low-density languages (Eidelman 2011).

5. Aggregating Multilingual Information to Stop Stock-Outs

The Stop Stock-Outs campaign is an initiative based in Kenya, Uganda, Malawi, Zambia and Zimbabwe to gather data from citizens on stock-outs of ten essential medications (first-line antimalarials, zinc, benzathine penicillin, first-line antiretrovirals, metronidazole, ciprofloxacin, amoxicillin, ceftriaxone, cotrimoxazole, ORS) at public health facilities. By gathering data via SMS and aggregating the information on interactive maps, the campaign has raised awareness about health rights and access to essential medicines and fostered a culture of greater community engagement and institutional accountability in the implementation of public health policy.

The Stop Stock-Outs campaign was born out of an epidemic sweeping across many regions of Africa, wherein pharmacies and health centres temporarily had no medicine on the shelf, often lasting several weeks. For example, in Uganda, prior to the 2009 launch of the initiative to combat these supply shortages, only 45.7% of public facilities had a basket of 28 essential medicines. In 2007-2008, stock-outs at public facilities averaged 72.9 days per year, as a result of poor funding, ineffective coordination or drug procurement and distribution, as well as due to gaps in management and pilferage (Medicines 2010).

Armed with the data, activists collaborated with Ushahidi and FrontlineSMS to create a project to map the availability of essential medicines at public health facilities in several African nations in real-time, supplying public and private health centres with vital information, while holding them accountable. By gaining citizen support and participation in the project, through grass-roots campaigns and local media, the campaign was also able to fight the resistance and denial of local officials faced early-on regarding the very existence of the pervasive and life-threatening issue.

In fact, the Kenyan Ministry of Medical Services released a press statement stating, "There are no stock-outs of essential drugs experienced in the country as reported by the media" (Ole Kiyiapi 2009). However, 5 months after the Stop Stock-Outs campaign was launched, including additional momentum gained during "Pill Check Week" (22 - 26 June 2009), garnering the attention of international media because of the innovative use of emerging technology in the developing world, the Minister of Medical Services, Prof. Peter Anyang' Nyong'o admitted on a Kiswahili radio station that this was indeed a potentially lethal problem threatening public health facilities in Kenya and, furthermore, he confirmed to lawmakers that additional funding would be made available to purchase essential medicines and pre-empt further stock-outs (Stopstockouts 2010).

In practice, Stop Stock-Outs was aimed at checking stock levels of essential medicines. These included first-line anti-malarials, zinc tablets, penicillin, and first-line anti-retrovirals (ARVs) to treat HIV/AIDS. All of these medicines are essential in varying degrees to fighting disease and illness, and widely used in the countries where the project was carried out. After visiting clinics and pharmacies, activist-participants used their mobile phones to report their results using structured, coded SMS messages: "*x, y, z*" – where *x* represented the country code (Kenya, Malawi, Uganda, Zambia or Zimbabwe), *y* the district or city and *z* the medicine that was found to be out of stock. The messages were then received by FrontlineSMS, a free open source software, which would then run an automatic script to validate the data before sending it over the Internet to a Ushahidi-powered website. From there, the results could be automatically aggregated and visually displayed on an interactive map, as seen in Figure 2.

