CASE STUDY
Category: ICTs, Climate Change and Disaster Management

Climate Change, Innovation & ICTs Project

Centre for Development Informatics (CDI), University of Manchester, UK
With the support of the International Development Research Centre (IDRC)

Pakreport: Crowdsourcing for Multipurpose and Multicategory Climate-related Disaster Reporting

Authors: Faisal Chohan, Vaughn Hester, and Robert Munro

Initiative Overview

Fierce monsoon rains in 2010 caused the worst flooding in Pakistan in 80 years (Aon Benfield 2010). Approximately one-fifth of Pakistan's total area was underwater at the height of floods. The floods displaced and affected 20 million people, mostly via destruction of property, livelihood and infrastructure. Close to 2,000 people died as a result. Although the specific link to climate change is unclear, it is generally agreed that climate change is already increasing the risk of flooding and that this risk will grow in future (IPCC 2007). In particular, "Pakistan stands among the group of developing countries which are extremely vulnerable to the adverse impacts of climate change." (Planning Commission 2010:13). In recent years, climate change in Pakistan is evidenced by an increase in temperature in summer and timing changes of the monsoon season (Dell'Amore 2010). These changes are affecting agricultural patterns, but they also threaten a greater incidence and intensity of natural disasters, including floods.

During such disasters, information is at a premium: there is an urgent need to know which areas are affected; how they are affected; what the priority problems are; and so forth. This typically coincides with the disruption of traditional lines of communication. The advent of mobile phones has provided a new digital development infrastructure, which may be of significant value to disaster response. Pakistan has seen particularly strong growth in mobiles, with roughly 110m subscriptions in mid-2011; well in excess of the adult population (PTA 2011).

This case study focuses on Pakreport, an ICT initiative between crisis mapping organisations, engineers, relief agencies and crowdsourcing companies that began in July 2010 as a response to the floods. We will review emergency communication using ICT tools and volunteers using crowdsourcing platforms to perform verification, categorisation, translation, and mapping of the information in real time. This initiative demonstrates ICTs' impact on climate change in the domains of both adaptation (disaster management) and monitoring (Heeks 2009). This case study will extend previous research on crowdsourced workflows for crisis relief in Kenya and Haiti.

