DATA-POP ALLIANCE SYNTHESIS REPORT

BIG DATA FOR CLIMATE CHANGE AND DISASTER RESILIENCE:

REALISING THE BENEFITS FOR DEVELOPING COUNTRIES

September 2015

DATA-POP ALLIANCE
SYNTHESIS REPORT

BIG DATA FOR CLIMATE CHANGE AND DISASTER RESILIENCE:

REALISING THE BENEFITS FOR DEVELOPING COUNTRIES

September 2015

Table of Contents

Foreword and Acknowledgements i

About this document i
About the authors i
Thanks ii
Disclaimer ii
Access and Citation ii

Executive summary iii

1. What is the overall context and message of this report? iii
2. How can Big Data help? Learning from existing applications iv
3. What are some of the key barriers, gaps and risks? v
4. What could a desirable and feasible roadmap entail and achieve? vi

Introduction 1

1. Setting the stage and stakes 4

1.1 Hazards, disasters, vulnerability and resilience in developing contexts 4
1.2 What is Big Data? 5
1.3 How can Big Data increase resilience? 7
1.4 Key actors and activities 8
1.5 The DfID case studies and pilot projects 9

2. Opportunities and potential of Big Data for resilience 11

2.1 Monitoring hazards 11
Improving existing systems 11
Using new sources of data to monitor risks 12
2.2 Assessing exposure and vulnerability to hazards 13
New satellite imagery 13
Crowdsourced mapping 14
Call detail records 15
2.3 Disaster response: Early warning, situational awareness, and immediate impacts 16
Situational awareness 17
Immediate impacts 17
2.4 Innovative approaches to assessing the vulnerability and resilience of natural systems 20
2.5 Beyond single events: Big Data and general disaster resilience 21
Feedback throughout the disaster cycle 21
Societal learning about risks 22
Collective action and accountability 23

3. Challenges in mobilizing Big Data for resilience 24

3.1 Data access and completeness 24
3.2 Analytical challenges to reliability, representativeness and replicabilty 25
3.3 Human and technological capacity gaps 27
3.3 Bottlenecks in coordination, communication, and self-organization 28

3.4 Ethical and political risks 30
4. Toward a roadmap 31
4.1 The state of the field 31
4.2 Specific recommendations 33
References 38
Endnotes 43

Boxes, Figures, and Tables

Box 1. Key Terms and Concepts at a Glance 2
Box 2. Overview of Case Studies 10
Box 3. Using Data Mining to Create Landslide Exposure Maps in Data-Poor Contexts 13
Box 4. Auditing Early Warning Programmes With CDR-Based Maps of Population Movement 16
Box 5. Leveraging Mobile Phone Data to Infer Post-Disaster Movements 18
Box 6. The Possibilities and Limitations of Digital Disease Detection 19
Box 7. Advancements in Open Data 22
Box 8. Lessons from the Ebola Outbreak 24
Box 9. Correcting Sampling Bias in CDR Data in Senegal 26
Box 10. Example of the 'Crowdseeding' Approach 27

Figure 1. Disasters and resilience 5
Figure 2. Three-dimensional model of the Groundwater (left) and Google Earth hydrography map (right) of the Al Assi (Orontes) river basin in Lebanon 11
Figure 3. Schematic display of a typical Twitter count pattern leading up to a flood event 12
Figure 4. The city of Brasilia from the current Global Human Settlement Layer, using a combination of images from different satellites with resolution ranging from 0.5 to 10 m. 14
Figure 5. Comparison of OpenStreetMap coverage of Kathmandu, before and after the 2015 Nepal earthquake and the efforts of the Humanitarian OpenStreetMap Team 15
Figure 6. Post-earthquake population movement in Nepal 19
Figure 7. Diagram illustrating a tipping point, where a system shifts rapidly from one equilibrium state to another 21
Figure 8. Definition of 'open data' for the OpenDRI Field Guide 23
Figure 9: Global telecommunication capacity by country income group 28
Figure 10. The Humanitarian Data Exchange's page for the Nepal Earthquake 29

Table 1: Sustainable Development Goals and selected targets related to resilience 3
Table 2. Taxonomy and Examples of Big Data Sources 6

Foreword and Acknowledgements

About this document

This document is a synthesis report for the "Big Data for Resilience: Realising the Benefits for Developing Countries" project funded by the United Kingdom Department for International Development (DfID), whose financial support is gratefully acknowledged.

It draws on an extensive review and analysis of the available academic and policy literature as well as the main findings of eleven contributions commissioned by DfID with the Natural Environment Research Council (NERC) and the Economics and Social Science Research Council (ESRC) as part of the Science for Humanitarian Emergencies and Resilience Programme.

The report is intended to provide evidence that will feed into the Post-2015 Development Agenda in September 2015, the Paris 2015 UN Climate Change Conference, and the 2016 World Humanitarian Summit to be held in Istanbul at the end of May 2016, as well as to support the implementation of the Sendai Framework for Disaster Risk Reduction; several dissemination and discussion activities are scheduled in the lead up to these events.

This document was developed by a team of researchers under the umbrella of Data-Pop Alliance of the Harvard Humanitarian Initiative (HHI), MIT Media Lab, and Overseas Development Institute (ODI) during May-July 2015.

About the authors

- Jesse Anttila-Hughes, Assistant Professor of Applied Econometrics, Department of Economics, University of San Francisco, Research Affiliate, Data-Pop Alliance, as synthesis report co-lead author and contributor
- Marion Dumas, PhD Candidate in Sustainable Development, Columbia University, Research Affiliate, Data-Pop Alliance, as synthesis report contributor
- Lindsey Jones, Research Fellow at ODI, as synthesis report contributor
- Gabriel Pestre, Research Scientist, Data-Pop Alliance, as synthesis report contributor
- Yue Qiu, Columbia University, as synthesis report contributor, especially on section 2
- Marc Levy, Deputy Director, CIESIN, Adjunct Professor, School of International and Public Affairs, Columbia University, as synthesis report co-lead author
- Emmanuel Letouzé, Director and Co-Founder of Data-Pop Alliance; Visiting Scholar, MIT Media Lab, Fellow at HHI, and Research Affiliate at ODI, as synthesis report contributor and principal investigator
- Simone Sala, Associate Director of the Dr. Steve Chan Center for Sensemaking, AIRS, at Hawaii Pacific University and Swansea University, Research Affiliate, Data-Pop Alliance, as synthesis report contributor
- Bessie Schwarz, Communications Coordinator, Yale Project on Climate Change Communication and Research Affiliate, Data-Pop Alliance, as communications strategy lead and synthesis report contributor
- Natalie Shoup, Program Manager, Data-Pop Alliance, as workshop lead and synthesis report contributor,
- Elizabeth Tellman, PhD student, Arizona State University, Research Affiliate, Data-Pop Alliance

- Patrick Vinck, Assistant Professor at Harvard University, Director of HHI's program for Vulnerable Populations, Co-Director and Co-Founder of Data-Pop Alliance, as synthesis report contributor and reviewer

Thanks

The authors wish to thank Max Baiden (Development Alternatives Inc.) for his support in the preparation of this report and process and Nicola Ranger (DfID) for her trust and detailed guidance during the course of the development of this report.

We are grateful to John Crowley for his input and guidance during the writing workshops for the report as well as to the participants of the dedicated workshop organised in London on June 5, 2015, including representatives from all teams that prepared the commissioned case studies used in this report and shared their perspectives: Craig Bardsley, Emily Boyd, Tao Cheng, Jonathan Cinnamon, Melanie Duncan, Serge Guillas, Henry Kiragga, Mark Iliffe, Kate Jones, Leanne Jones, Brendan Jongman, Sriganesh Lokanathan, Sue Loughlin, Markus Luczak-Roesch, Katy Mee, Seb Mhatre, Lucy Moore, Chloe Onoufriou, Mamun Rashid, Silke Roth, Rohan Samarajiva, Gemma Sou, Elisabeth Stephens, Andrew Tatum, Mark Thorley, Erik Wetter, Jenny Wilson, and David Wrathall.

We also wish to thank Jan Gallagher who edited this document, and Kofo Sanusi and Suzi Nandera (ODI) for their support managing the process.

Disclaimer

The views presented in this document are those of the authors. They do not necessarily reflect the view of the institutions with which they are affiliated or of DfID, NERC, or ESRC. Any errors or omissions are those of the authors.

Access and Citation

The full report will be available at http://www.datapopalliance.org/projects

Suggested citation: Data-Pop Alliance. *"Big Data for Resilience: Realising the Benefits for Developing Countries"*. Synthesis report. September 2015.

Executive summary

1. What is the overall context and message of this report?

Scores of recent events have revealed the vulnerability of poor communities to natural hazards such as earthquakes, storms, and epidemics while less well-publicized longer-term hazards, such as desertification or threats to staple food sources, can have even more devastating effects. Several factors are expected to raise disaster risk in developing countries over the next few decades. Higher urban densities and larger coastal settlements for example will increase the size of vulnerable populations and assets exposed to hazards in developing countries, while climate change will in all likelihood increase the frequency and intensity of hydrometeorological hazards in varying, complex ways that are expected to only worsen the situation. This confluence of factors has led to increasing calls to make disaster risk reduction a core development concern, as well as to promote an understanding that disaster risk reduction is a development investment.

This growing emphasis on increasing resilience has occurred since about the end of the last decade at roughly the same time as the emergence of 'Big Data'. We conceptualize Big Data not just as large datasets, some of which have been used for decades in climatology, but as a new socio-technological phenomenon resulting from the emergence and development of an ecosystem made up of the new kinds of data 'crumbs' about human behaviours and beliefs generated and collected by digital devices and services, ever more powerful computing power and analytics tools, and a vibrant community of actors in this field.

This report explores the opportunities, challenges and required steps for leveraging this new ecosystem of Big Data to monitor and detect hazards, mitigate their effects, and assist in relief efforts. Ultimately the goal is to build resilience so that vulnerable communities and countries as complex human ecosystems not only 'bounce back' but also learn to adapt to maintain equilibrium in the face of natural hazards.

An overall conclusion is that Big Data for resilience, as with nearly everything with Big Data, is still "in its intellectual and operational infancy"[1]; most existing applications are small pilots, few formal evaluations exist, and much of the field consists of studies from the grey and white literature, case studies, and reports from NGOs, humanitarian organizations, and private companies. But based on the evidence available so far, Big Data does show real value and potential as a force for increasing social resilience, provided it is approached and promoted not merely as yet another technological fix.

Generally speaking, and particularly in disaster-prone regions, we find that Big Data can have four main roles or functions:

1. *Descriptive,* involves narrative or early detection such as using data from satellite imagery to identify flooded areas or identifying areas in need from crisis maps;
2. *Predictive*, includes what has been called 'now-casting' - to make real-time inferences on population distribution based on cell-phone activity before, during or after a shock, for example – as well as forecasting sudden and slow onset hazards;
3. *Prescriptive* (or *diagnostic*), goes beyond description and inferences to establish and make recommendations on the basis of *causal* relations, for instance by identifying the effects of agricultural diversification on resilience;
4. *Discursive* (or *engagement)*, concerns spurring and shaping dialogue within and between communities and with key stakeholders about the needs and resources of vulnerable populations such as crowdsourcing maps to assist disaster relief efforts;

The key stakeholders in this ecosystem who would have the ability to take small pilots to scale are numerous – private firms including telecoms and information technology companies, natural and social scientists, donors or investors, and a plethora of government agencies at all levels. However, no structures have been established to bring together these core stakeholders. Rather than collaborating to scale up pilot programs into sustainable systems, they generally continue to work in silos on projects driven by specific available technologies rather than by the needs and knowledge of at-risk communities.

Our research did find many examples of isolated experiments using specific technologies in response to specific events but virtually no insight on how to scale and connect these experiments. Our recommendations therefore focus on the need for investments and mechanisms to improve coordination among actors and technologies to realize the promise of Big Data in building resilience, tackle both 'single event' resilience and 'general purpose' resilience with the critical role of social learning, and above all place at-risk individuals and communities at the centre of these efforts to ensure their context-appropriateness and sustainability.

2. How can Big Data help? Learning from existing applications

It goes without saying—but it is worth restating—that Big Data is not a panacea. It is one potential force for increasing social resilience to disasters. The use of Big Data to build resilience generally falls into one of five categories throughout the disaster cycle, which rely on the four functions of Big Data described above. These five categories are:

1. **Monitoring hazards.** Seismographs, satellites, and drones offer ever-improving remote sensing capabilities. Adding vibration data from citizens' smartphones or information from their Twitter feeds offers tremendous potential for monitoring such hazards as earthquakes and floods.

2. **Assessing exposure and vulnerability to hazards.** Satellite images enable experts to identify geographical and infrastructure risks. Crowdsourcing initiatives like the OpenStreetMap project empower volunteers to add ground-level data that are useful notably for verification purposes. Call detail records – phone metadata tracking numbers and times of calls– have been used to estimate population distribution and socioeconomic status in places as diverse as the U.K. and Rwanda.

3. **Guiding disaster response.** Social media can be monitored to provide early warning on threats ranging from disease outbreaks to food insecurity. Remote sensing has been used to provide early assessment of damage caused by hurricanes and earthquakes. Mobile phone data provide precious information on population movements and behavioural response after a disaster.

4. **Assessing the resilience of natural systems.** Satellite images revealing changes in, for example, soil quality or water availability have been used to inform agricultural interventions in developing countries. Citizen science reporting via social media and other platforms can radically expand scientists' observations of ecological systems.

5. **Engagement of communities.** Building long-term resilience takes more than enhancing the ability of both external and local actors to react to single events. Resilient communities manage their natural systems, strengthen their infrastructure, and maintain the social ties and networks that make communities strong. The longer-term potential of Big Data lies in its capacity to raise citizens' awareness and empower them to take action. Decisions that facilitate or hinder this capacity are fundamentally political ones.

The full report provides numerous examples of such applications.

3. What are some of the key barriers, gaps and risks?

Despite some promising results, there are barriers, gaps and risks associated with the application and use of Big Data in supporting resilience in developing countries. Many of these challenges are similar to those that have emerged in related areas– notably in the ICT for Development (ICT4D)[2] and Participatory GIS (PGIS)[3] field. This includes for example, human and institutional capacity gaps and lack of access to internet and IT infrastructure.

Big Data also comes with specific technological, political, and economic hurdles to implementing and scaling new approaches as well as new risks. For example, cell-phone data's usefulness is currently hampered by factors ranging from the large size of data sets to be analysed to uncertainties surrounding individual and group privacy risk. Social media analyses that work well in upper and middle income countries may falter in poorer countries with much thinner and more skewed user bases.

These are essentially the same well-documented problems that affect Big Data for development and policymaking purposes generally. But leveraging Big Data in often highly complex and volatile environments adds to the need to be especially mindful of these factors when attempting to leverage it to build resilience:

1. **Constraints on data access and completeness.** For all the talk about the 'data deluge', most Big Data sets are in the hands – or, rather, on the servers – of private corporations, and as of yet no comprehensive frameworks and principles for data sharing exist. The tools to gather and process these data also tend to be difficult to use and expensive.
2. **Analytical challenges to actionability and replicability.** Big Data sets and streams face issues of reliability and representativeness that may hamper internal and external validities of findings derived from their analysis. Approaches to mitigate these effects such as verification techniques and sample bias correction methods have been or are being developed.
3. **Human and technological capacity gaps**. At present the capacity to gather and analyse data, as well as the ability to integrate it into policy making and programming are still largely lacking – especially among the institutions of the Global South. As stated by Claire Melamed of ODI and Data-Pop Alliance, ***"the explosion of big data has far-outpaced our ability to make sense of it in poorer nations that already lack human and technical capacity."***
4. **Bottlenecks in effective coordination, communication and self-organization.** The knowledge people need to inform risk assessment, preparedness and response efforts come from many sources that are rarely coordinated and socio-cultural and psychological factors are too often ignored, notably the need to build knowledge and exchange *networks* rather than provide information *products*.
5. **Ethical and political risks and considerations.** The potential for unethical or even dangerous use of Big Data grows exponentially in developing countries and there is an urgent need for developing ethical guidelines rooted in the long history of ethics in social science and medical research. Participation must be voluntary, users' data must be protected, and the needs of people without access to technology must be addressed.