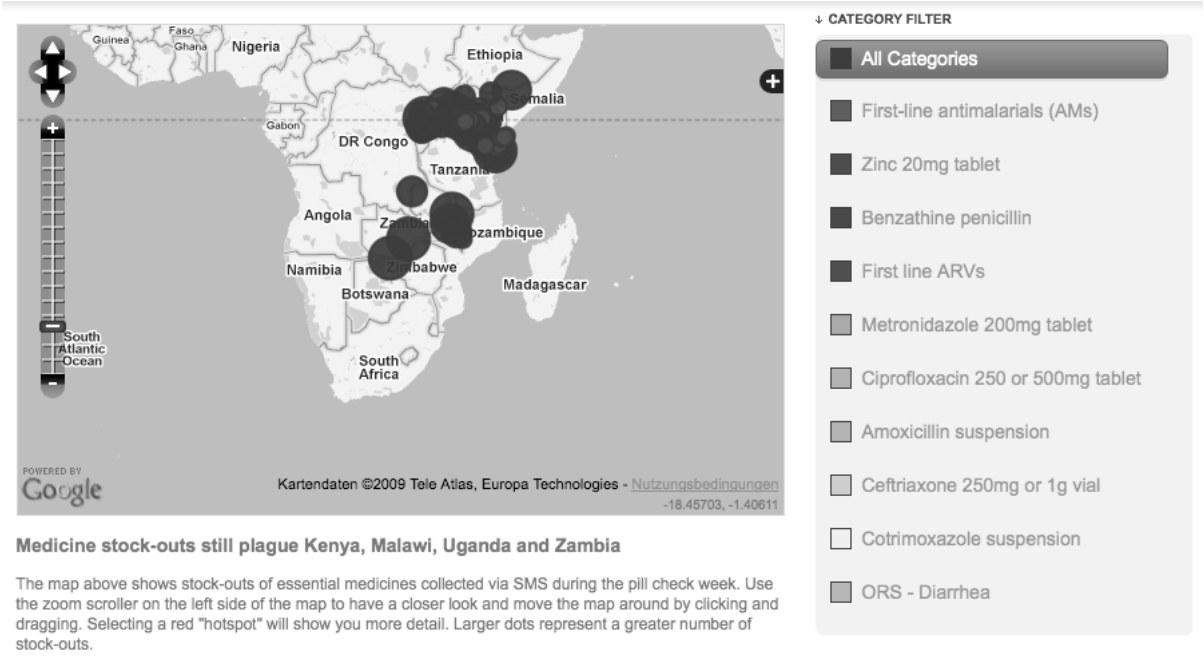

Figure 2: The Stop Stock-Outs campaign started in Kenya and spread to Uganda, Malawi and Zambia, gathering citizen data on the stock-outs of ten essential medications at public health facilities (Stopstockouts 2010)

In this project, while crowdsourcing was used to engage the participation of locals in the initiative, the encoded numerical messages served as a workaround to the problem of engaging a population speaking a variety of local languages. The project organisers determined that the availability of real-time information on medicine inventory superseded the need to gather more complex multilingual data from citizens. Instead, information on the initiative could be translated for local health clinics, explaining how to use the service and input data via a mobile network.

By implementing this translation strategy upstream from the crowdsourcing, crowdmapping and data analysis steps, the linguistic barrier was removed, along with the challenge of sourcing a network of on-call crowdsourcing translators. Although, this encoded system also removed the possibility for two-way communication, framing the data around extremely specific data points. Furthermore, since only numerical data was gathered according to a key, the Stop Stock-Outs programme was somewhat error-prone, as a typographical error, too few or too many input numbers or a failure to adhere to the requested code order could potentially generate erroneous and misleading data. Ultimately, the campaign, combining off-line translation and multi-lingual community outreach with crowdsourcing mobile data collection was effective in obtaining real-time reports from those on the ground, thereby including a cross-lingual population on the ground that is sometimes missed with more targeted online viral marketing campaigns.

6. ProMED-mail: An Adaptable Health Initiative with Far-Reaching Potential

Another crowdsourcing-translation initiative launched in the healthcare sector, serving to deliver real-time intelligence on a range of emerging infectious diseases and outbreaks is ProMED-mail. The completely human-based information collection programme, established in 1994, uses reports gathered from local media, experts and eyewitness reports, which are then processed and translated by humans for systems in English, French, Russian and Spanish. This

crowdsourced translation application for global public health differs from many related projects in its ability to organise multilingual open-source data and in its ability to offers users complete and unfiltered control of searching the aggregated data, rather than framing and constructing fixed analysis tools (Madoff 2004).

In the simplest of terms, ProMED-mail is an e-mail list used to gather data and provide early warnings of disease outbreaks around the world 7 days a week, 24 hours a day. Whereas World Health Organization surveillance reports require official clearance before posting, and are then only posted in French and English, ProMED-mail is able to pick up and disseminate reports over the Internet much sooner, through its network of global multilingual moderators (Lyon 2012). Moderators start by triaging reports in their own language(s), rejecting them or forwarding them to subject moderators. The subject moderators then edit the reports and add commentary, subsequently returning the report to the top moderator, who assigns a colour-coded level of urgency (green, yellow or red – the most urgent) (Cowen *et al*. 2006). Green reports are published within 24 hours, while red and yellow reports receive expedited review. On a typical day, 7 reports are published: 1 red, 1 yellow and 5 green (Madoff 2004).

Since disease and epidemics know no borders, ProMED-mail runs simultaneous language-specific regional networks in Portuguese, Spanish, Russian, English for Southeast Asia, French for Francophone Africa and English for Anglophone Africa. Local languages are translated by regional specialists in infectious diseases, epidemiology and public health into the official language of the local ProMED-mail network. Then, particularly emergent or cross-border reports are translated by a moderator-expert for the main English-language site. Based on the urgency of these reports and the availability of further emerging information, the main English-language page and the local network pages may report translations with slight variations, and may also have been edited to highlight particularly pertinent information for a more global or more local audience. One such example, regarding the recent outbreak of the Ebola virus in West Africa shows two similar, but somewhat divergent reports on the Ebola outbreak, as of 7th April 2014, as per the Francophone and Anglophone Africa pages:

> A cumulative total of 151 clinically compatible cases of Ebola virus disease (EVD) including 95 deaths had officially been reported from Guinea. 65 samples were tested of which 34 tested positive for Ebola.
>
> *La Guinée avait notifié officiellement un total cumulé de 151 cas cliniquement compatibles, incluant 95 décès. Les investigations se poursuivent au laboratoire de l'Institut Pasteur de Dakar à Conakry (65 échantillons analysés, dont 34 ont été positifs à la PCR pour le virus Ebola) et au Laboratoire mobile de l'Union européenne (EMLab) qui a mis en place une équipe à Guékédou (36 échantillons analysés/20 positifs)* [Guinea officially reported a cumulative total of 151 clinically compatible cases, including 95 deaths. Investigations are continuing in the laboratory of the Dakar Pasteur Institute in Conakry (65 tested samples, of which 34 are positive from PCR for the Ebola virus) and in the European Union Mobile Laboratory (EMLab), which has dispatched a team to Guékédou (36 tested samples/20 positive).] (ProMED-mail 2014).

Rather than providing perfect, linear translations of cross-national reports in its epidemiological surveillance, ProMED-mail focuses on rapid-fire, cross-border, multilingual communication, with mapping capabilities and logical information filters (Lyon *et al.* 2011). It is a sensitive, but not necessarily specific system, using local scientific and linguistic intelligence to pinpoint outbreak vulnerabilities, which may in turn be visually mapped and processed, as shown in Figure 3.

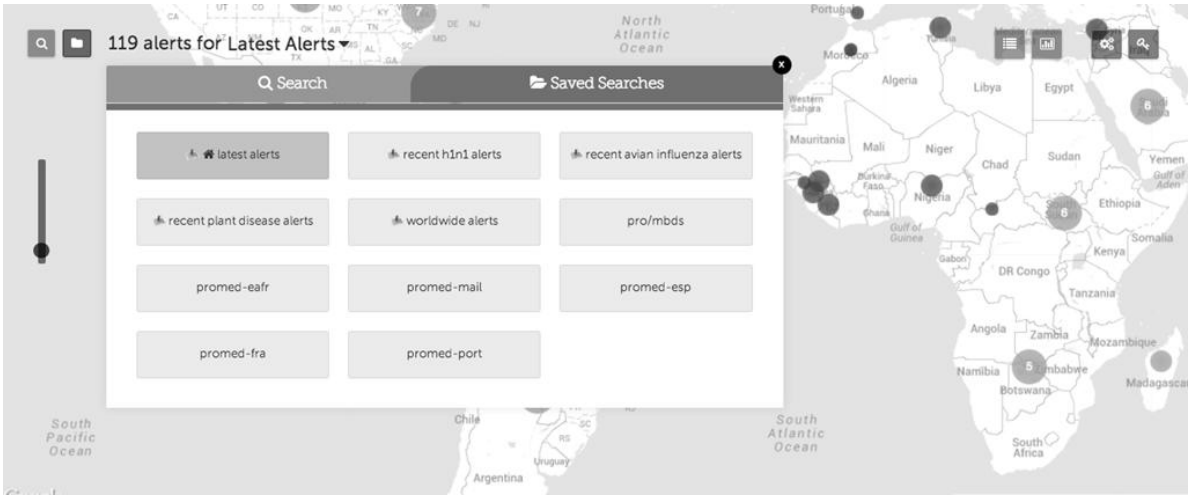

Figure 3: ProMED-mail delivers real-time intelligence on emerging diseases and outbreaks, based on reports from local media, experts and eyewitness reports, which are then translated by humans for systems available in multiple languages (Lyon *et al.* 2011)

Not constrained in its reporting by the need for official clearance, as with the World Health Organization (WHO), ProMED-mail allows a human-based multilingual crowdsourcing project to outperform more automated processing systems. For example, specialists in Ukraine, Russia and Uzbekistan post an average of 2-3 daily reports with expert comments 7 days a week. These selected reports are then translated form the Russian on ProMED-RUS post to the main ProMED-mail English page in an average of 0-4 days, following a translation and editing lag time (Rakhmanova *et al.* 2014). This far-reaching cross-lingual communication project outperforms WHO reporting, despite its human component, ensuring massive multilingual cover through its crowded resources. In 23 outbreaks on WHO's Outbreak Verification List (OVL), ProMED-mail reports preceded the OVL listing, versus 5 outbreaks in which the OVL preceded ProMED-mail (Woodall 2001).

In both the Stop Stock-Outs and ProMED-mail crowdsourced translation campaigns for healthcare initiatives and reporting, human-centric multilingual systems designed around major and minor language processing systems were means of engaging users and non-professional translators to bridge socio-linguistic gaps in framing and contextualising reports from otherwise geographically and/or linguistically excluded populations. Such non-profit health campaigns transformed new technology and the buzzword of the new millennium into powerful transcendental tools to use crowdsourced translators as a vehicle for social impact and inclusive change.