Application Description

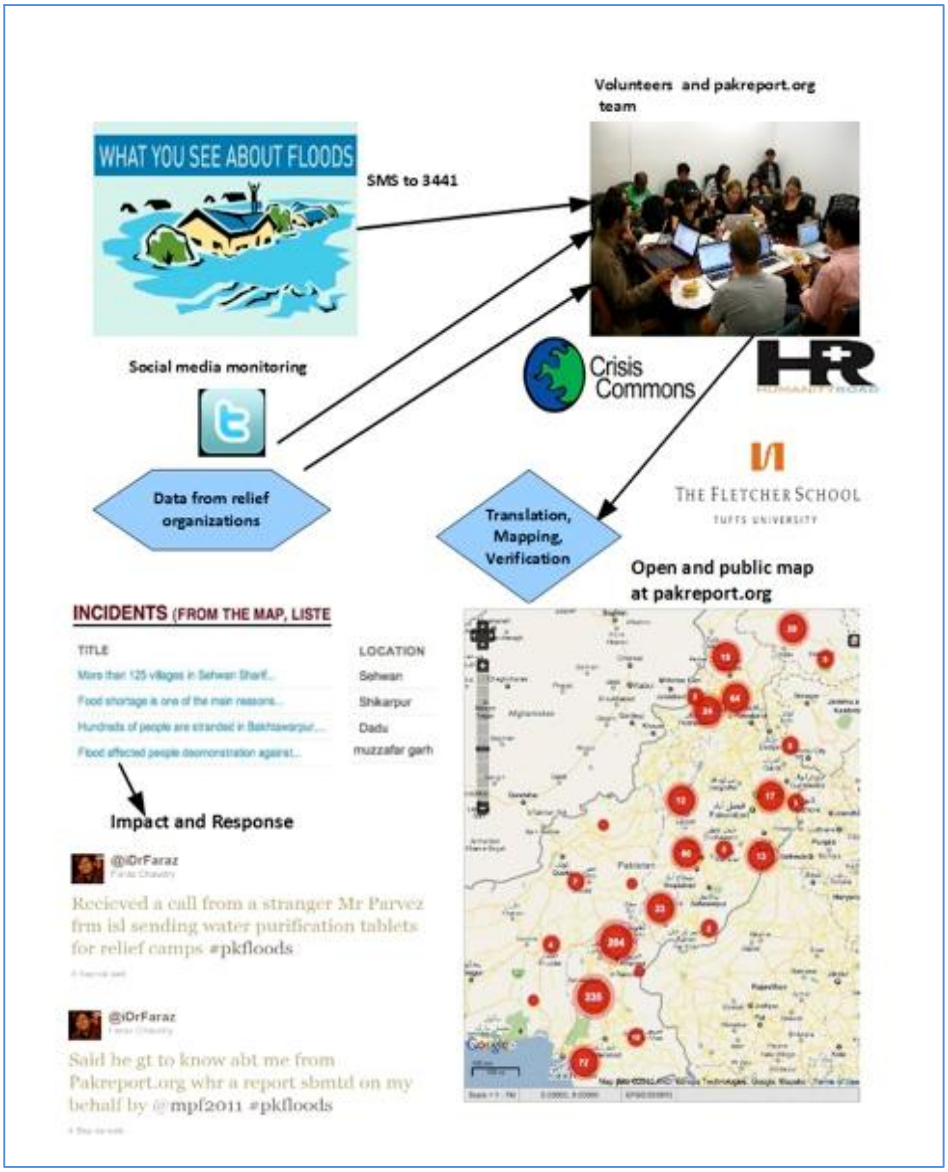
Pakreport is a customisation of Ushahidi[1] software in Pakistan which employs two forms of crowdsourcing. First, the use of a distributed group of people to provide data reports from the ground. Second, the use of a (very different) distributed group to translate, categorise and geolocate the incoming messages; this being undertaken via a CrowdFlower microtask[2]. Once the information was processed, it was input and displayed on the Pakreport.org platform; most visibly via an online map: see Figure 1.

[1] Ushahidi is an open source application which allows users to collect crisis information from large numbers of people.
[2] CrowdFlower is a company that crowdsources projects by breaking them into microtasks: a microtask is another name for a short, online form that can be completed by a member of the crowd performing a small analytical task.

The Pakreport platform was deployed on 8th August 2010, and could accept data input in many different forms: via radio channels, via social media sites such as twitter, and directly online via the web site itself. However, the primary source of initial information was a set of village-level assessments direct from the relief agencies which detailed the situation and damage on the ground. These reports were categorised and mapped via Pakreport staff.

The team also wanted a simple way for ordinary people caught up in the disaster to report their situation. SMS was seen as the most effective means for this, and the team at Pakreport set up a short code – 3441 – that was available on four of the five mobile companies in Pakistan. It was a shared short code, so the "FL" tag needed to be appended before the message.