Much of the return to investing in Big Data thus revolves around simply facilitating the management and use of existing data, or in simply increasing the likelihood that data known to be useful are being gathered and prepared before the next disaster hits.

4. What could a feasible roadmap entail and achieve?

The proven or potential benefits of Big Data has not yet translated into a clear roadmap indicating practical ways to build disaster resilience on the scale necessary to counter the risks faced by the world's most vulnerable. Remedying this gap requires investments in Big Data technologies, in the communities that support and use these technologies, and in the future of the field.

Within these three main areas, this report identifies the following **12 priorities:**

Invest in Big Data technologies

Many cutting-edge technologies have huge potential but need to be tested and refined in the field to scale-up.

1. **Get early warning systems and risk maps into the hands of the people who can use them**. Techniques to process crisis data from satellite images and the Internet already exist. Human-centred design and wide dissemination would make these models more useful.
2. **Invest in basic forms of existing applications that have high returns,** such as social media, mobile call record data, and crowdsourced approaches that combine machine and human computing.
3. **Identify high value-add contexts.** An example of a high value-add context is vulnerable populations in middle-income countries that are experiencing the greatest increase in disaster risk and, simultaneously, rapid growth in cell and social media technology.
4. **Facilitate the proper management and use of existing Big Data resources** by developing data sharing guidelines and common standards and designing innovative models and partnerships to enable rapid release of crisis data.
5. **Shift to an integrated "data portfolio" approach.** The most promising uses of Big Data combine data from a variety of sources. The effectiveness of resilience strategies will be shaped by the value of the data portfolio as a whole, not by individual technologies.

Invest in Big Data communities

Societal learning and a shared understanding of risks and opportunities are important components of resilience. The people, as much as the technology, drive the success of Big Data innovations. To build resilience, investments should focus on:

6. **Facilitating coordination among stakeholders** by, for example, fostering regional data ecosystems around key actors and activities to link grassroots groups and start-ups with large corporations, organizations and agencies.
7. **Spurring dialogue on ethics and privacy** with and between public officials and civil society organizations to understand and address privacy and other political and legal risks;
8. **Promoting and incentivizing private sector involvement,** via the organization of data challenges and promotion of financial and in-kind support to local start-ups and organizations;
9. **Spurring data literacy.** Big Data for resilience should not be left to experts only; a major requirement is to enhance people's willingness and ability to engage with and via Big Data to shape the future of the field.

Invest in the future of the field

The field of Big Data for resilience is young but growing rapidly. Investments in its future can have big payoffs for developing communities and the humanitarians who serve them. Steps should be taken to:

10. **Facilitate feedback in the disaster response community.** Simple mechanisms can be used, for example, to enable researchers and humanitarian agencies to share new knowledge and best practices.
11. **Tap mobile phone data more fully and rapidly.** Mobile phones are a critical technology entry point for people in developing countries, but the data infrastructure and processing capability lag behind.
12. **Synchronize Big Data sources.** Basic data from mobile call data records could be combined with social media and Internet data to inform policy by, for example, providing updated demographic data for risk assessments.

Much of the return on investment in the use of Big Data for resilience revolves around simply managing and using existing data before the next disaster hits.

Introduction

Natural hazards pose major risks to developing countries. Long-standing patterns of economic development expose poorer communities to more natural hazards and leave them less resilient than developed countries when hazards do occur.[4] The resulting disasters take more lives and lead to more damage.[5] Furthermore, growing evidence indicates that natural disasters leave difficult-to-observe indirect damages that can affect development outcomes ranging from economic growth to individual and community health.[6]

Several factors are expected to increase disaster risk in developing countries over the next few decades. Higher urban densities, larger coastal settlements, and increased investment are all expected to increase the number of people and assets exposed to hazards in developing countries. Climate change is increasing the frequency and intensity of weather-related hazards in varying, complex ways that can only be expected to worsen the situation.[7] This confluence of risk factors has led to increasing calls to make disaster risk reduction a core development concern[8] and position it as a development investment.[9] Several—close to 30 by our estimates—of the agreed 169 targets of the Sustainable Development Goals (SDGs) agenda[10][11] relate, directly or indirectly, to disaster risk reduction and increased resilience (see Table 1).

This growing emphasis on increasing resilience has occurred roughly at the same time as the emergence, since the end of the 2000s, of 'Big Data', which we conceptualize as an ecosystem made up of three factors: digital data from sources as diverse as satellites and mobile phones, the capacity to analyse and use that data, and the people who produce, analyse, and/or use the data (see **Box 1**). Big Data has opened up promising approaches to disaster resilience. Mobile phone data, for example, can provide an incredibly detailed view of population behaviour and movement in areas that were previously observed infrequently and indirectly. Social networks like Twitter are already improving the ability of humanitarian and disaster risk reduction organizations to monitor and respond to hazards. Further, opportunities are increasing as mobile phone penetration and access to internet, for example, are increasing significantly in developing countries. Over the 10 years there is likely to be an explosion of new data.

At the same time, leveraging and scaling Big Data approaches to increasing resilience requires navigating and linking highly complex technological, political, and socieconomic systems. On a basic level, the possibilities of using mobile phone data is hampered by factors ranging from access to the data, the size of the data, and privacy concerns. Methods of mining social network data that work well in upper- and middle-income countries are less adequate in poorer countries with thinner user bases. The limitations and requirements are multiple, making advanced assessment of the promise of various methods difficult even for specialists.

'The explosion of big data has far outpaced our ability to make sense of it in poorer nations that already lack human and technical capacity.'

—Claire Melamed

This synthesis report sheds light on this rapidly changing area by highlighting the growing body of empirical work that explores ways in which Big Data has been used to increase resilience.

Box 1. Key Terms and Concepts at a Glance

- **(Natural) hazard**
  A sudden or slow-onset natural event or process that may cause harm to humans or other organisms. Examples are floods, drought, earthquake, desertification, landslide, epidemic, and locust invasion.
- **(Natural) Disaster**
  As distinct from a hazard, a disaster consists of the combination of a natural hazard and its effect on the population and assets.
- **Vulnerability**
  The characteristics and circumstances of a community or asset that make it susceptible to the damaging effects of a hazard.
- **Exposure**
  The actual extent to which assets or populations are likely to experience a given set of hazards over time. Exposure plus vulnerability equals disaster risk.
- **Risk**
  The combination of the probability of an event and its negative consequences.
- **Impact**
  The sum of the consequences if the risk does occur.
- **Resilience**
  The ability of a system or community to resist, absorb, accommodate, and recover from the effects of a hazard, including preservation and restoration of essential structures and functions.
- **Big Data**

  An ecosystem made up of the combination of three factors: digital data from sources as diverse as satellites and mobile phones, the capacity to analyse and use that data, and the people who produce, analyse, and/or use the data. The concept of Big Data goes well beyond the datasets themselves—regardless of their size.

This report is organized as follows. The first section outlines the key concepts (see **Box 1**) and questions at stake. The next two sections draw from the academic and 'grey' literatures and from a set of case studies specifically commissioned for this project by the UK Department for International Development (DfID), the project sponsor, with NERC and ESRC. Section 2 provides an analysis of the potential of Big Data to increase resilience. Section 3 outlines the many pitfalls involved in transitioning from understanding to action. A constant theme is the extremely early stage of the field.

Table 1: Sustainable Development Goals and selected targets related to resilience

Goals	Selected targets
Goal 1. End poverty in all its forms everywhere	Target 1.5 - By 2030 build the resilience of the poor and those in vulnerable situations, and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters
Goal 2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture	Target 2.4 - By 2030 ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters (…)
Goal 3. Ensure healthy lives and promote well-being for all at all ages	Target 3.8 - Achieve universal health coverage (UHC), including financial risk protection (…)
Goal 4. Ensure inclusive and equitable quality education and promote life-long learning opportunities for all	Target 4.7 - By 2030 ensure all learners acquire knowledge and skills needed to promote sustainable development (…)
Goal 5. Achieve gender equality and empower all women and girls	Target 5.5 - Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic, and public life
Goal 6. Ensure availability and sustainable management of water and sanitation for all	Target 6.6 - By 2020 protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers and lakes
Goal 7. Ensure access to affordable, reliable, sustainable, and modern energy for all	Target 7.1 - By 2030 ensure universal access to affordable, reliable, and modern energy services
Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	Target 8.2 - Achieve higher levels of productivity of economies through diversification, technological upgrading and innovation, including through a focus on high value added and labour-intensive sectors
Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	Target 9.1 - Develop quality, reliable, sustainable and resilient infrastructure (…)
Goal 10. Reduce inequality within and among countries	Target 10.2 - By 2030 empower and promote the social, economic and political inclusion of all (…)
Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable	Target 11.5 - By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations
Goal 12. Ensure sustainable consumption and production patterns	Target 12.8 - By 2030 ensure that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature
Goal 13. Take urgent action to combat climate change and its impacts	Target 13.1 - Strengthen resilience and adaptive capacity to climate related hazards and natural disasters in all countries and Target 13.3 - Improve education, awareness raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction, and early warning
Goal 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development	Target 14.3 - Minimize and address the impacts of ocean acidification, including through enhanced scientific cooperation at all levels
Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	Target 15.3 - By 2020, combat desertification, and restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land-degradation neutral world
Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	Target 16.7 - Ensure responsive, inclusive, participatory and representative decision-making at all levels
Goal 17. Strengthen the means of implementation and revitalize the global partnership for sustainable development	Target 17.18 - By 2020, enhance capacity building support to developing countries (…) to increase significantly the availability of high-quality, timely and reliable data (…)

The report concludes with specific recommendations for operationalizing Big Data approaches to increase resilience, identifying both promising areas where Big Data adds value and the many practical hurdles to implementation. The capacity to gather, analyse, and use data is still largely lacking, especially in the Global South. The biggest returns on investment in Big Data thus involve managing and using data that already exist or ensuring that the kinds of data already known to be useful are being gathered before the next disaster hits.

Big Data is not a panacea. It is one potential force for increasing social resilience to disasters. Strong investment in Big Data and resilience is necessary to 'fully integrate statistics into decision making [and] promote open access to, and use of, data', according to the U.N. High-Level Panel report on the post-2015 agenda. This report highlights the need to invest in people and institutional capacities, make governance of technologies more open and transparent, and place attempts to strengthen resilience using Big Data under a cohesive and participatory framework. Building more adaptable societies and investing in effective long-term solutions is more complex than simply developing an app or performing an analysis.

1. Setting the stage and stakes

This section outlines key contexts, concepts, and questions necessary to understand the potentials and pitfalls of Big Data approaches to resilience.

1.1 Hazards, disasters, vulnerability and resilience in developing contexts

Disaster risk management and response have long been core humanitarian concerns. However, recent decades have seen substantial advances in our understanding of how natural hazards become deadly and costly disasters. The key is to recognize that the social impact of a disaster is a product of the interaction between natural hazards and vulnerable human communities (**Figure 1**). Disaster risk thus depends on a variety of social factors ranging from population density to urban planning to disaster warning and response systems. Improving a society's ability to withstand disasters requires greatly improved risk management. It requires changes in the many social variables that increase resilience.

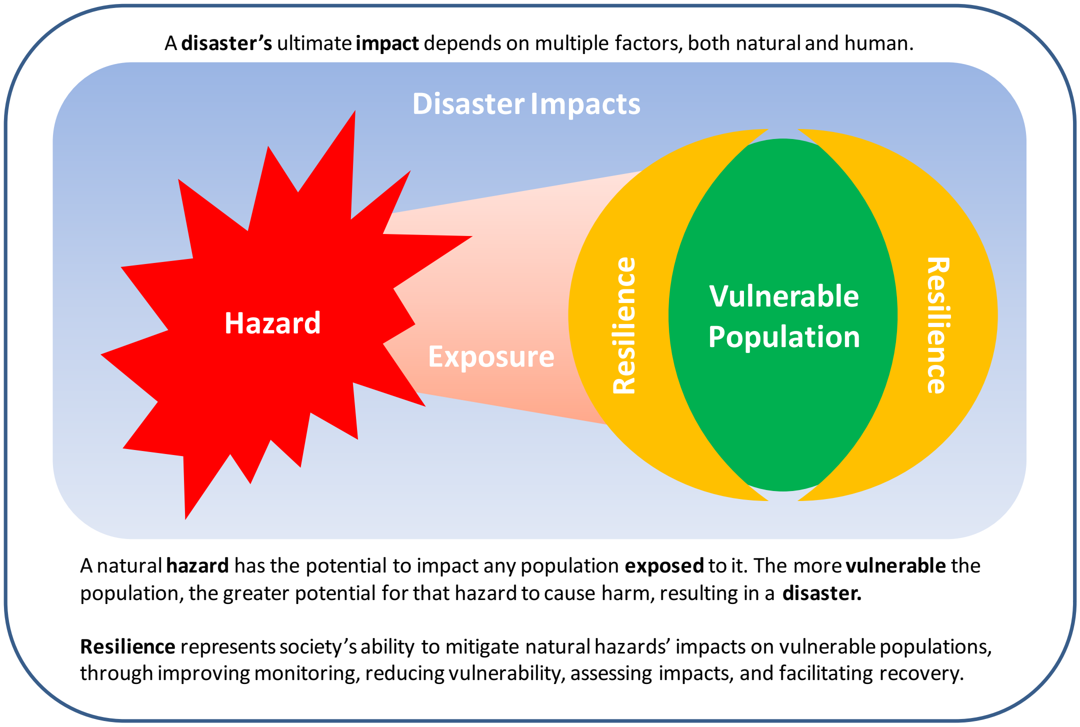
Resilience is the sum of the many factors and processes that can reduce vulnerability defined as *"the susceptibility of [a] system or any of its constituents to harmful external pressures"*[12], or simply the *"propensity for loss"*[13]. Whereas in most of the policy and social science literature and discourse, resilience is understood as the (considered normatively desirable) capacity of a social system to 'bounce back', in the ecological science where it was first developed the concept rather describes a system's ability to *"maintain structure and function"* while 'bouncing back' would be closer to a system's *robustness*.[14] As such, and fundamentally, a distinctive feature of a resilient system is *adaptability*.

Improved warning systems that enable populations to better respond to risk information, earthquake-resistant architecture, and disaster drills could all be expected to increase resilience. So would more subtle factors such as effective health systems and well functioning social networks.

In some cases, technical fixes may reduce system vulnerability in the short term, but increase it in the long term. For example, relying on groundwater may increase resilience to droughts in the short run but make more people more vulnerable to severe damage when groundwater reserves are depleted. Assessing a system's resilience requires consideration of possible unforeseen consequences of current actions in the larger context. Ultimately, societal learning is the key to building the flexibility that makes systems resilient.

The relationship between resilience and development is not simple. As populations become richer, their ability to respond and adapt improves due to increased income and access to technology, but so does exposure of vulnerable assets and populations, for example, in densely packed cities. Climate change further complicates this scenario by increasing the likelihood of extreme events in uneven patterns around the world.