7. Monitoring Outbreaks and Epidemics with a Click

Diseases and epidemics know no borders, making collaborative translation projects with integrated data components that deliver real-time intelligence on topics such as emerging infectious diseases and outbreaks invaluable. Crowdsourced translation applications for global public health can be used to organise multilingual open-source data and give users complete and unfiltered control of searching the aggregated data, rather than framing and constructing fixed analysis tools. Custom tools supply alternatives to perfect, linear translations of cross-national reports on epidemiological surveillance, instead focusing on rapid-fire, cross-border, multilingual communication, with mapping capabilities and logical information filters. It is a sensitive, but not necessarily specific system, using local scientific and linguistic intelligence to pinpoint outbreak vulnerabilities.

8. Conclusions

Crowdsourced translation initiatives in the healthcare field in the developing world, as with Stop Stock-Outs and ProMED-mail, have all contributed to pioneering and co-innovating crowdsourced translation as a tool for inclusion, access and assistance, quickly assembling the data, know-how and resources necessary to carry out a task or solve a problem by facilitating multi-channel communication among people and organisations, collaborating across disciplinary, geographic and linguistic boundaries. However, despite reductionist descriptions of crowdsourcing in the context of language processing, this type of translation is not a singular task, but a translation eco-system, combining human and automated work-flows and interfaces to harness professional, crowdsourced and machine translation that remain inclusive and flexible enough to motivate contributors and engage broad general participation in such relief, development and aid initiatives.

Crowdsourced translation will continue to have varied and cross-platform applications, some of which we can likely not even predict at present; however, what remains clear is that a connected world means translation is now a much broader and scalable possibility, regardless of the number of speakers, the far-flung location or the lack of or blocked social-media footprint. We are already speaking to one another cross-linguistically and this is changing very quickly, as are the means used to do so. The place for MT, HT and crowdsourced translation in humanitarian work and social initiatives will continue to be re-shuffled in coming years as technology gains precision and humans are better able to integrate automated technology into existing work-flows. Regardless, the outlook is optimistic for the harnessing of mass communication to serve as a mechanism of inclusion through translation and cross-lingual communication in a growing digital world that has made our geographic footprint that much smaller.

References

Cowen, P., T. Garland (2006) M. High-Jones, A. Shimshony, S. Handysides, D. Kaye, L. Madoff, M. Pollack and J. Woodall 2006. “Evaluation of ProMED-mail as an Electronic Early Warning System for Emerging Animal Diseases: 1996 to 2004”, Journal of the American Veterinary Medical Association 229(7), 1090-1099.

Crowley, J. and Chan J. (2011) “Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies”, UN Foundation-Vodafone Foundation-UNOCHA.

Eidelman, V., Hollingshead, K. and Resnik P. (2011) “Noisy SMS machine translation in low-density languages”, 344-350.

Joseph, S. (2014) “The marketers’ game plan to scoring big at the World Cup”, Marketing Week.

Junglelightspeed.com (2014) “Sudden onset translation”. Available online at: [http://www.junglelightspeed.com/sudden-onset-translation]. (accessed on 6 June 2014).

Lorentz, F. (2004), Digiworld 2004. 1st ed. Montpellier.

Lyon A., Nunn M., Grossel G., Burgman M. (2011) “Comparison of web–based biosecurity intelligence systems: BioCaster, EpiSPIDER, and HealthMap. Transboundary and Emerging Diseases”. Doi: 10.1111/j.1865-1682.2011.01 258.x

Lyon A., Nunn M., Grossel G., Burgman M. (2012) “Using internet intelligence to manage biosecurity risks: a case study for aquatic animal health. Diversity and Distributions”.

Madoff LC (2004) “ProMED-mail: An Early Warning System for Emerging Diseases”. Clin Infect Dis 39:227-32.

Medicines Transparency Alliance (2010) “Case Study: Stop Stock-outs Campaign A CSO initiative to increase access to medicines”.

Rakhmanova N., N. Pshenichnaya, B. Aslanov, and V. Melnik (2014) “Global surveillance of emerging diseases: Contribution of the Russian regional ProMED-mail network”, International Society for Infectious Diseases.

StopStockouts.org (2010) “Access to Medicines – African Civil Society Role and Response”, Available online at: [http://stopstockouts.org/2010/01/18/access-to-medicines-%E2%80%93-african-civil-society-role-and-response] (accessed 7 June 2014).

Woodall, J. (2001) “Global surveillance of emerging diseases: the ProMED-mail perspective”, Cadernos de Salude Publica [Reports in Public Health], vol. 17, 147-154.