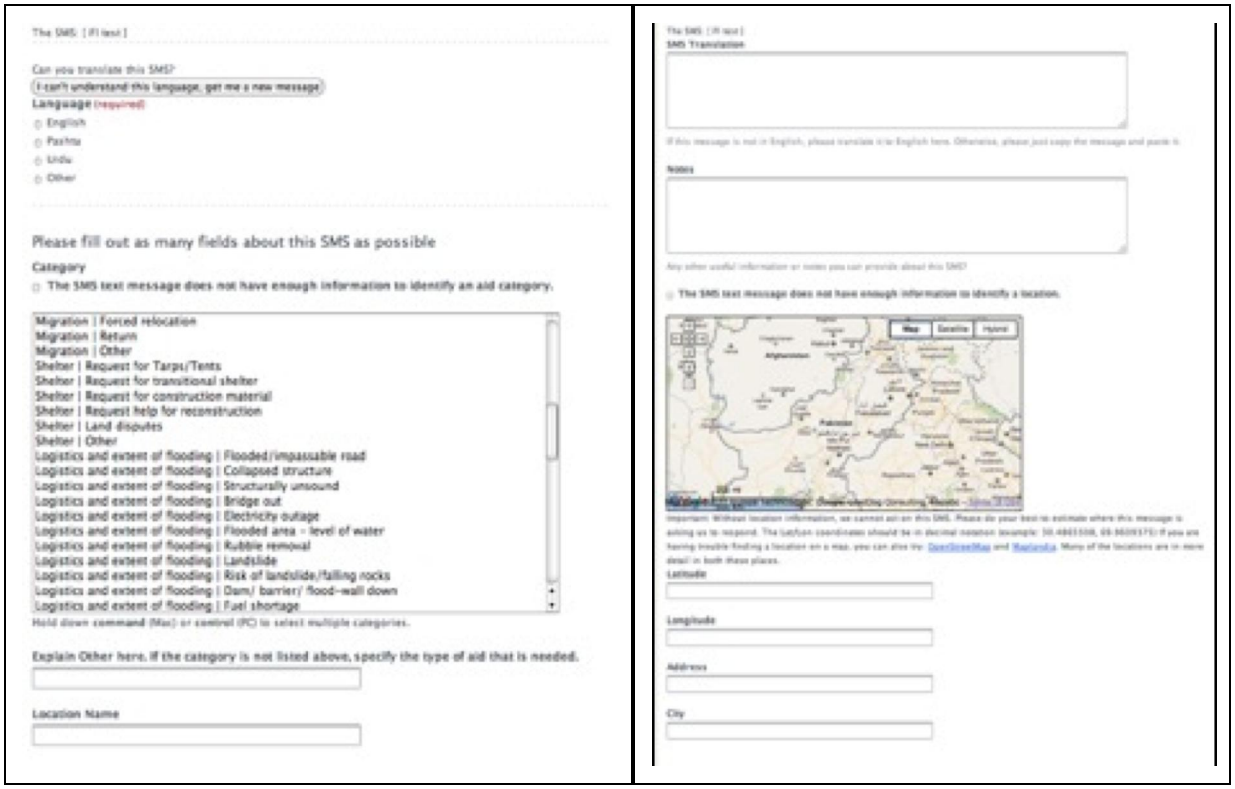
Figure 1: The Pakreport.org Operational Flow and Stakeholders



Details of the short code with the message "what you see about floods" was spread via the mass media – in particular via a partnership with the BBC World Service – and via relief agency workers. This led to a substantial increase in the volume of data being received, with the SMS channel taking over as the primary information source. This created a need to categorise the incoming messages depending on what was being reported, and then (if locational details were available with the SMS) geolocate the message so that it could be mapped.

Messages might also need to be translated from Pashto or Urdu to English. It was decided to crowdsource these microtasks via partner CrowdFlower. Volunteers from around the world participated in completion of the microtasks (see Figure 2). Because accuracy was critical to these efforts, CrowdFlower added a level of redundancy for improved quality control (as compared to previous disaster relief workflows), meaning that multiple volunteers evaluated each SMS message. This resulted in the collection of multiple points on a map, which the CrowdFlower platform then used to calculate the centroid of the points. This increased the accuracy of the final results being placed onto the online map.

Figure 2: CrowdFlower Disaster Message Analysis Microtask Forms



Because maps of and details about certain locations (especially rural villages and small towns in remote areas of Punjab and Sindh provinces) were not well developed or readily available, the Pakreport team created detailed online documentation (http://groups.google.com/group/PakReport-volunteers/web/mapping-links) to supplement these gaps. The team also provided training through Skype to manage 40 volunteers who assisted with this part of the process.

The map – with details, labels and annotations from the messages and other incoming data flows – was made available to relief agencies in Pakistan, providing real-time reporting for them. It was integrated with various ICT implementations; nationally in the UN and NGO sector (e.g. http://www.pakresponse.info) and the National Disaster Management Authority (NDMA); and provincially in the Provincial Disaster Management Authorities (PDMAs). The map enhanced the response efforts by improving prioritisation and coordination of the disaster response.