Figure 1. Disasters and resilience



Source: elaborated by the authors

1.2 What is Big Data?

The term 'Big Data' has come into wide use as an umbrella term for both the new technologies that generate large-scale data on social outcomes (see **Table 2**) and the opportunities those data engender. Unfortunately, Big Data too often continues to be reduced to 'big data' characterised by the volume, velocity, and variety of the data, which overlooks most of Big Data's novelty and complexity. Instead, Big Data must be conceptualized as a new socio-technological phenomenon resulting from the emergence and development of an ecosystem made up of the new kinds of data 'crumbs' about behaviours and beliefs generated and collected by digital devices and services, ever more powerful computing power and analytics tools, and a vibrant community of actors in this field.

Following a previously developed framework,[15] this report conceptualizes Big Data as the complex social system created by the emergence of 'the 3 Cs of Big Data':

- Digital bread **crumbs**[16]**:** pieces of data that are the digital translation of human actions and interactions captured by digital devices[17], the majority of which are passively[18] emitted by users of digital devices and services. These fall under 3 main categories of 'Exhaust data', 'Digital content', and 'Sensing data'[19] (see **Table 2**)
- Big Data **capacities** or analytics: the set of tools and methods, hardware and software, and know-how and skills necessary to process and analyse these new kinds of data. The tools and methods include visualization techniques, statistical machine learning, algorithms, and the like.
- Big Data **communities**: the actors involved in the Big Data ecosystem, from the generators of data to their analysts and end users – potentially the whole population.

Table 2. Taxonomy and Examples of Big Data Sources

Types	Examples	Opportunities
Category 1: Exhaust data		
Mobile-based	Call Details Records (CDRs) GPS (Fleet tracking, Bus AVL)	Estimate population distribution and socioeconomic status in places as diverse as the U.K. and Rwanda
Financial transactions	Electronic ID E-licenses (e.g. insurance) Transportation cards (including airplane fidelity cards) Credit/debit cards	Provide critical information on population movements and behavioural response after a disaster
Transportation	GPS (Fleet tracking, Bus AVL) EZ passes	Provide early assessment of damage caused by hurricanes and earthquakes
Online traces	Cookies IP addresses	Mitigate impacts of infectious diseases through more timely monitoring using access logs from the online encyclopedia Wikipedia
Category 2: Digital Content		
Social media	Tweets (Twitter API) Check-ins (Foursquare) Facebook content YouTube videos	Provide early warning on threats ranging from disease outbreaks to food insecurity
Crowd-sourced/ online content	Mapping (Open Street Map, Google Maps, Yelp) Monitoring/ Reporting (uReport)	Empower volunteers to add ground-level data that are useful notably for verification purpose
Category 3: Sensing data		
Physical	Smart meters Speed/weight trackers USGS seismometers	Sensors have been used to assess the demand for using sensors to estimate demand for high efficiency cook-stoves at different price points in Uganda or willingness to pay for chlorine dispensers in Kenya
Remote	Satellite imagery (NASA TRMM, LandSat) Unmanned Aerial Vehicles (UAVs)	Satellite images revealing changes in, for example, soil quality or water availability have been used to inform agricultural interventions in developing countries

Together, these three parts form a complex system in which feedback loops in data generation, use, and assessment produce new data and techniques. At the most basic level, organizations generate new kinds of data that lead to the development of new kinds of analytical tools, and then various actors interact with those tools.

Despite the high potential for use of Big Data to improve global well being, implementation of Big Data initiatives raises a number of concerns in both the private and public sectors.

- Most Big Data investment has been made by the private sector, and corporations own much of the data. Although corporations have a legitimate right to act in their commercial interests, the privacy and rights of citizens and organizations must be protected.
- Technical and legal data sharing frameworks that would facilitate use of private corporations' data in the public interest do not currently exist. The concept of data philanthropy is being promoted to foster data sharing; corporate social responsibility can also be a driver.
- Government institutions generally lack both the IT infrastructure to collect and analyse Big Data and the financial resources to invest in technology development. They also often do not have the resources to distribute available data or foster interoperability among datasets.
- Most government organizations need to build the capacities of their staff and incentivize data sharing across departments and units.
- Governments also face challenges in establishing regulatory frameworks and work procedures to respect the privacy of citizens and organizations while handling their data.

1.3 How can Big Data increase resilience?

The value of Big Data for disaster and climate resilience lies in the insights that can be gained from combining data *crumbs* with enhanced *capacities* or analytics. There are four main types of analytics for resilience:

- **Descriptive analytics** is concerned with describing situations and critical concerns, for example, assessing damages from a disaster or combining satellite imagery and social media data for early detection of a flood.
- **Predictive analytics** involves making inferences about unobservable or difficult-to-measure concerns. For example, changes in call frequency, movements of mobile phones and mobile recharges ('top-ups') have been used to assess mobility and interaction patterns in response to disasters. Predictive analytics is also concerned with what may happen in the future. For example, Big Data is central to enabling granular, early, and accurate weather forecasts and can increasingly predict both sudden and slow-onset disasters.
- **Prescriptive analytics** goes beyond description and inferences to examine likely futures by identifying causal pathways. For example, to identify the most promising policies, analysts might explore multiple likely forecasts or run predictive analyses under different policy scenarios. Another prescriptive application is 'behaviour nudging', in which individuals' data are used in personalized reports on, for example, their energy consumption or exposure to health risks.
- **Discursive analytics** generates value for resilience through the third C, *communities*. Using Big Data for community engagement includes raising awareness about disaster risks and providing real-time feedback to enhance response and community-led preparedness.

The role of Big Data in enhancing society's ability to avoid disasters and improve resilience is primarily a long-term one. It follows an action cycle with three main components: prevention or mitigation, monitoring or prediction, and response or recovery.[20]

Big Data can help with *prevention or mitigation* by helping communities to map hazards and to characterize their exposure and vulnerability. For example, analysis of patterns from mobile data and online content can help policymakers understand the behaviour of communities and test their response to emergency plans and training.

Big Data also has potential to improve *monitoring or prediction*. In case of an earthquake or volcano, for example, crowdsourced hazard detection techniques can enlist citizens to provide information by sending pictures of the volcano's activity. Other techniques involve mining user-generated content on hazards, such as social media posts, or aggregating data produced by a variety of sensors. For example, the motion detectors built into mobile phones can help with seismic detection.

Big Data also allows better *response and recovery* efforts. Sensing data generated by electronic devices and information posted on social networks can be used to channel search and rescue activities. Online communities can help develop maps of affected areas or match requests for help with offers of assistance. Satellite and drone images can facilitate quick, large-scale assessment of the impact of a disaster by comparing pre- and post- event images of damaged buildings and infrastructure. When the focus shifts from response to recovery, Big Data can help local communities return to normal. For example, analysts can identify areas whose recovery is lagging behind through polls via mobile networks or social media mining. Volumes of digital transactions or data on supply chains before and after a disaster can improve understanding of the interaction between humanitarian assistance and local systems.

1.4 Key actors and activities

Currently, the main players in the field of Big Data for resilience are mix of development agencies, governmental agencies, local and international non-governmental organizations (NGOs), public and private donors, private companies, philanthropic foundations and academic initiatives.

Various United Nations agencies have begun experimenting with using Big Data methods and tools to build resilience. The UN Global Pulse consists of a network of innovation labs established in New York, Jakarta, and Kampala, where experts from the UN, governments, academia, and the private sector collaborate on research and projects on Big Data for development.[21]

The UN Development Programme (UNDP) is also active in Big Data for resilience. The UNDP regional office in Europe and Central Asia piloted the application of data science techniques to identify socioeconomic vulnerabilities and organized a 'data dive' in Vienna[22] to improve poverty mapping. The UNDP office in China established a laboratory in partnership with a local internet services company to experiment with using Big Data to support development goals.[23] Other promising Big Data work is being done by UNDP branches in Macedonia,[24] Armenia,[25] and Kosovo.[26]

Another UN agency, the Office for the Coordination of Humanitarian Affairs, in late 2014 set up a data lab in Kenya and released the Humanitarian Data Exchange,[27] an open platform to promote humanitarian data sharing, automation, and interoperability. More than 40 organizations joined the platform, which was tested during the Ebola outbreak in West Africa and the earthquake in Nepal.

In the non-profit non-governmental sector, the Rockefeller Foundation notably is supporting studies and applications in the field. Its 100 Resilient Cities initiative is aimed at strengthening the resilience of 100 urban areas worldwide not only against disasters but also against stresses that can undermine a city's stability, such as unemployment. Flowminder Foundation (which developed a case study reviewed in this report) is a registered non-profit entity supporting NGOs and governments in leveraging anonymised mobile phone location data and satellite data to improve public health. In academia, an active player is the Harvard Humanitarian Initiative, with which Data-Pop Alliance and some of the authors of this report are affiliated.

Among private sector entities, the activity of Orange Group is notable. As part of its corporate social responsibility programme, it made anonymised mobile phone datasets from Ivory Coast and Senegal available to allow research groups worldwide to explore how Big Data could tackle global development challenges.[28] Among the hundreds of submissions, a few were specifically related to resilience—including identifying areas vulnerable to floods and to monitor call behaviour during floods,[29] detecting anomalies in human mobility patterns,[30] and linking changes in collective mobile data to emerging crises.[31]

Similarly, broadband and telecommunications provider Telefónica has also been contributing to research in this area. In partnership with the Universidad Politécnica de Madrid, the firm gained insight into the effects of floods in the Mexican state of Tabasco in 2009 by analysing, after the fact, millions of mobile phone datasets.[32] The study enabled identification of the most affected areas, provided knowledge on the size and behaviour of affected populations, and highlighted the mismatch between the population's awareness of risk and civil awareness-raising activities.

'Technology – no matter how well designed – is only a magnifier of human intent and capacity. It is not a substitute.'

—Kentaro Toyama

Despite some promising results, these efforts to leverage technology for global development have revealed challenges and lessons learned – notably in the fields of information and communication technologies for development [33] and participatory geographic information systems.[34] Such efforts have highlighted a number of political, economic, scientific, and technological issues that prevent the smooth transfer of technologies to alleviate poverty. As Toyama[35] notes, 'Technology – no matter how well designed – is only a magnifier of human intent and capacity. It is not a substitute.' Solutions developed by the richest countries or international NGOs often fail – either because of the gap between design and reality[36] or because the least developed countries and regions do not have the resources to collect and analyse data or to leverage local communities to use the data that exists.

1.5 The Uk-funded case studies and pilot projects

In addition to existing research, this report draws on 11 case studies and pilot projects commissioned by DfID with NERC and ESRC to explore Big Data for resilience; see **Box 2**. The case studies highlight the extent to which Big Data approaches to increasing resilience are still in early stages and face unusual hurdles. Even basic uses of mobile phone data records, for example, require a great deal of effort to acquire and analyse the data. These concerns were echoed at a workshop DfID held to inform this report in London on June 5, 2015. Sections 2 and 3 draw on feedback from the case studies and the workshop, which also informs this report's final recommendations.

Box 2. Overview of Case Studies

The field of Big Data for resilience is rapidly evolving. In the 11 case studies and pilot projects DfID with NERC and ESRC commissioned to explore the links between Big Data and resilience, four themes emerged.

Theme 1: Building resilience through crowdsourcing

A number of papers investigate the use of crowdsourcing to characterise hazards, identify localised needs, and diagnose disaster response on the ground. Social media and text messages are key tools. The case studies in this group document both rich promise and significant limitations in use of crowdsourcing.

1. Early Flood Detection for Rapid Humanitarian Response: Harnessing Big Data from Near Real-Time Satellite and Twitter Signals (Jongman et al.)
2. Increasing Resilience to Natural Hazards Through Crowd-Sourcing in St. Vincent and the Grenadines (Mee & Duncan)
3. Inclusiveness in Crowdsourced Disaster Response (INCROWD) (Roth & Luczak-Roesch)

Theme 2: Using mobile network data to understand actions, behaviours, and attitudes

In developing countries, observational data is often lacking, and mobile phone usage is growing rapidly. Mobile networks have the potential to generate a clear picture of actions and contexts on the ground. The potential and challenges emerge in the following case studies:

4. Mobile Network Data and Climate Resilience: Analysis of Cyclone Mahasen in Bangladesh Using De-Identified Data of Five Million Phones in the Grameen phone Network (Bengtsson et al.)
5. Big Data for Flood Resilience in East Africa (Iliffe et al.)
6. Leveraging Mobile Network Big Data for Disaster Risk Reduction: Minimizing Harms and Facilitating Access (Samarajiva & Lokanathan)

Theme 3: Improved statistical methods for defining disaster risk

A critical area of innovation is the identification of statistical tools for analysing the vast quantities of information available. Drawing on lessons from other fields, the commissioned papers highlight promising statistical techniques for improving the relevance and accuracy of data in support of resilience and reducing the computational needs and time required for Big Data analysis. These improvements will be especially crucial for near real-time analysis in developing countries.

7. Landslide Susceptibility Mapping in Data-Poor Environments (Cheng)
8. Big Data for Tsunami Hazard Warnings in India (Guillas)

Theme 4: Big Data and communication technologies for awareness raising and disaster relief and recovery

Big Data approaches, especially use of mobile phones, offer unique opportunities for disseminating information and raising risk awareness at scale, as the papers in this area highlight.

9. Mobile-Based Disaster Risk Monitoring System: An Innovative Approach to Enhance Community-Led Disaster Preparedness in Uganda (Kiragga et al.)
10. The Potential of Big Data to Encourage Long-Term and Preventative Disaster Risk Reduction Behaviours: Evidence from Cochabamba, Bolivia (Sou)
11. Big Data in Disease Disaster Management in Developing Countries: A Mobile Phone Data Use Framework (Cinnamon et al.)

2. Opportunities and potential of Big Data for resilience

This section explores applications of Big Data to provide information about the variables that constitute the hazard equation: hazard risks plus the exposure and vulnerability of communities. It next reviews an area in which development and grassroots communities have already made important strides: using Big Data to detect and respond to single events. Turning to longer time scales, we show how information on exposure, vulnerability, and disaster impact can be used to build resilience over time. Finally, this section emphasizes the importance of social learning and the potential of Big Data for strengthening democracy and enhancing communities' capacity to act.

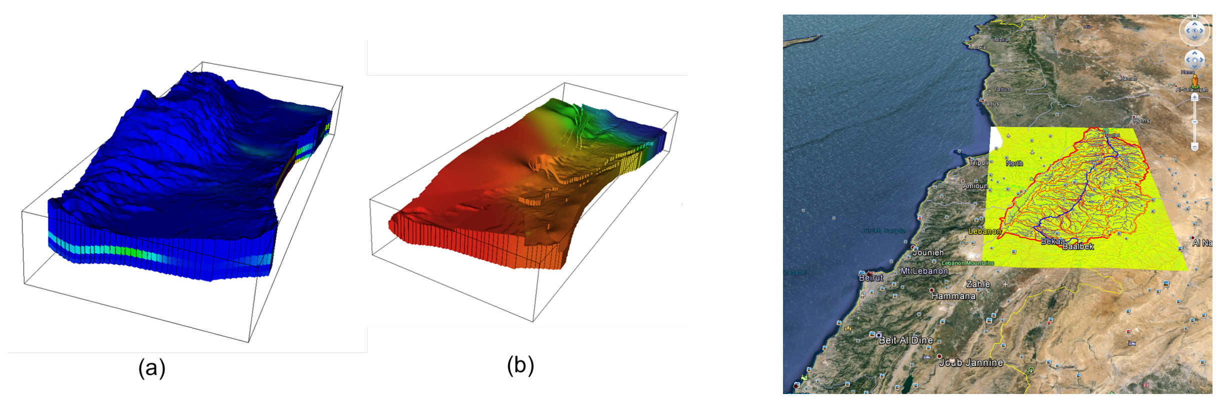
2.1 Monitoring hazards

Effective monitoring is an important foundation for improved management of disasters. Many aspects of hazard monitoring, particularly in the geosciences, are already heavily driven by Big Data sources such as remote sensing. The added value of new Big Data techniques in this area breaks down into advances that improve existing systems and advances that use new sources of data to monitor risks.