Formal Drivers and Objectives/Purpose for ICT Usage

Climate change is increasing the dangers of natural disasters for developing countries, such as flooding. Although the specific link to climate change is uncertain, the impetus for Pakreport was the massive flooding of 2010; in particular, the sense of isolation felt by those affected, and the limits on quality information for those responding. The objectives of this initiative were therefore a) to create an SMS-based line of reporting for the flood-affected people to communicate with the outer world about their situation on the ground; and b) to connect this information with the appropriate disaster response stakeholders to enable improved decision-making and relief efforts.

Stakeholders

As illustrated in Figure 1, this project brought together a very wide range of stakeholders. The main stakeholders were the local communities and flood-affected people who used mobile phones to communicate their emergency situation. Relief agencies and the general public who sought information about the crisis situation and who could respond to help requests were the end-stakeholders who received the information. Between these were many others:

- Media entities (e.g. BBC World Service radio, Internews) who helped spread the word about Pakreport
- The core Pakreport team
- Various volunteer groups (Crisis Commons, Crisis Mappers, Humanity Road) who participated via CrowdFlower in data analysis or who gave direct assistance to Pakreport
- Individual volunteers (from Fletcher University) who had worked on similar exercises in the past
- Technology partners (Pakistani cell phone providers, and CrowdFlower)

Impact: Cost and Benefits

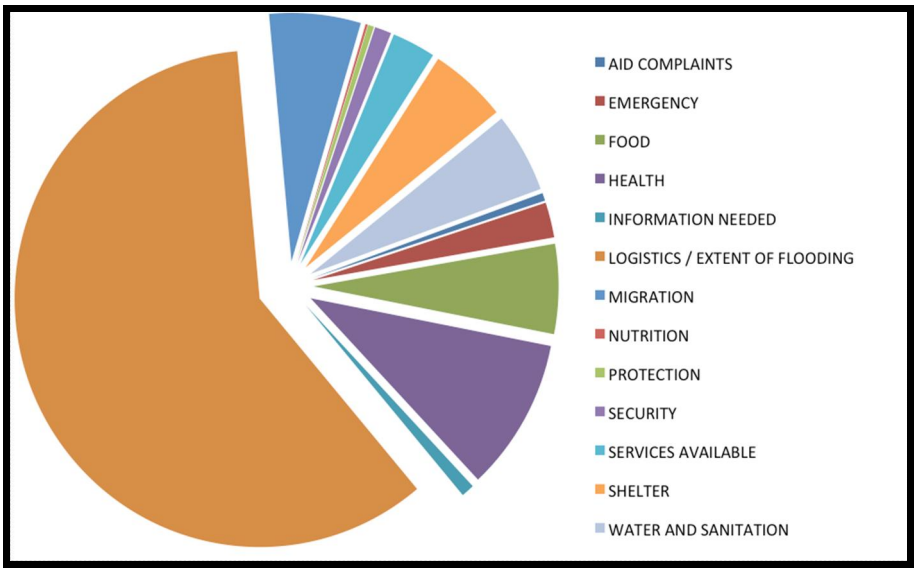
Costs of Pakreport included: set up of the short code; monthly rental charges for the short code, hosting of the website on Amazon EC2 and managing logistics of the team working in Pakistan. Usually, the amount of technical and mapping resources utilised at Pakreport.org can cost in the range of US$10,000-15,000, but most of the team members and partners volunteered their time. Total cost of the project was US$7,000. The funding came from a fundraising campaign at globalgiving.org. Work space and office supplies were provided by Cogilent Solutions. The microtasking platform and technical services for integration were provided by CrowdFlower free of cost. Ushahidi instance customisation and code development for integration with CrowdFlower were also provided gratis. Three independent engineers also donated services: Chris Blow, George Chamales and Robert Munro.

Pakreport created a number of benefits:

- Collection of 1500 real time reports from the people on ground through SMS.
- Translation, categorisation and mapping of reports in near real time. Crowd volunteers completed over 2500 labels or categorisations of reports.
- Detailed knowledge and mapping resources organised by the Pakreport team (to view these map resources, please visit www.pakreport.org or refer to Figure 1.)
- A base for future crowdsourcing and mobile implementations in Pakistan. It would be easy to repurpose or duplicate this type of project. The growth and socialization of short code use and short code disaster reporting is an important component of the future uptake of similar implementations.
- Awareness of and dialogue about the effects of climate change. The collaborative reporting about the flooding and its aftermath contributed to environmental awareness as well as a clear channel for reporting and monitoring the environmental changes throughout Pakistan. Pakreport also represents an important and innovative technological foundation for ongoing, national-level reporting, monitoring and/or early warning efforts.

The breakdown of categorised messages in Figure 3 shows that the majority of categories were related to the floods themselves and their immediate effects like migration, shelter, water and sanitation. Among the largest categories, however, there was also security and protection. The affected populations were clearly concerned about their vulnerability to deliberate physical threats that were not directly related to the floods, especially among frequent rumours of aid groups becoming the targets of insurgent attacks.

Figure 3: Share of Messages by Category (Source: Pakreport)



Evaluation: Failure or Success

Evaluating Pakreport

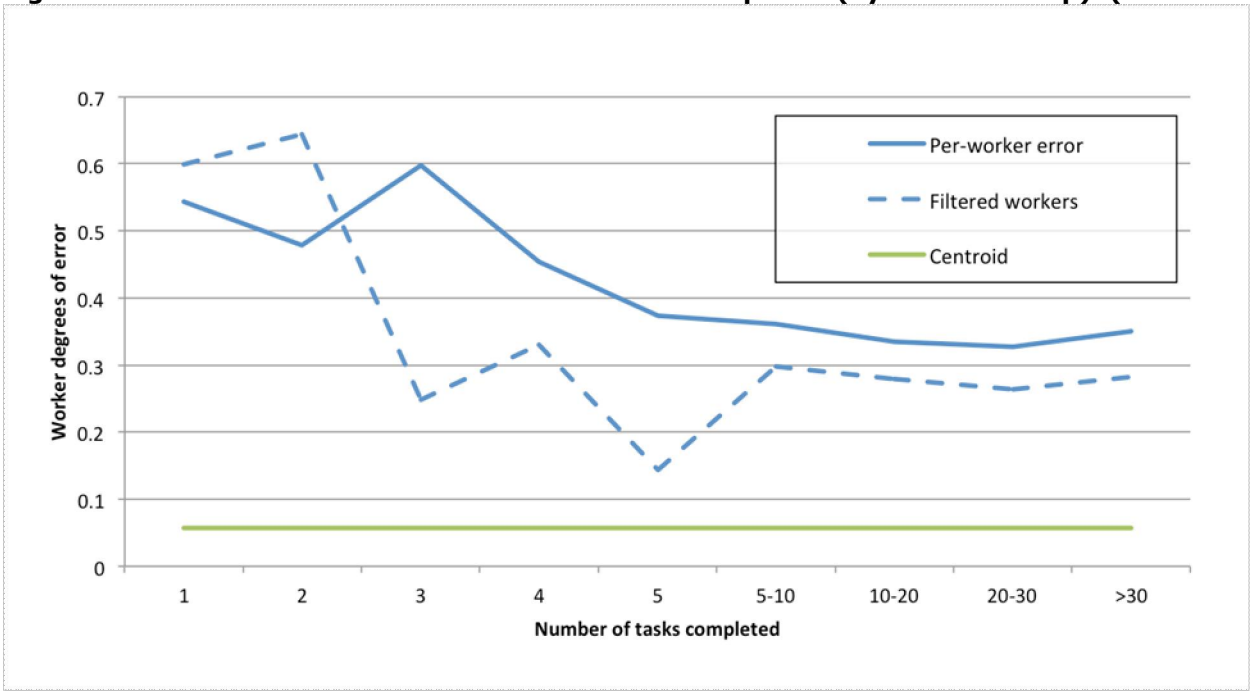
The project was successful in enabling rapid data moderation as a foundation for information exchange among various agencies and actors. Flood-affected individuals did not know about any platform to connect with the relief providers before or during the floods. The first responders were the Pakistan Army; but they did not have any mechanism to share ground-level information with other relief agencies. The information within the UN was managed through – but largely restricted within – the UN Cluster system (with a call centre to receive calls from ground and route this information to the UN agencies).