Improving existing systems

One of the prime opportunities in using Big Data to monitor hazards lies in advancing geoscience risk measurement systems. In the past decade, increased availability of high-resolution satellite sensors has contributed to great improvements in hazard detection and mapping. For example, data collected from the NASA Gravity Recovery and Climate Experiment satellites launched in 2002, enable effective monitoring of groundwater depletion. New sensors from the NASA Soil Moisture Active Passive mission will soon provide soil moisture data with unprecedented accuracy, resolution, and coverage. Jointly, these data enable analysts to detect loss of resilience in dryland ecosystems, monitor drought, and estimate yields in data-sparse and food-insecure regions (see **Figure 2**).

Figure 2. Three-dimensional model of the Groundwater (left) and Google Earth hydrography map (right) of the Al Assi (Orontes) river basin in Lebanon



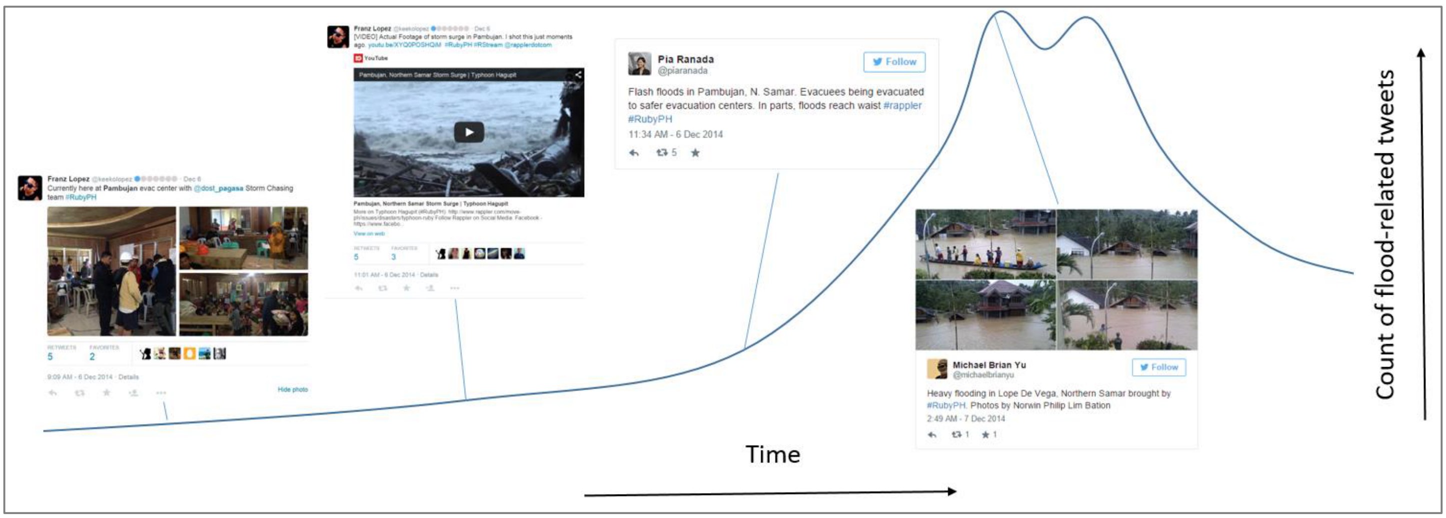
Source: UNESCO, "Science diplomacy and transboundary water management – The Orontes River case", 2015

Another opportunity lies in expanding the use of Big Data to enhance early warning systems. The commissioned case study 'Big Data for Tsunami Hazard Warnings in India' (**Case study 8 above**) presents a hazard model for assessing tsunami risk using satellite surface wave information and GPS observations. It finds that reasonably accurate estimates of inundation risk are calculable even in a data-poor environment like India during the 2004 Indian Ocean earthquake. However, the complexity of the necessary computations would make it difficult to use the technique on current data. Advances in technological infrastructure and capacity seem poised to radically expand the accuracy and usefulness of monitoring systems in coming years.

Using new sources of data to monitor risks

New sources of Big Data can facilitate innovative solutions to compensate for the difficulties of hazard detection and mapping in data-scarce environments. For example, the US Geological Survey now integrates social media surveillance into its network of seismometers to improve tracking and real-time mapping of landslides and earthquakes. The case study 'Early Flood Detection for Rapid Humanitarian Response: Harnessing Big Data from Near Real-Time Satellite and Twitter Signals' (**Case study 1 above**) provides another example of how social media can be used to gather real-time images and descriptions of developing situations (see **Figure 3**).

Figure 3. Schematic display of a typical Twitter count pattern leading up to a flood event



Source: Jongman et al. "Early Flood Detection for Rapid Humanitarian Response", 2015

In the Philippines and Pakistan, two sources of near real-time data, the Global Flood Detection System and the Twitter-based analytics platform Floodtags, were combined to support disaster monitoring. The approach worked well in unexpected and contentious flood events, such as intentional breaching of flood defences, as well as in densely populated urban areas. Questions about processing of social media content, identification of critical thresholds, and sampling and distributional issues require answers before similar applications can deliver results. The success of similar schemes may ultimately depend on working through local organizations to make people aware that their communications are being used to inform effective disaster response.

Social network data need not be passively emitted to be useful. Crowdsourcing, or soliciting data from the public, is another key opportunity. The citizen science literature has demonstrated that certain types of scientific data can be reliably gathered by distributed networks of non-specialists. One of the UK-commissioned reports outlines how crowdsourced data has contributed to monitoring of volcano hazards (**Case study 2 above)**. It shows how even a small number of dedicated users can bring large increases in understanding in data-scarce areas. Uganda's use of uReport, the cell-phone based two-way governmental messaging system (**Case study 9 above**), provides another example. Ongoing efforts highlighted in the commissioned report on mobile-based disaster risk monitoring in Uganda demonstrate the potential for crowdsourced reporting of local hazard outcomes.

Additional technologies currently in development that are likely to shape hazard monitoring during the next decade, though the uncertain nature of technological progress makes it difficult to tell how any of these technologies will play out. One of the most obvious changes are major advances in sensor networks. The accelerometers that detect motion in mobile phones can be used to get very rapid data on earthquake occurrence and intensity. Hand water pumps fitted with sensors that report on use can generate a cascade of potential hydrological data for hazard monitoring. In the longer run, embedding sensors into entire cities, as in India's 100 Smart Cities project, will provide novel opportunities to monitor hazards in real time, particularly when they are combined with existing monitoring systems.

2.2 Assessing exposure and vulnerability to hazards

Detailed information about the exposure of communities and assets to potential hazards is essential to estimating risks and monitoring resilience. Big Data has tremendous potential to identify the exposure and vulnerability of communities and countries, providing information on, for example, the presence of infrastructure, population density, and the socioeconomic status of the population. Large returns could be realized, even in the short term, by investing in Big Data applications and scaling them up. Furthermore, through traditional methods of mapping hazard exposure typically require vast quantities of geo-environmental data and inventories of past data, many of the methods described below show promise in data-poor environments where historical records may not exist (see **Box 3**).

In recent years, very different approaches have made great strides in mapping settlements, buildings, and infrastructure at risk, especially in settings with insufficient capacity to survey the land systematically. First, new satellite imagery products available at relatively low or no cost have spurred the development of new algorithms to map urban extent and building types. Second, communities around the globe have used crowdsourcing approaches to create detailed maps for vulnerability assessment. Finally, call detail records – phone meta-data – have also been used to assess exposure and diagnose disaster preparedness.

New satellite imagery

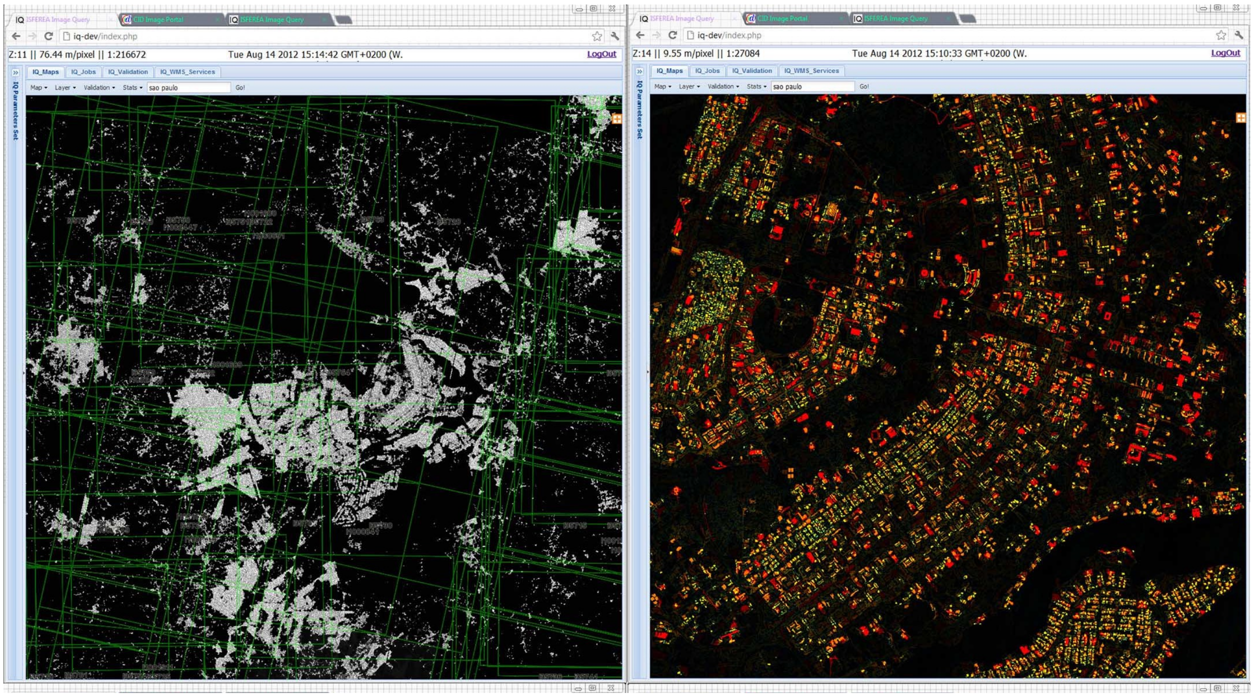
Box 3. Using Data Mining to Create Landslide Exposure Maps in Data-Poor Contexts

Reducing the risk and impact of landslides is a top resilience priority for many developing countries. Producing susceptibility maps typically requires a great deal of current and historical data – which few developing countries have. The UK-commissioned study 'Landslide Susceptibility Mapping in Data-Poor Environments' describes how a data-mining algorithm called Random Forest can be used to produce landslide susceptibility maps in contexts with poor access to data. Focusing on Piedmont, in Northwest Italy – where a comprehensive landslide inventory and full geo-environmental data are available as benchmarks – the paper shows that data mining can successfully predict landslide susceptibility, producing risk maps at up to 75% accuracy. Addition of landslide data would bring significant improvement.

Previous approaches to remote sensing, which used coarse data and relied on physical models built and interpreted by experts, rarely allowed global coverage, frequent updates, or the fusion of data from multiple platforms.[37] Improvements in the velocity, volume, and variety of satellite imagery data, along with automated methods for processing and aggregating data, have been a boon for exposure mapping. In the last decade, new land cover global layers including GlobCover[38] and MODIS Land Cover Type[39] have classified urban areas with about 96% accuracy.[40] Together with global population layers, these tools are well suited for systematic risk analysis in data-poor countries, though only at large spatial scales.

Even more significant in scope and ambition are two projects that will provide worldwide mapping with unprecedented spatial detail: the Global Human Settlement Layer (GHSL) by the European Union's Joint Research Centre and the Global Urban Footprint (GUF) by the German Aerospace Center (see **Figure 4**).

Figure 4. The city of Brasilia from the current Global Human Settlement Layer, using a combination of images from different satellites with resolution ranging from 0.5 to 10 m.



Right: the presence of buildings GHSL layer represented at 1:50K scale (the dark green shows all the input scenes used); Left: a zoom into the city center showing the average building size at 1:10K scale.

Source: Pesaresi et al., 2013

Both projects are in the testing phase, yet recent validation shows them to be reliable.[41] Both provide unprecedented ability to detect small and informal settlements. The GHSL will likely be the more flexible and detailed project, as it can quickly store, retrieve, and integrate large amounts of heterogeneous image data. It will provide fine-grained data on building sizes, types, and numbers.[42] Its very high-resolution imagery will even produce maps that show how vulnerable buildings are based on such factors as roof quality and building age.

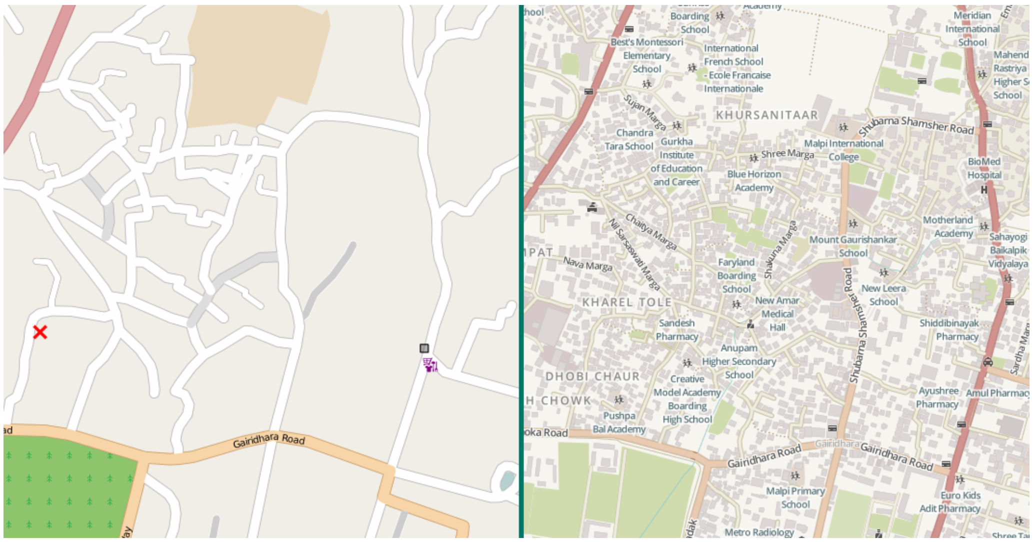
Data at this level of detail is costly. As part of the Group on Earth Observation (see **Box 7**), GHSL will be adapted for use in developing countries. Users will be able to model population distribution, plan censuses, map poverty and slums, model urban climate change, and much more. The success of this ambitious project requires commitment from funders and commercial satellite data companies to make the very high-resolution imagery available and to train development professionals to use these new fine-scaled maps.

Crowdsourced mapping

Crowdsourced or participatory mapping is the second important breakthrough in exposure mapping. Crowdsourced projects enlist volunteers to map geographic information, based on local knowledge, imagery data, social media including spatial information (in text, videos, photos), or a combination of these. The most well-known initiative is OpenStreetMap (OSM). Created in 2004, OSM is now a globally distributed organization – counting 1.5 million registered users and local groups in over 80 countries – working to create a common open digital map of the world.[43] OSM consists of a single database that users edit remotely, digitizing the presence of roads, buildings, and so on. OSM users self-organize into sub-groups that focus on geographic areas. Currently, OSM coverage is uneven, but its quality is high in comparison to official data.[44]

OSM data is useful for both hazard exposure and humanitarian relief efforts. The Humanitarian OpenStreetMap Team (HOT), a sub-branch of OSM, focuses on disaster applications. HOT quickly provided relief organizations with detailed maps after the 2010 earthquake in Haiti, Typhoon Yolanda in 2014, and the 2015 earthquake in Nepal. HOT's Missing Maps Project is mapping the most vulnerable places in the developing world to facilitate better response to future crises[45] (see **Figure 5**).