A primary success of the Pakreport platform is thus the creation of mapping knowledge and information in Pakistan that did not exist previously. A second success was the widespread use of mobile devices and a crowdsourcing platform to connect with people on the ground in a disaster situation. Mobile technologies, crowdsourcing and open data represent three emerging trends that have yet to be adopted by many nonprofits and relief agencies, especially in Pakistan. The collection and dissemination of this information created more focused, targeted and informed relief and response efforts. Relief actors were better able to direct time, resources and personnel as a result of open access to real time reports and requests for assistance from throughout the country. These efforts were also highly participatory in that they incorporated the survivors of the floods as well as flood-affected communities into the response efforts.

Evaluating Crowdsourcing Strategies

Figure 4 shows the average degrees of error between the volunteers working on the CrowdFlower platform and the final coordinates published in the Pakreport instance. The two most common types of error were omitting a category or confusing the services with requests. A typical example of the former is someone reporting "we need food and water", but only the "food needed" category is selected ("Water needed" is a separate category as water is a more time-critical need and also because some response agencies will focus only on ensuring clean drinking water). A typical example of the latter is someone reporting "There is a makeshift shelter treating the wounded" which is categorized as "medical attention needed" rather than "medical attention offered".

Figure 4: Worker Error Rates vs. Number of Tasks Completed (by Worker Group) (Source: Pakreport)



The first line shown, *per-worker error*, is simply the average degrees difference per worker after they have completed 1 task, 2 tasks, 3 tasks, etc. It shows improvement from the 3rd to the 10th task, indicating the volunteers became better at the task once they became more familiar with it, but not immediately so. The second line, *filtered workers*, excludes workers who averaged more than 1 degree of error. Predictably, after a few tasks filtration is a consistently more accurate strategy. The final graph, *centroid*, shows the method actually used in the deployment: the centroid of the locations identified by different workers (it is the average over all tasks, as the number of tasks is not as meaningful). It clearly shows that this was more accurate than taking any single worker's locations, even if that worker was substantially experienced.

To our best knowledge, no humanitarian organization has previously attempted to increase the accuracy of encoding by giving real-time work to multiple workers (crowdsourced volunteers or otherwise) so we hope that our analyses of different filtering/aggregation techniques can positively influence the decisions of humanitarian organizations that are considering possible information processing strategies. We did not filter workers during PakReport. Rather, we simply took the super-set of all categories they selected and the weighted average across locations. We explored the potential for filtering workers in post-hoc analysis of the data, simply because filtering high-error workers is standard practice in commercial microtasking platforms.

Enablers/Critical Success Factors

Crowdsourcing was critical to the success of this disaster response system. It was integral to the data input model, which would otherwise have relied on much more limited inputs from individual relief agency workers. It was not integral to the data analysis model, but its use greatly reduced the costs of analysis and mapping, and increased the timeliness and accuracy of those processes.

Although there was a need for innovation to fit the particular requirements of the Pakistan floods, in large part Pakreport **used existing technologies**. The bulk of its digital infrastructure was provided by the country's mobile phone system, which remained sufficiently operational; so there was no requirement for new infrastructural investments. Data input relied on a technology – mobile – that had already diffused to almost all parts of the country, and which was already familiar to, and in use by, the majority of the population.

Pakreport also made use of Ushahidi – a well-tried and well-trusted technology platform – rather than seeking to develop technology from scratch (something which, given the disaster timescales, would in any case have been impossible).

Related to this, the project **utilised established organisational expertise** in a number of ways. Because Ushahidi was the foundation, then members from Ushahidi Haiti and Chile implementations of crisis reporting and mapping platforms along with Silicon Valley engineers joined the Pakreport team in the first week after deployment. The team's expertise was substantial, and it included lessons learned first hand from the Haiti and Chile implementations. Similarly, partnering with CrowdFlower allowed the data analysis microtasks to be rapidly and scalably rolled out, rather than necessitated a new organisational infrastructure to be created. And linking up with radio broadcasters made use of their expertise and their existing technologies to publicise use of the SMS message system.

Finally, **altruism** can be seen to have played a role. Of course, some of those providing the SMS inputs were motivated out of their personal need for relief from this climate-linked disaster. But others were simply reporting what they saw for the benefit of others. Much of the technical development and all of the data analysis work was undertaken by volunteers; providing their time in the service of those affected. And individuals also took action on the basis of the reports they saw mapped, in order to organise their own direct relief activities.

Constraints/Challenges

The project had an **informational and technological not disaster response** focus at least in the initial stages. That is, the main work that had to be done was in setting up the software, web platform, data input and analysis and mapping processes in order to produce the map-based information. But this was separate from the disaster response effort so that coordination with the disaster relief agencies was at first quite limited. In large part this was due to the nature of application deployment – during the disaster rather than prior to the disaster when there could have been time to make relief agencies aware of the application, and to help ensure the information produced was being used to guide field decisions and actions.

A major challenge of any crowd reporting project is **verifying the accuracy and authenticity of the data** coming from the ground. In general the hope is that the volume of data will be such that good data drives out bad, but this may not always be the case, especially as crowdsourcing of climate-related data increases. There is no authoritative answer yet (and the dangers are probably greater during political crises rather than natural disasters) but some suggestions include weighting based on past data inputs, use of additional data such as locational or photographic, direct questioning of the source, and the possibilities for using language analysis software (Meier 2011).

Using mobiles and SMS enables any climate change-related project to reach a very large user population. However, there are still **digital divide** problems. For example by restricting inputs to three languages and by requiring a text-based message to be sent, Pakreport recognises that it did exclude some members of the population, such as illiterate mobile owners. Use of the web as the main reporting tool similarly meant that only a certain subset of the population was able to utilise the results.

Although recognising the great value of the complementarities brought by having so many stakeholders, this also brought the challenges of **negotiating agreements and reaching consensus**. With telecommunication providers, government departments, international agencies, the national media and others all involved, this added greatly to the transaction costs of the Pakreport initiative; something particularly challenging given the ever-pressing timescale of disaster relief. This required a great deal of time and effort; not just when formal agreements were needed but also in seeking to create less-formal consensus; for example around the rather-radical notion of collecting and then disseminating potentially unverified citizen reports.

Recommendations/Lessons Learned

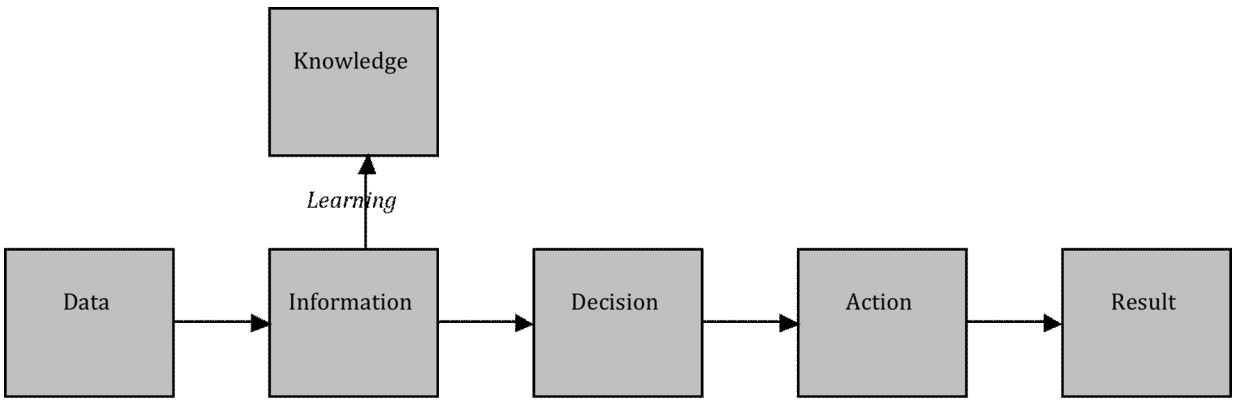
The following lessons were learned from this case study:

1) **Crowdsourcing can be highly powerful and effective for ICT-enabled climate change applications**. As seen, crowdsourcing can be used in at least two different ways – for gathering disaster / climate change data from a very broad set of users and locations; and for then analysing that data so that it can be effectively displayed and utilised. Although this particular application related to a climate-linked disaster, it is entirely feasible to use the same model for monitoring climate change e.g. by asking populations to provide information on local drought, rainfall, temperature, waterflow, etc. This would of itself also help to raise awareness about climate change; something that could be enhanced with a feedback loop by which those providing data would themselves also be sent short reports and climate alerts via SMS, web, etc. Locations at high risk of natural disaster and/or climate change effects should consider proactive establishment of this type of simple reporting workflow.