Figure 5. Comparison of OpenStreetMap coverage of Kathmandu, before and after the 2015 Nepal earthquake and the efforts of the Humanitarian OpenStreetMap Team



Source: Humanitarian OpenStreetMap Team website, accessed August 2015.

Other HOT projects map specific vulnerabilities. For example, a HOT team in Tanzania recently started mapping infrastructure vulnerable to flooding in Dar es Salaam.[46] Since 2011, HOT in Indonesia has been collecting exposure data to feed into open-source risk modelling software.[47]

Crowdsourced mapping provides local exposure information that governments may not have. However, little is known about how coordination works in crowdsourced projects like OSM, who participates, and how the organization could be improved to increase accuracy, consistency, and coverage.[48] More thorough evaluation of crowdsourced mapping projects can identify ways to scale up these initiatives and make them more useful.

Call detail records

To add to population maps like LandScan, which is based on census data, data scientists have experimented with using call detail records (CDRs) from telecom operators to estimate population densities,[49] movements of people,[50] and socioeconomic status. CDRs, or call meta-data, are available at the level of the individual or of the tower; they provide information on the number of calls between towers and on airtime purchases. With mobile-phone penetration rates greater than 90% in developing countries[51], CDRs could provide extremely fine-grained and dynamic data on large populations.

CDRs have many potential applications for exposure and vulnerability assessment, some of which have been well tested, while others have not yet been tried. CDRs were used to study population dynamics in Europe, Haiti, and New York.[52] The information available, although it has not yet been applied in real decision-making contexts, could be useful for designing evacuation routes, providing exposure maps that are sensitive to regular population movements such as daily or seasonal migrations, evaluating the effectiveness of early warning systems (see **Box 4**), or assessing risks of disease outbreaks following a disaster.[53] A number of projects (in the UK,[54] Ivory Coast,[55] Rwanda,[56] and Latin America[57]) have shown that measures of socioeconomic status can be derived by coupling CDRs with census or survey data to determine the relationship between calling patterns and wealth or income. CDRs can then be used to interpolate between censuses and extrapolate to populations that are not covered by official surveys, such as informal settlements.

Box 4. Auditing Early Warning Programmes With CDR-Based Maps of Population Movement

In one of the case studies commissioned for this report, a team led by Flowminder examined information collected from de-identified data of five million phones in the Grameen phone network in the wake of a cyclone in Barisal and Chittagong (**Case study 4 above**). The data included changes in call frequency, SIM movements within the network, and mobile recharges. Findings suggest that spikes in call frequency precede users' exposure to the storm, suggesting increased communication as communities prepare to be impacted. Similarly, interruptions in network function can be used to infer damage to infrastructure or power grids as towers go off line. Interestingly, despite early warnings, the data show a lack of mass displacement from coastal areas in the weeks and days preceding the cyclone. The case study therefore showcases how mobile network data can be used to audit the performance of early warning programmes. As mobile phone use proliferates in developing countries, such methods may help with assessing the impacts of extreme events and evaluating the effectiveness of disaster response. However, another of the commissioned case studies[58] stresses that telecom information is proprietary and subject to strict rules (**Case study 5 above**), so that relationships between researchers and network providers must be carefully managed and negotiated.

Several potentially powerful applications of CDRs to vulnerability assessment should be tested in the years to come (see **Box 6**). In particular, CDRs can be used to assess the features of social networks, which are vital when disaster strikes.[59] Quantitative studies based on CDRs confirm the importance of social networks: Rwandans with on-going economic relationships were more likely to receive remittances to help them after a disaster, and inhabitants of Port-au-Prince were found to take refuge with family members in other regions of Haiti following the 2010 earthquake. CDRs have been used to map friendship networks at the scale of whole societies.[60] This capacity could be leveraged to create measures of social capital and of the economic or social marginalization of specific communities. CDR population mobility data could also be used to infer the size and stability of market hubs and the robustness of transportation and service infrastructures.

Advances in the use of satellite imagery, crowdsourced maps, and CDRs allow analysts to map informal settlements and track poverty; they could easily be used to measure social and economic marginalization. However, these data do not capture some aspects of vulnerability, such as individuals' age, education, and health. Other aspects, such as access to water and electricity or the size and health of livestock, have not yet been tackled.

2.3 Disaster response: Early warning, situational awareness, and immediate impacts

The use of Big Data to respond to natural hazards depends on who has access to the data and tools and how effectively the analysis can inform decision making. The three parts of disaster response covered in the literature and the UK-commissioned case studies are early warning, situational awareness, and immediate impacts.

Early warning

In the wake of the International Decade for Natural Disaster Reduction (1990–1999) and subsequent international efforts, effective institutions have been created to coordinate the collection, analysis, and sharing of weather data to predict floods, storms, and droughts – though obviously major gaps still exist, subject to local political and technological capacities. The telecommunications systems in several of the world's least developed countries require upgrade, and progress can still be made in the lead-time of warnings, especially for floods and storm surges in coastal areas. Still, these institutions are effective at channelling data and at forecasting and detecting many climate-induced hazards, and they continue to improve.[61]

Situational awareness

As discussed above, crisis detection increasingly draws on digital social data. For example, the US Geological Survey monitors tweets worldwide to detect earthquakes and issue alerts.[62] The Billion Prices Project at the Massachusetts Institute of Technology (MIT) monitors prices posted online to detect inflation trends and monitor food security.[63]In public health, digital disease information is being used to alert individuals and governments to possible outbreaks (see **Box 6**). Lessons from the 11 commissioned case studies show what is required in order to capitalize on these opportunities in an emergency: technical capacity to design dissemination platforms and tools; adequate framing delivery mechanisms; and, above all, clear, standardized guidelines for the dissemination of disaster-related information.

Immediate impacts

Multiple sources of data can help with timely assessment of the effects of a disaster.

Remote sensing in particular serves three main functions:[64](a) providing large-area reconnaissance to enhance situational awareness and map damage; (b) assessing damage to properties critical for livelihoods and stability, such as homes and businesses; and (c) determining impact on critical infrastructures including roads, energy grids, and water pipes. In one project, remote sensing provided positive preliminary results in estimating the impact of several types of disasters by means of specific applications such as the use of "structured light" laser scanning devices for generating 3D spatial models of damaged areas.[65] Satellite imagery has enabled timely post-hurricane damage assessment of tropical forests[66] and accurate damage assessment after a tsunami.[67] Relatively high accuracy in the detection of damage after earthquakes was reported in Haiti[68] and Japan.[69] Novel methods for automated processing are being developed.[70]

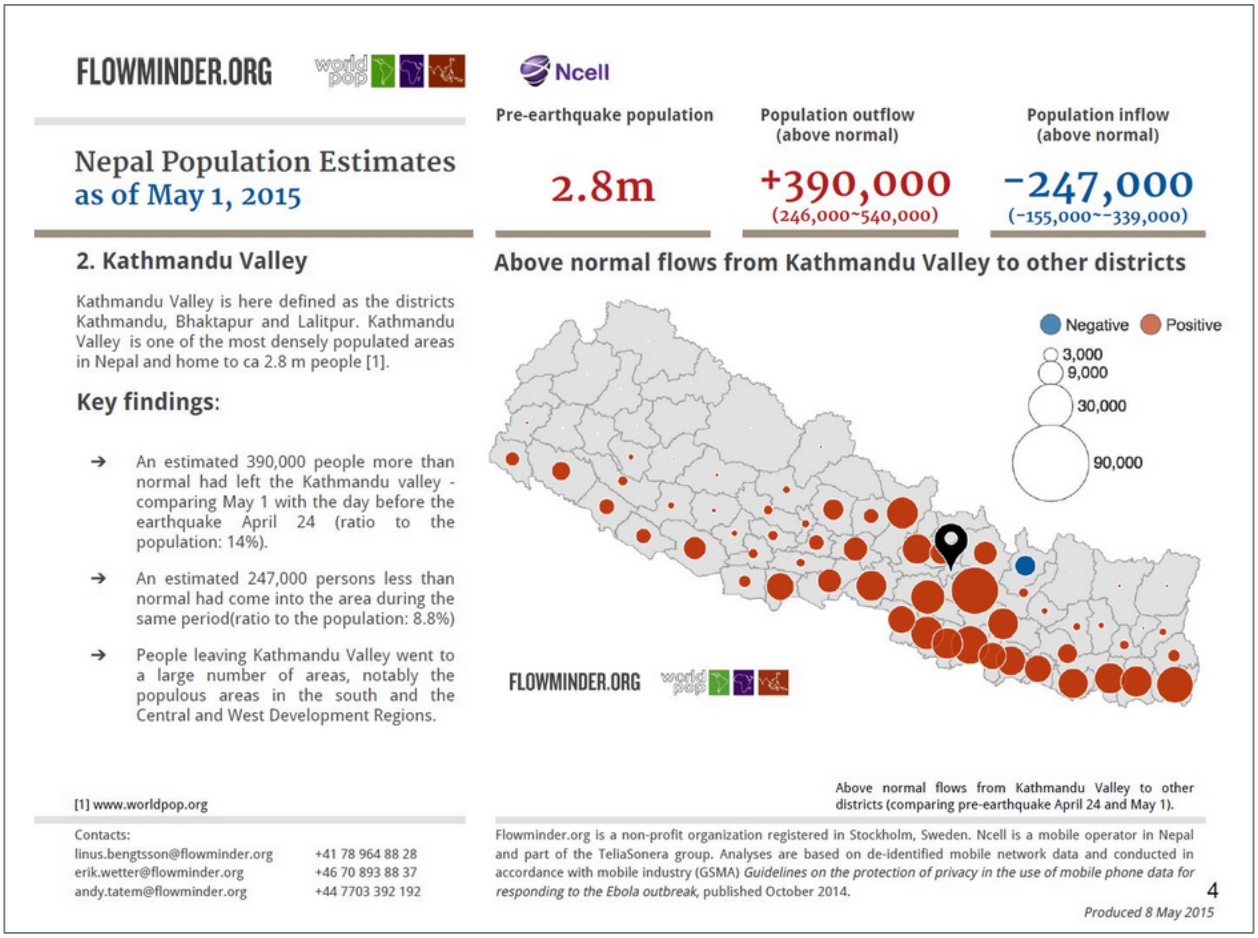
Crowdsourced approaches can complement ground and remote sensing data to reach a more fine-grained and dynamic understanding of impacts. Due to the increasing availability of satellite imagery from government sources as well as images from drones, crowd-based micro-tasking, a crowd-based approach that break the analysis process down into smaller steps that can be carried out by a network of simultaneous contributors – often volunteers – provide promising ways to identify damages, has become a promising way to identify damages.[71] In the Russian wildfires of 2010, a group of digital volunteers used social media to coordinate volunteer firefighters on the ground and provide relief assistance.[72] When they are open, decentralized, and interactive, Big Data and digital technologies can improve the ability of individuals affected by disasters to mobilize their social networks and get help from the 'crowd' of first responders. They can also help to coordinate the efforts of relief organizations. Indeed, health workers in the Haiti earthquake spontaneously attempted to coordinate their activities on Twitter, thereby revealing a latent demand for such an information platform.[73] One of the UK-commissioned case studies showcases how social media data facilitates a grassroots approach to digital humanitarianism (**Case study 3 above**) by giving local actors the ability to voice their needs.[74]

Box 5. Leveraging Mobile Phone Data to Infer Post-Disaster Movements

Data from mobile phones constitute a precious source of information to infer population mobility after disasters.

- Tracking the position of mobile phones in Haiti before and after the 2010 earthquake allowed experts to estimate population displacement and helped build an effective real-time monitoring system to track the outbreak of infectious diseases. [75]
- Scholars used CDR to infer internal migration patterns in Rwanda[76], determining a baseline for mobility that may be useful to assess after a disaster. [77] This baseline may also support policymakers in defining where to invest to strengthen local infrastructures.
- CDRs used to understand human behaviour during the 2009 floods in Tabasco, Mexico, proved to be representative of the population when compared with official census data[78];; furthermore, the data offered interesting insights on the impact of the disaster and on the citizens' awareness. [79].
- After a 2011 earthquake, Statistics New Zealand mapped population movements by tracking text messages and voice calls.[80] The experiment showed 'which geographical areas attract high percentages of people, patterns of return movements over time, and flows of non-residents into the emergency zone'. However, it did not assist in 'verifying residential areas people leave from, which areas people relocate to following an event, or the actual number of people who relocate, temporarily or permanently'.
- More recently, Flowminder used a similar approach to study population movements in Nepal in the aftermath of the April 2015 earthquake (see **Figure 6**).

Figure 6. Post-earthquake population movement in Nepal



Source: Flowminder, 2015

Box 6. The Possibilities and Limitations of Digital Disease Detection

Digital platforms such as HealthMap and the Global Public Health Intelligence Network have shown promise in detecting disease outbreaks. These platforms mine disparate web sources, using advanced natural language processing tools, to alert users of possible outbreaks. Such systems are credited with helping to detect outbreaks of severe acute respiratory syndrome (SARS) in November 2002. Some systems look for keywords in social media and search engine entries. For example, Google's Flu Trends, launched in 2008, initially garnered much attention for its predictive power. Although research has demonstrated a correlation between search queries and the number of physician visits related to flu-like symptoms in a given week, this approach has limitations.[81] For instance, Google Flu Trends does a much better job of predicting nonspecific respiratory illness, like a bad colds or SARS, than actual influenza. Another issue is confirmation bias: If the arrival of influenza is widely covered in the media, more people will search terms related to the virus, not because they are sick but because they are more aware of the disease—which was the case with Google Flu Trends. Thus, although digital social data has the potential to help detect public health crises, it is also subject to external factors that can lead to false positives – as is the case with most malfunction detection systems and forecasting models.

Big Data approaches can help coordinate social response in more structured ways as well. In a conflict-related example, the Nairobi-based NGO Sisi ni Amani sends text messaging to help resolve violent riots and human abuse in Kenya, using its over 30 million mobile phone subscribers to tap both pre-existing communities and groups recruited explicitly for this digital peace effort. Tangible work so far mostly has had to do with land disputes, which often lead to unrest in poor Kenyan communities. For example, in response to a land dispute in the Mulot-Narok region in October 2011, Sisi ni Amani sent text messages to the individuals involved. One read, 'We the people of Mulot shall resolve all boundary issues peacefully, for only with peace can we find lasting solutions...'. Texts like this one help to alleviate pressure and end regional clashes. Phones and existing networks spread the message deeply into the community at a relatively low cost.