2) **Mobile-plus-Internet-plus-servers equals a system with reach and power**. On their own, mobile and Internet and server technologies have great value. However, the technological key to Pakreport has been its combination of the three. Mobile provided the reach down to the "bottom of the pyramid" populations who are on the front line of disasters and other climate change-related vulnerabilities. The Internet provided the reach and power to help coordinate volunteers across the world, and disseminate results to relief agencies. As the foundation, servers provided the power to collect, analyse, store and display the processed information. Other ICT-based disaster and climate change applications can therefore identify how to combine the reach and power of these technologies into an overall system.

3) **The full "information chain" must be in place**: the provision of information on climate-linked disasters is critical to effective disaster response and broader management. But the latter are only possible if there is a full "information chain" (see Figure 5); that is if there is a mechanism by which that information is turned into decisions about what to do and where and how to do it; and those decisions are then turned into actions on the ground. ICTs and climate change projects must therefore be designed around the entire chain, typically starting that chain backwards from the results that are sought and the actions necessary to achieve those results.

Figure 5: The Information Chain (Source: adapted from Heeks & Kanashiro 2009)



Data Sources & Further Information

Unless otherwise noted, the data and figures herein came from the CrowdFlower and Pakreport systems / databases and from the experiences of the authors: Faisal Chohan is a co-founder of Pakreport and a TED Fellow in Pakistan; Vaughn Hester is a Program Manager at CrowdFlower and helped set up the CrowdFlower task used to process Pakreport data; Rob Munro, a computational linguist, was a member of the Pakreport.org team.

For Further Information:

Faisal Chohan, Pakreport, faisal@pakreport.org
Vaughn Hester, CrowdFlower, vaughn@crowdflower.com
Robert Munro, Stanford University, rmunro@stanford.edu

References

Aon Benfield (2010) *Pakistan Flood Event Recap Report*, Aon Benfield, Chicago, IL. http://www.aon.com/attachments/reinsurance/201008_pakistan_flood.pdf

Dell'Amore, C. (2010) Russia fires, Pakistan floods linked?, *National Geographic*, 11 Aug. http://news.nationalgeographic.com/news/2010/08/100812-russian-wildfires-pakistan-floods-global-warming-science-environment

Heeks, R.B. (2009) Overview model of ICTs, climate change and development, *Notes on ICTs, Climate Change and Development*, 14 Dec http://niccd.wordpress.com/2009/12/14/overview-model-of-icts-climate-change-and-development/

Heeks, R.B. & Kanashiro, L.L. (2009) Remoteness, Exclusion and Telecentres in Mountain Regions: Analysing ICT-Based "Information Chains" in Pazos, Peru, IDPM Development Informatics Working Paper no.38, University of Manchester, UK http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/

IPCC (2007) *Working Group II Contribution to the Intergovernmental Panel on Climate Change Fourth Assessment Report*, Intergovernmental Panel on Climate Change, Geneva. http://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-spm.pdf

Meier, P. (2011) How to verify social media content, *iRevolution*, 21 June. http://irevolution.net/2011/06/21/information-forensics/

Planning Commission (2010) *Task Force on Climate Change*, Planning Commission, Islamabad http://www.planningcommission.gov.pk/usefull%20links/Taskforces/TFCC%20Final%20Report.pdf

PTA (2011) *Telecom Indicators*, Pakistan Telecommunications Authority, Islamabad. http://www.pta.gov.pk/index.php?Itemid=599

EDITORS:
Richard Heeks
Angelica Valeria Ospina

The **Climate Change, Innovation and ICTs** Project is an initiative led by the Centre for Development Informatics (CDI) of the University of Manchester, UK, with funding support from Canada's International Development Research Centre (IDRC). Further information about the project and related resources can be found at: http://www.niccd.org

SOME RIGHTS RESERVED BY NC **2011**