Furthermore, humanitarian organizations have pioneered new methods to assess impact during and immediately after disasters. Data from social media, for example, increased the accuracy of disaster impact assessment during the 2013 floods in Colorado.[82] Geo-tagged tweets with pictures of flooded areas were combined with remote sensing imagery to optimize post-disaster reconnaissance. Recent experience in Indonesia also suggests that mining data from tweets can support emergency response by fine-tuning ground relief efforts.[83]

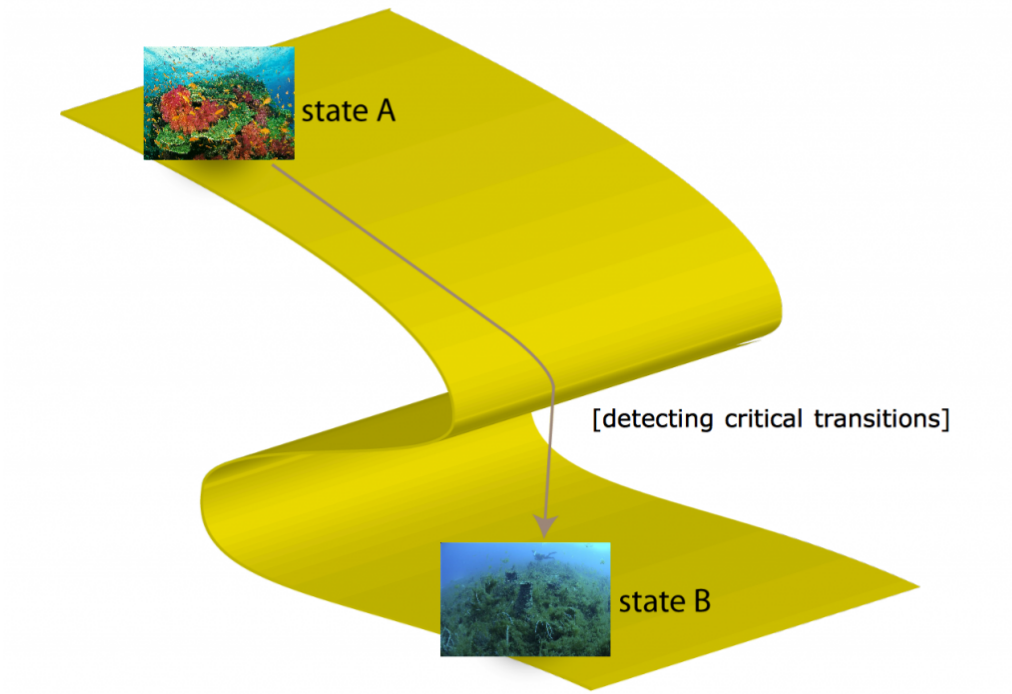
2.4 Innovative approaches to assessing the vulnerability and resilience of natural systems

Big Data can help monitor the resilience of both agricultural systems and natural ecosystems. Sustainable intensification of agriculture can make a significant contribution to resilience by mitigating climate change, the water crisis, extreme weather events, biodiversity loss, underemployment, inequality, and food crises.

Scientists at the Consultative Group on International Agricultural Research have recently introduced the Global Intervention Decision Model to help improve agro-ecosystem interventions.[84] This model flexibly incorporates ecological, economic, and social data, as well as the qualitative insights of practitioners. In 2013, the model was applied to nine pilot projects ranging from an irrigation project in Ghana to the intensification of rain-fed agriculture in the Tana River Basin. The model informed practitioners' views of project risks and told them what to measure. For example, it revealed the need for more information on 'points of failure' of the seed distribution network in Ghana. Evaluation of this model concludes that it should immediately be used in future interventions.[85] The model is more powerful when social and environmental variables are taken into account over time and space rather than using instead of using averages. This is where Big Data – particularly satellite imagery that provides dynamic, high-resolution data on such vital factors as soil quality and water availability – can make a big difference.

Big Data can also contribute to early-warning indicators for systems that are approaching a 'tipping point'—a threshold beyond which the system takes on different behaviour as it moves swiftly to a different equilibrium[86] (see **Figure 7**).

Figure 7. Diagram illustrating a tipping point, where a system shifts rapidly from one equilibrium state to another



Source: Early Warning Signals Toolbox

Examples of tipping points include grasslands turning into shrub land,[87] groundwater tables shifting to a different equilibrium height,[88] and the collapse of food webs. Such changes make systems less resilient, yet early detection is difficult because the changes are often sudden. However, patterns in the time series of certain variables can indicate that the system is approaching a critical transition.[89] Big Data methods, though they are still in development, have been applied with promising results to, for example, the sustainability of harvested fish stocks[90] and the weakening of the El Niño weather pattern.[91] The approach could also be used in drylands to monitor desertification.[92]

2.5 Beyond single events: Big Data and general disaster resilience

A resilient community can and must do more than face one discrete disaster. It learns from past disasters about its own vulnerabilities. It develops the capacity to detect and monitor emerging hazards and vulnerabilities – some of which may be caused by the recovery process itself – that may build up over time.[93] It then acts to reduce its vulnerability by improving the management of natural systems, strengthening infrastructure, and strengthening social networks. These activities are fundamentally political, requiring that members of the public have three attributes: awareness, capacity for collective action, and ability to weigh in on decisions.

This section describes ways in which Big Data can improve societal awareness of longer-term trends that create vulnerabilities and highlights how Big Data and digital communication platforms create opportunities for citizens to enhance their own agency and thereby build their resilience.

Feedback throughout the disaster cycle

The advances discussed above have obvious applications to disaster preparedness. Big Data can help decision makers to identify critical infrastructure that is at risk, define the optimal positioning of levees and shelters, or design robust evacuation routes. Furthermore, these new sources of data have great potentially to increase *understanding* of resilience and vulnerability. For example, they can help decision makers see whether exposure maps and vulnerability indicators predict the actual impacts on the ground during the crisis. Were assumptions about population movements and evacuation strategies correct?

Case studies that use Big Data to demonstrate the importance of social ties in a disaster[94] show how analysis of social responses to disasters can improve future assessments. Beyond academic studies, information about exposure and vulnerability, followed by data on disaster response and impacts, should feed into the next cycle of risk assessment and preparedness to refine the field's understanding of vulnerability and resilience.

Societal learning about risks

Resilience requires that all actors who make decisions – from households to ministries – be empowered to understand the risks they face and act on them.[95] One of the four pillars of the UN's International Strategy for Disaster Reduction approach to people-centred early-warning systems is understanding risk,[96] because people tend to act on early warnings only if they already understand the risk. Social science research also shows that, in order to act together effectively, communities need a shared understanding of goals, risks, and options.[97] At a minimum, then, building resilience requires all citizens to have access to data. Several global efforts aim to make large amounts of environmental and social data available to the public (see **Box 7** and **Figure 8**).

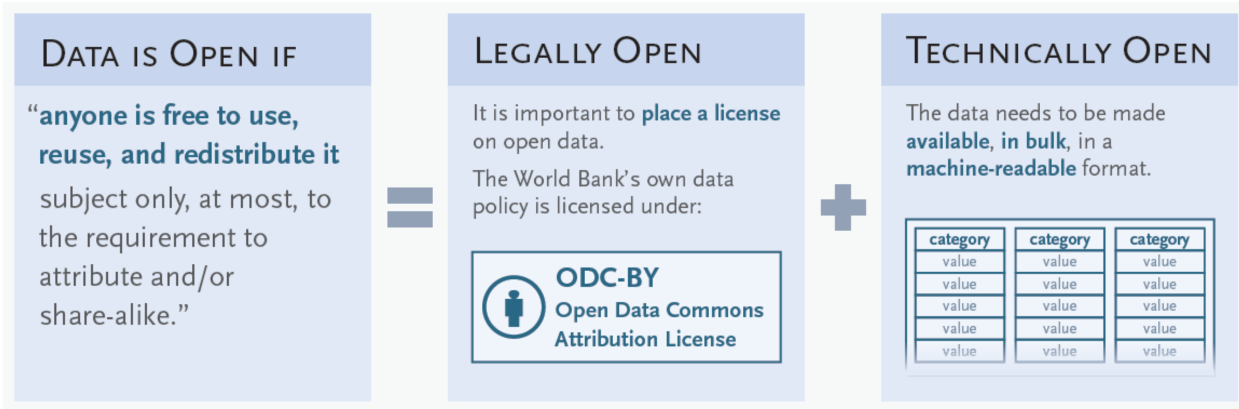
Box 7. Advancements in Open Data

In the last decade, governments have made unprecedented commitments to 'open data' –- in, for example, the G8 Open Data Charter. They are now increasingly focusing on climate and disaster datasets.

- Particularly important is GFDRR's Open Data for Resilience Initiative (OpenDRI)[98], born out of the 2010 earthquake in Haiti. For the first time, private satellite imagery companies released data under an open license to help with disaster response. The satellite data enabled the successful deployment of volunteer mapping efforts. OpenDRI now helps governments streamline open data projects. Using an open-source application called Geonode, OpenDRI initiatives help local actors process their existing data, engage communities in mapping and curating data about their changing exposure to natural hazards, and catalyse a community of practitioners interested in developing risk communication tools and monitoring efforts. Open-source risk communication software called InaSAFE facilitates risk analysis once the data is collated.
- The Group on Earth Observation (GEO) – a partnership of 76 member countries; the European Commission; and 51 intergovernmental, international, and regional organizations – is working toward implementing the vision of the 'digital Earth' advanced by Al Gore in 1998. Currently, only about 1% of ecological data collected is accessible after publication. GEOSS, the GEO System of Systems, is a worldwide effort to connect existing spatial data infrastructures. GEOSS is designed to be highly dynamic, creating a framework for any community, government, or research team to integrate its open data and use the system to model projects. If a very wide range of actors learns how to operate with GEOSS, it can be expected to revolutionize the world of Big Data for earth system management.

The presence of a framework does not ensure its use. Actors must have incentives to spend time curating datasets. Scientists, for example, would be more encouraged to participate if their professional advancement were based on citations for their datasets as well as for their papers.

Figure 8. Definition of 'open data' for the OpenDRI Field Guide



Source: Open Data for Resilience Initiative Field Guide, 2014

They also aim to facilitate the integration of these data for use in risk assessment. However, information alone is not sufficient. Communities must be able to make sense of the data. For that, they need analytical, organizational, and educational support.

Collective action and accountability

There can be no accountability without information. Big Data can strengthen communities' resilience by improving the responsiveness of institutions at all stages of the disaster cycle. As Buchanan-Smith and Davies note,[99] the most essential factors in the success of early warning and early response efforts are political will and institutional capacity. Before the 'data revolution', newspaper circulation was shown to be instrumental in making state governments more responsive to the needs of people affected by floods and droughts in India.[100] In an analysis of Internet penetration, Groshek [101] provides evidence that the Internet can strengthen democracy, while Miard [102] documents cases in which mobile phones seem to have helped citizens' political mobilization. As information technologies spread, this efficacy is likely to increase, at least in relatively democratic countries. However, no quantitative research has yet studied the effect of digital technologies on accountability in the context of disaster management.

Information enables collective action and empowers people to address the risks they face only in so far as they trust the sources of the information. In order to trust, people must know where the information comes from. Deliberate investment in making Big Data analytical tools transparent is therefore likely to pay off in increased disaster preparedness and resilience. The technical complexity of Big Data analytical tools could make the public feel that they represent yet another technocratic approach to development. However, the open and participatory nature of Big Data could increase the trustworthiness of data sources. In particular, Big Data can help people verify information issued by governments, which often manipulate data for political purposes.[103] Little research has been done on how Big Data might affect people's trust in hazard-related information outside of their personal networks.

3. Challenges in mobilizing Big Data for resilience

A number of bottlenecks, gaps, and risks affect the feasibility or desirability of Big Data approaches to disaster resilience. Most are well-documented issues common to Big Data applications in development and humanitarian efforts.[104] This section places these issues in the specific context of disaster resilience, pointing out mitigation strategies where possible. The literature review and case studies identify 5 main classes of issues: Constraints on data access and completeness; Analytical challenges to validity and replicability; Human and technological capacity gaps; Bottlenecks in coordination, communication, and self-organization; and Ethical and political risks, which are discussed in turn.

3.1 Data access and completeness

An obvious impediment to use of Big Data for resilience is access. A large share of Big Data 'crumbs' remains in the hands – or rather, on the servers – of private corporations. The question of opening up CDRs, in particular, has received a lot of attention. During the Ebola crisis, calls to share CDRs with researchers and responders went unanswered[105] (see **Box 8**). The fundamental question of legal ownership will become more urgent in the next few years.[106] Private corporations protect their commercial interests and their reputations by refusing to share CDRs and similar data, despite their potential social value. Though some telecom operators have, like the Orange Group (see Section 1.4), opened up data in a controlled and time-bound manner, no comprehensive guidelines yet exist to facilitate such data sharing and ensure stability and predictability of access. A number of organizations are working on data sharing protocols – notably a joint project by the Leiden University, the World Economic Forum, and New York University's GovLab.[107] Whether, how, and when such efforts come to fruition remains to be seen.

Box 8. Lessons from the Ebola Outbreak

The Ebola outbreak took place in one of the most highly connected and densely populated regions of Africa. Researchers, respondents, and journalists argued that CDRs should be made available to trusted organizations to help with the response. Accurate information on population movements and interactions would have helped in monitoring the progress of the outbreak, predicting its spread, and facilitating interventions. However, despite the efforts of a large number of powerful actors, the governments of affected countries denied all requests. The main reason was that appropriate legal frameworks and institutional processes were not in place.[108] Another factor was risks to the reputations and security of the governments and telecom operators involved. The underlying concern is that, in the absence of clear ethical guidelines, Big Data—at times referred as the "new oil"[109], could lead to extractive or predatory behaviour on behalf of poorly regulated firms.[110]

Accessing some social media data raises challenges too. Social media platforms offer access to part of their data through dedicated APIs (Application Program Interfaces) allowing the automated sharing and standardization of data, but many allow querying only of an archive of past messages. Only a few platforms offer public streaming with a real-time data feed. The most widely used source of social media data is Twitter,[111] which provides a random sample of 1% of all postings that can be filtered by keyword or location. One way of increasing the quantity of data pulled from Twitter in a disaster is to increase the number of terms being searched by, for example, using a disaster lexicon.[112] Another approach is to aggregate feeds from various sites to increase the volume of data being analysed.

A new platform called CrisisNet – an initiative of Ushahidi – provides a single stream of all crisis data from a wide variety of websites and social media platforms, restructuring the data into a single format. The UK-commissioned case study on crowdsourcing[113] highlights limitations of social media data in developing countries, including skewed user bases and accessibility (**Case study 3 above**). Probably as a result of these limitations, only a handful of relief organizations have incorporated social media monitoring into their emergency response efforts.[114]

Use of satellite imagery also comes with challenges. Partial coverage and contamination can impair the accuracy of damage mapping.[115] Furthermore, though satellite data is usually less expensive than ground mapping, notably those provided for free by the United States National Space Agency (NASA)[116], some remote sensing products can be costly. Public-private partnerships that take into account corporations' financial constraints could help improve disaster response. Open access to databanks would enable testing of new methods, but private satellite companies have generally not been willing to provide free or low-cost data even after disasters, much less for testing. Another possible solution to the cost issue is use of new-generation nano-satellites and drones, which could provide cheaper aerial imagery especially useful for hazard monitoring. They can also help establish distributed wireless sensor networks that would lower the costs of monitoring hazards by radically reducing infrastructure, management, and physical connectivity requirements. But these UAVs although also raise ethical and privacy questions as discussed below.

3.2 Analytical challenges to reliability, representativeness and replicabilty

A related and well-established set of challenges pertains to the analysis of and via Big Data – both in terms of internal and external validities.[117] Although strictly speaking, both kinds of validities refer to the ability to make and generalize *causal* claims, i.e. to the realm of *prescriptive* uses of Big Data, we expand their meaning and scope to include predictive and descriptive uses as well, i.e. the extent to which useful insights can be gleaned from these data.

The basic question researchers must always answer is, 'What does the data tell us?' Big Data doesn't magically answer this question; in fact, the answer is usually harder to find than in the case of controlled data collection. Tweets and text messages collected during or after an emergency may be deliberately misleading or false – as may any call for help. More often, the challenge is finding the 'signal' in the 'noise.' Automated processes for dealing with large quantities of unstructured data try to answer the key question through careful extraction, verification, and classification of data.

Some applications call upon volunteers to help classify data; fully automated systems are vastly faster and cheaper, but less flexible. Promising applications for testing the plausibility of claims in real time, based on human and machine computing, include Verily, developed by the Qatar Computing Research Institute.[118] Another example is Artificial Intelligence for Disaster Response (AIDR).[119] It uses volunteers to tag a small subset of the data in order to train a classifier, which then automatically processes the rest of the data. This system enables rapid scaling of data processing and makes it possible to reuse the classifier algorithm in similar future hazards. Though it has been deployed in only a few crises, AIDR has achieved high accuracy in its message classification rate.[120] Since each disaster is unique, systems that combine human and machine computing seem to offer the best results.

Another main challenge to analysis is statistical bias, which comes in several forms. One big problem during disasters is selection bias resulting from attrition. For example, in the aftermath of an earthquake, more phone signals are likely to come from less affected areas than from areas that have been devastated. Assessing need based on the number of phone signals may send aid to the wrong places. Selection bias highlights the need to verify conclusions, most likely by correlating additional data sources, before acting.

Another issue is sampling bias. Despite huge sample sizes, most Big Data sets are not representative. People's decisions to use the technology in the first place are largely determined by characteristics that affect the behaviours under study; Self-selection biases the sample. For example, factors such as age and income affect whether people use mobile phones. Using CDRs to study, say, mobility patterns after a disaster may give a misleading picture because the sample – mobile users – skews younger and richer than the general population[121]—although in some cases it may.[122] These biases will tend to be greater with technologies that have lower penetration rates because there is then more room for highly skewed usage between different social groups.[123] It is unclear, nonetheless, whether adoption patterns and trends of newer and future technologies across groups will mirror those observed in the past. Techniques to correct for sampling bias use standard statistical models and methods to control for mobile or internet penetration rates in, for example, a given area or age group (see **Box 9**).[124] Refining such approaches requires calibrating new data with reliable target data from official or vetted sources. Even then, the ability to generalize the models and their results to other times and places is limited.

Similarly, predictive modelling based on a large number of variables (known as features) may have a very high predictive power in a given setting, but their data requirements and changes in the relationships at play will make them hard to replicate over time and space.

Box 9. Correcting Sampling Bias in CDR Data in Senegal

Based on CDRs of more than 9 million mobile customers in Senegal, Letouzé and colleagues are currently developing a new methodology to correct for sampling bias in CDR datasets. They elaborate a CDR-based estimate of population density and demonstrate how this proxy can be applied to real-time mapping of flood vulnerability.

Correction of sampling bias correction is necessary because factors such as income can affect whether and how much people use mobile phones. Combining the two main approaches to bias correction – a statistical model and machine-learning techniques – Letouzé and colleagues used such variables as mobile penetration rates to partially correct for the sampling bias. As data from reliable sources becomes available to calibrate with their model, they hope to improve the proportion of bias that can be corrected.

An approach to misreporting and bias that has a relatively long history is 'crowdseeding', which combines the strength of crowdsourcing – the ability to quickly generate detailed real-time data – with traditional data gathering that relies on known sources and representative samples. One well-known example is the *Voix des Kivus* project, which uses crowdseeding for real-time monitoring of conflict in the Congo (see **Box 10**). Issues of scalability aside, the project seems to have been a success. However, it generated discontent and mistrust among users[125] because the data was not used to trigger responses to the events reported.

Another challenge to analysis is the sheer complexity of human systems and the inherent difficulty of understanding them through models. One example is the role of social cohesion. Social networks are key determinants of resilience,[126]since safety nets are often provided by neighbours, families, and friends. A community's social capital and leadership constitute one of its most effective means of adapting to change[127] and promoting disaster recovery.[128] However, understanding of the exact processes at play remains limited.[129] Research needs to focus on using 'social sensors,' such as social media and mobile data, to measure social cohesion and then use the results to improve resilience.

Box 10. Example of the 'Crowdseeding' Approach

The *Voix des Kivus* project employs crowdseeding for real-time monitoring of conflict in the Democratic Republic of the Congo.[130] The pilot project selected three representatives in each of 18 villages: the chief of the village, the head of the women's association, and one person elected by the community. These informants were trained to report conflicts, health issues, and natural hazards by text on a weekly basis and were given free airtime to do so. After verifying the data through site visits, the researchers reported that data collection during the 18-month study period was satisfactory, although they found variation in villages' propensity to report. The main challenge was access to electricity, which researchers resolved by distributing $25 solar chargers. Because of the context of conflict, the project included a protocol to protect sensitive information in order to avoid reprisals on reporters. However, such a protocol would not necessarily be robust in a larger-scale project.

3.3 Human and technological capacity gaps

A related well-established set of challenges pertains to local human and technological capacities. By and large, exposure to hazards, income, human capital and technological capacities tend to be correlated, both cross-sectionally and longitudinally. These relationships are not linear and straightforward, but apart from what may appear as exceptions—e.g. the oil-rich desert Gulf countries—places and people that are most exposed to sudden and slow onset natural hazards are typically those least able to leverage the opportunities of Big Data.

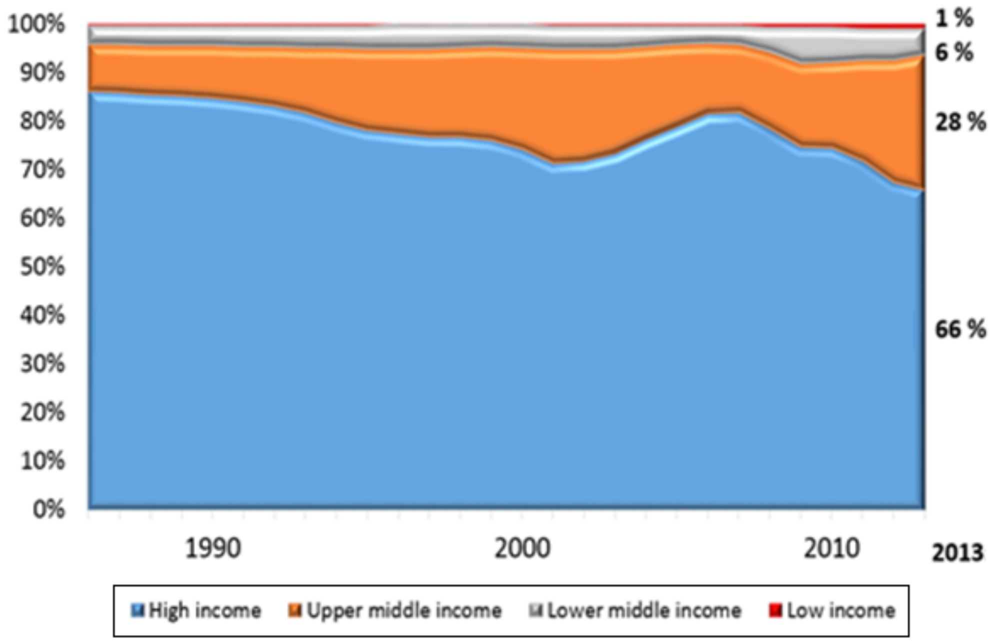
Limited data transfer capabilities (i.e. low bandwidth), a problem often experienced in low-income countries, will evidently hinder the implementation of Big Data applications that rely on cloud computing. More generally few institutions in developing countries can afford the kinds of equipment on which Big Data analytics is performed by top global universities and corporations. Low income countries have only 1% of the world's capacity to transmit data via Internet and phone capacities (see Figure 9). Progress is being made, however, as in the case of Africa, which has enjoyed a 20-fold increase in bandwidth between 2010 and 2015 thanks to submarine cables).[131]

Another limitation is human capacities, perhaps best exemplified by the well-known dearth of skilled staff in statistical offices in developing countries—cause and effect of a brain drain—for which Big Data may seem like a distraction. Several proxies, including standard literacy rates around the world, point to the skills gap. A popular analytics software such as R may be entirely free and open source, but other barriers limit its adoption.

This has obvious short-term effects—Big Data techniques will simply not be part of the potential toolkit of at-risk populations and their institutions. In the face of emergencies where time is of the essence, this simple fact may lead to decisions with detrimental long term effects: the parachuting or even distant engagement of external Big Data experts bypassing local structures.

For Big Data to have a significant and lasting positive impact, investing in technical and above all human capacities will of course be key. In recent months, the notion of 'data literacy' has received increasing attention—a welcome development. But it must be clear that enhancing data literacy, a key requirement for building resilience through Big Data, is not reducible to training world-class computer science PhDs in developing countries.

Figure 9: Global telecommunication capacity by country income group



Source: Hilbert, 2015

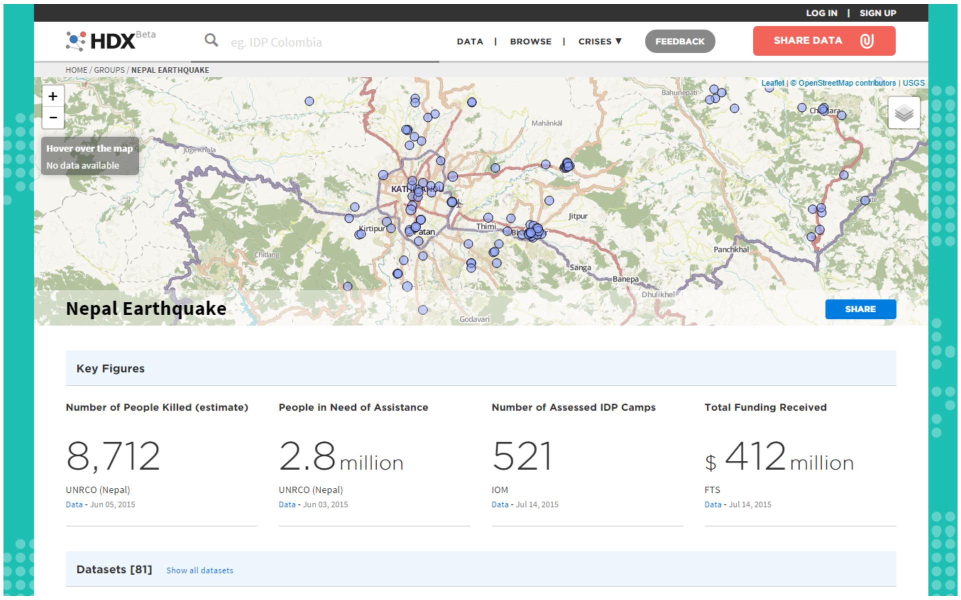
3.3 Bottlenecks in coordination, communication, and self-organization

The 'decision gap' – the disconnect between information and action – is one of the most detrimental features of the disaster cycle. A number of people-centred early-warning systems, defined as systems 'whose warnings are timely and understandable to those at risk (...) including guidance on how to act upon warnings,' have emerged out of the Hyogo Framework.[132]

However, even when hazard detection capacities are good, warnings do not reach all people at risk. Those that do may not be clear and may not address people's concerns[133] and most pressing needs.[134] These issues must be addressed if Big Data is to contribute to bridging the decision gap.

For example, a preliminary impact evaluation compared the work of the digital humanitarian network (DHN) in the 2010 earthquake in Haiti with the more traditional information system deployed by an on-the-ground NGO.[135] It found that, although DHN had quicker deployment, it was actually less responsive to informational needs on the ground. The needs of local users were not always clear to digital volunteers who were processing Big Data online, and the distance between them impeded feedback. Feedback loops rarely remain active when data is being generated by different actors in different circumstances. Cross-communication between agencies and data platforms is a major challenge, although a number of standardized ways of classifying disasters have been developed for this purpose. Investing in open data initiatives is in itself an investment in disaster preparedness; a noteworthy initiative is the Humanitarian Data Exchange (HDX)[136] established by the UN Office for the Coordination of Humanitarian Affairs in 2014 to make humanitarian data easy to find and analyse (see **Figure 10**).

Figure 10. The Humanitarian Data Exchange's page for the Nepal Earthquake



Source: Humanitarian Data Exchange homepage, accessed August 2015.

The real potential for Big Data to help coordinate disaster response so far depends on advanced systems and intensive human resources to staff a hotline and process each message. Such systems are not replicable in developing countries. Improving the flow and coordination of data – and local capacity to use the data – are central to building resilience.

Beyond issues of coordination and capacity, the process by which information leads to action at the individual or group level is highly complex. For example, when considering a complex issue like climate change, individuals tend to rely on their immediate personal experience – today's temperature – rather than the latest report on global atmospheric trends.[137] Enhancing people's capacity to relate their personal experiences of, for example, above-average heat with the effects of climate change has been shown to improve their awareness of and concern about climate change.[138] Whether and how Big Data can help is unclear. However, visualization – a technique that can be facilitated by, for example, a satellite-generated map – is a powerful communication tool that may help people relate their experiences to global trends.

Decades of research in the cognitive and social sciences have shown that, even when they have full information about and understanding of risks, individuals may make decisions that seem irrational to external observers, especially in contexts of high uncertainty and complexity.[139] Patt and Schröter[140] provide the example of a failed flood resettlement programme in Mozambique. Policymakers believed that the farmers' flood losses outweighed the benefits of living in a fertile river bed, but the farmers themselves disagreed.

Many studies point to the critical importance of nurturing active relationships between information providers and the people who use the information.[141] Behavioural and cognitive research shows that people can understand risks much better when quantitative information is presented in interactive simulations attuned to their mental models and preferences.[142] Fusing crowdsourced data with traditional data, as in the crowdseeding method (see **Box 11**), offers an opportunity to build knowledge and exchange *networks* rather than providing information *products*. For example, the Grameen Foundation Application Laboratories built a network of 'community knowledge workers' across Uganda[143] who collect agricultural information through phone surveys and share the results with farmers in their community.

Knowledge networks may be ill-suited to contexts where the effects of poverty strain people's ability and energy to engage in deliberative thinking.[144] Furthermore, they are sustainable only when they are formalized around a small set of people who can propagate the practices over time,[145] so scalability is a challenge. Experimentation, future evaluation, and tailoring to local conditions by local actors are necessary to build appropriate structures. Still, the success of OpenStreetMap and Ushahidi demonstrates that local communities have the capacity to produce knowledge for disaster resilience.

Incorporating knowledge networks into open Big Data mapping initiatives like GEOSS and OpenDRI (see **Box 8**) offers huge potential for resilience building. The frontier in this area is to create open software that not only collects and maps crowdsourced information but also assimilates environmental data and climate forecasts to support participatory risk analyses and generate concrete options. Any projects in this area must create strong relationships among all actors in the knowledge system. Vigorous evaluation of knowledge systems and open data projects would allow for rapid development and learning about what works.

One of this report's case studies systematically reviews the promises and challenges of using mobile phone data during disease outbreaks in data-poor environments (**Case study 11 above**).[146] Advantages included better situational knowledge and increased efficiency, consistency, and ease of communication. However, standardized protocols are needed to facilitate rapid release, collection, and processing of mobile data in an emergency. In addition, the community's existing knowledge of disaster risks and its culturally accepted responses must be taken into account.

The next generation of Big Data crisis management tools should be built around users' needs and goals. Hughes[147] describes participatory methods for the design of such tools in various contexts. In order to devise effective communication strategies, developers must take into account the cognitive barriers to understanding risk.

3.4 Ethical and political risks

Ethical concerns constitute one of the biggest challenges in developing the next generation of Big Data crisis tools. In recent years, for instance, the notion of 'anonymised' data has been torn to pieces; with sufficiently large connected datasets, whether from cell-phones[148] or credit card transactions[149], re-identification is almost always possible. Critically, this development was hardly foreseeable a few years ago; by extension, it is nearly impossible to foresee future technological advances.

Another risk is that some leaders could consider Big Data to be the perfect technical fix for all of the world's problems, looking for a '30,000 feet view' that ignores the critical need for local engagement and investment.[150] Vague discussions about the ethical use of data, often reduced to 'anonymisation', overlook deeper ethical considerations that ought to shape the future of Big Data for resilience, especially in complex and volatile contexts. Arguments over core aspects of data use such as the "right to be forgotten" are still unsettled and ethical norms are evolving quickly, necessitating a cautious and ethically conscious approach.

Building on the principles of the Menlo Report,[151] Pham and Vinck[152] and Letouzé and Vinck[153] outline key requirements for technology-enabled approaches and Big Data analytics.

- Participation must be voluntary and should respect individual autonomy. People must be aware of the risks and benefits of sharing their information based on how it will be used, and they must have the right to withdraw their data;
- Responsible organizations must actively weigh the risks and benefits of their data collection and analysis. They must put a high premium on validation, report potential biases, and prioritize

security over speed;

- The needs of people without technology access are paramount;
- Context must dictate the level of data protection necessary to prevent breach of confidentiality. In repressive environments, for example, protecting information and informants is a priority;

Clearly, in the age of Big Data, meeting these requirements means overhauling current practices and frameworks. However, intermediate steps can be taken to enhance citizens' awareness and persuade governments, NGOs, and corporations to engage on and invest in these issues.

APIs could be part of the answer to some of these ethical concerns as well as to challenges with shared communication and interoperability standards. APIs provide an architecture for the interaction between, on the one side, a data source and, on the other, a programme that uses the data. They can structure the interactions among actors in the informational network – from people's tweets or text messages, through the various organizations that collect the information and use it to inform disaster preparedness or response, and back to the public, who choose how to act on the information. Agreed-upon standards of communication that, for example, protect data privacy, can be hard-coded into these APIs to govern data exchange at every point in the information supply chain. APIs are already widely used to coordinate data flows in the private sector. The next generation of Big Data tools in the humanitarian sector could be based on a shared set of APIs to address issues of data privacy, communication, and coordination.

4. Toward a roadmap

4.1 The state of the field

The menu of options for data-driven resilience is expanding rapidly, and interest in using resilience as a policy framework is on the rise. At the same time, challenges are growing. The evidence base on Big Data for resilience is generally recent, based on incomplete information, and not embedded in well-structured programmes of discovery and evaluation. These circumstances make it difficult to arrive at firm answers to high-priority questions. However, they create opportunities to add dramatically to the knowledge base.

The literature search and commissioned case studies lead to four conclusions about the current state of knowledge about Big Data for resilience.

1. **Big Data has demonstrated its value and is now a permanent feature of the data-for-resilience landscape.**

    1. Big Data has put new metrics into our toolkits. It has shown its ability to make possible the collection of data that otherwise would be completely out of reach.
    2. Big Data has made our metrics more valuable. It has shown its ability to dramatically reduce the time needed to collect data, to increase the spatial resolution of maps, and to target data more precisely at groups and questions of interest.
    3. Big Data has made our data systems more participatory. It has shown its ability to open data collection, interpretation, and use to a broad set of stakeholders.
    4. Big Data has made data systems more responsive. It has shown its ability to dramatically shrink the innovation cycle by leveraging rapidly changing technology that is grounded in a culture of continual reinvention.
    5. Advances in technological infrastructure and capacity seem poised to radically expand the accuracy and usefulness of monitoring systems in coming years.

6. None of these positive effects are universal or automatic. Not all efforts to deploy Big Data on behalf of improved resilience achieve all these benefits. Delivery on the promise requires careful attention to norm- and value-driven design and implementation along with adoption of good management practices. Still, the value proposition has clearly been demonstrated.

2. **The proven benefits of Big Data for resilience do not translate into a clear roadmap showing how to reap these benefits operationally and at scale.**

   1. The literature review did not find a single study or project that addressed how to use Big Data for resilience operationally and at scale. The literature is dominated overwhelmingly by isolated experiments or pilots and highly general paeans.

   2. The literature is a direct reflection of the funding. Funders pay for rapid deployment of new approaches to new disasters and for specific applications in particular cases, but not generally for work aimed at the bigger picture.

   3. The literature and the pilot projects are overwhelmingly dominated by supply-driven or technology-driven questions. Few studies asked, "What information is needed for resilience?" and then looked for effective strategies to meet those needs. There is a clear need to re-imagine approaches to bridging the dispersed elements of the scientific community whose expertise is relevant and the practitioners seeking practical guidance on these questions from a demand-driven perspective.

   4. Studies identify bottlenecks and obstacles, but no one does anything about them. For example, studies frequently note that bandwidth and storage limits in developing countries limit the ability to achieve Big Data's potential, but follow-up projects are far more likely to explore yet another new technology than to look for practical ways to overcome the bandwidth and storage limits. Similarly, efforts to address the high price of commercial satellite data are scattered, ad hoc, and ineffective. Another clear bottleneck concerns IT expertise – efforts to overcome limits are frequently undone by the common practice of newly trained personnel being hired away by commercial firms; there are experiments to work out incentive schemes to retain IT professional in the public sector, but they are not having impacts act scale.

   5. Public-private partnerships proliferate, but there is no progress on viable models to make them effective and transparent or to develop standards on protection and sharing of data.

3. **No organizational structures exist to bring together in a sustained manner the core stakeholders who need to guide the transition from ad hoc pilots to at-scale operations.**

   1. The case studies reveal that many actors have a stake in Big Data for resilience. The most impressive projects have been based on time-consuming engagements of a subset of partners that are not scalable or on specific situations that are not replicable.

   2. A large number of key stakeholders have varying degrees of capability, authority, and incentive to make the transition to at-scale operations: commercial firms, scientists and experts, government agencies and bodies, standards and regulatory bodies, national and international NGOs, community organizations, and donors or investors, among others. To realize the potential of Big Data for resilience, these stakeholders need to come together as a community. Discussions on global partnerships or world forums[154] reflect a growing awareness of the need to bring this latent community to life and give it power.

4. **The current focus on narrow elements of the Big Data ecosystem ignores the strong complementarities that can emerge from a coordinated approach.**

   Exploring the value of a single data technology, or even identifying the 'best' data technology, is not the path to effective use of Big Data for resilience. What will shape effective resilience strategies is combining emerging technologies into a *data portfolio,* as discussed below.

4.2 Specific recommendations

This report's recommendations fall into three broad areas: investments in Big Data technologies, investments in the communities that support and use these technologies, and investments in the future of the field.

1. **Invest in Big Data technologies**

Invest in basic forms of existing and proven applications

The rapidly changing nature of this field means that many promising cutting-edge technologies have not been field-tested and may face difficulty in scale-up. This situation suggests the following recommendations:

- *Share early warning systems and risk maps with stakeholders.* Academic groups and technology start-ups have developed techniques to quickly acquire, process, and analyse Big Data about crises from the Internet and from satellite imagery. The level of development of these technologies ranges from proof-of-concept to well-tested. After being designed for human-centred usability, they should be put into the hands of those who need them for disaster preparedness.
- *Develop tools that use social media to raise awareness and spread information.* Social media-based warning systems and risk maps, for example, have particularly high traction because they do not require local capacity for use, especially for open data sources like Twitter. Furthermore, the user base in developing countries is already expanding, thanks to growing web access and mobile phone adoption.
- *Use call data records to provide insight into population vulnerability.* Processing times currently limit the ability to use CDRs in response to sudden-onset hazards. However, the use of CDRs in Rwanda (see **Box 5**)[155] suggests a way to tap this new source of information: The internal migration patterns identified using CDRs provided a baseline measure that may be useful in future disaster prevention or post-disaster recovery.
- *Facilitate crowdsourcing.* OpenStreetMap and similar efforts demonstrate the potential of combining human and machine computing. Crowdsourcing seems to work particularly well post-disaster; large gains can often be made with few users.

Identify high value-add contexts

Local technological and analytical capacities govern the feasibility of Big Data approaches. However, Big Data methods have great potential in data-poor areas, where they offer a huge improvement in the ability to observe and follow events. The key to progress is investing in areas where such improvement can be achieved most efficiently, as outlined in the following recommendations:

- *Identify the minimum infrastructure required for each technology.* In the case of CDRs, for example, the existence of a phone network is not enough. Also necessary is the ability to either process data on site or quickly transfer it elsewhere.
- *Compare the outputs of traditional and Big Data approaches.* The landslide susceptibility mapping project described in **Box 3** is an example. Using data mining techniques to produce exposure maps in an area where historical data was readily available demonstrates the viability of using this method in data-poor contexts.
- *Invest in high value-add contexts,* such as:
    - Social media, which can provide local reports and insights in data-poor regions

- Middle-income countries, which face the greatest increase in disaster risk and simultaneously are experiencing rapid growth in mobile and social media technology
- Technologies that provide large returns with small numbers of early-adopting data volunteers or super-users

Facilitate proper management and use of existing Big Data resources

The sources and types of Big Data being used to analyse climate change and to foster resilience are expanding rapidly, creating high demand for data sharing among public and private institutions. However, the rights and interests of individuals and groups whose data are being shared must be protected, as recommended here:

- *Develop specific data protection and sharing guidelines.* Legal concerns that prohibit data sharing are a major barrier to development. Pilot projects like the CDR-based population estimates in Senegal described in **Box 9** often access data through ad hoc channels that end when the project does. Common, clear standards and guidelines could encourage companies to share data, knowing that their interests are protected.
- *Establish models and standing public-private partnerships for the rapid release of crisis data.* To make Big Data available during an emergency, legal and political concerns must be dealt with ahead of time, and appropriate infrastructure must already be in place. The effectiveness of such arrangements is demonstrated by the Open Data for Resilience Initiative (see **Box 7**), which made private satellite imagery available to help with response to the 2010 earthquake in Haiti.
- *Facilitate the spread of best practices in gathering and using Big Data.* For example, the UN's International Strategy for Disaster Reduction includes a recommendation for 'people-centred', rather than 'top-down' early warning systems.[156] As the 2014 Ebola outbreak demonstrated (see **Box 8**), governments cannot be relied on organise data collection or data sharing in a crisis.

Shift to a 'data portfolio' approach rather than individual data project approach

Newer technologies, like the CDRs and social media streams that dominate current research and applications, are more powerful when they are combined with traditional approaches, as in the case of crowdseeding (see **Box 10**). Selectively bringing in emerging devices such as advanced sensors and drones is likely to add even more value. In the absence of research and pilots that uncover the knowledge needed, the field should concentrate on establishing and testing portfolio approaches. At a minimum, an effective data portfolio would incorporate sensor networks, use both historical data and information from current or ongoing events, build operational linkages to traditional administrative data processes, and operationalize high-value recent technology such as CDRs. A portfolio approach would expand the existing focus on sudden-onset disasters to include slow-onset disasters and long-term trends in risk and vulnerability. The following recommendations can guide development of a portfolio approach to Big Data:

- *Use all existing sources of data across platforms.* In a crisis, disaster relief and response efforts should tap additional easily available data sources.
- *Make use of complementarities* between data with different strengths and weaknesses.
- *Establish and test sensor networks.* Technologies that generate environmental, infrastructural, and behavioural data—such as the motion detectors in mobile phones, home environmental monitoring systems like Nest, and large-scale sensor networks like the ones being created in India's 100 Smart Cities project—offer great promise as part of a larger data portfolio.

- *Develop efficient ways to correct for bias.* Patterns of adoption and use of such data sources as social media and mobile phones often produce selection or sampling bias. The case study on CDR-based population estimates in Senegal described in **Box 9** shows that such bias can be corrected.

2. **Invest in Big Data communities**

Facilitate coordination and communication among stakeholders

In a crisis, information is valuable only to the extent that it addresses the needs of affected populations, first responders, and relief organizations. To build long-term resilience, many more actors must be included in participatory design processes that democratize access to Big Data and make the information useful for societal learning. The following recommendations aim at facilitating coordination and communication:

- *Create new avenues and means of cooperation.* Regional hubs could help to connect communities and planners across jurisdictions.[157] Two UN Global Pulse labs have been established in Jakarta and Kampala and other organizations including Data-Pop Alliance are developing multi-partner and interdisciplinary Data Spaces in selected major cities of the Global South starting in Bogotá and Dakar.

- *Facilitate communication and exchange between affected communities.* Particularly when they are affected by a disaster, communities need ways to receive services meant to help them, provide feedback on the effectiveness of those services, and identify resources suited to their circumstances.[158]

- *Promote coordination of common standards.* Standards governing data format, documentation, and access are weakly enforced, largely due to low levels of capacity and mismatched incentives. Standards governing legal use of data would help avoid unnecessary obstacles to access and integration.

Promote and incentivize private sector involvement

The private sector could do much more to help leverage the power of Big Data for resilience. Private sector organizations should be encouraged to:

- *Connect to NGOs and international organizations.* Private companies would have to invest time and energy in identifying the needs of non-profits, understanding their capabilities and constraints, and earning their trust.

- *Develop ways to target donations where they are needed.* Several organizations have active initiatives in this area that could provide expertise. For example, GiveDirectly has a widely acclaimed and empirically tested cash transfer programme for mobile money. Post-disaster Big Data recovery efforts could use similar platforms.

- *Identify and support promising technology start-ups.* Investments in Big Data for resilience can be a win-win situation.

- *Support or organize data challenges for development.* The success of the Orange Group's challenges in providing data for resilience efforts in developing countries show that such initiatives can help unite previously disconnected groups.

Engage with public officials and civil society representatives to address privacy and other political and legal issues

While the value of using personal data for resilience has been demonstrated, it is not currently possible to scale up any meaningful applications because of a range of obstacles that include measures to protect privacy, to limit security threats, to communicate risks meaningfully to the public and to data holders, to provide means of redress in the event of breaches, to provide procedures over oversight and accountability to enable regulators to exercise due diligence in protecting the public welfare, for spreading proven technologies and mechanisms for protecting privacy, among others. For most of these things, knowledge of what to do is relatively robust, but there is much work to be done to put everything to use in a comprehensive policy framework that enables the power of private data to be unleashed in a manner that will not trigger a public backlash over concerns about privacy and security. Succeeding at this challenge requires bringing together expertise in data science, ethics, law, risk management, communications, and public policy.[159]

To date, there has been significant progress within these separate communities elaborating elements of solutions at the blueprint level. And there has been modest progress at working out higher-level architectures for how to combine such elements effectively. But there has been very little deep engagement within national governments to roll out reforms based on these exercises.

Spurring data literacy.

Big Data for resilience should not be left to experts only; a major requirement is to enhance people's willingness and ability to engage with and via Big Data to shape the future of the field. The urgent need for focusing on and investing in data literacy of various social agents and groups is now undisputed. The report[160] of the UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG)[161] published in the fall of 2014 mentioned *"data literacy"* 4 times and put forth *"(a) proposal for a special investment to increase global data literacy"*, advocating for the development of *"an education program and promote new learning approaches to improve people's, infomediaries' and public servants' data literacy"* adding that *"(s)pecial efforts should be made to reach people living in poverty through dedicated programmes"*. Yet, as of now, the proposal has not been picked up.

In recent months however, a coalition of stakeholders led by Data-Pop Alliance and PARIS21 have been developing a global education program to spur broad-based data literacy by tackling the methodological, technical, political, ethical dimensions at various levels of societies.

Invest in the future of the field

Facilitate knowledge sharing within the disaster response community and cycle

Knowledge and expertise are scarce in the rapidly evolving field of Big Data for resilience. New understandings often arrive unevenly as individual communities use specific technologies to respond to disasters. This situation reveals a crucial need to facilitate feedback between, on the one hand, Big Data innovators and researchers and, on the other, stakeholders who prepare for and respond to disasters. Big Data researchers could take the following steps:

- *Identify and spread best practices.* New findings could be spread through such familiar technologies as websites and email lists.
- *Conduct event 'post-mortem' analysis to evaluate specific approaches.* For example, the Flowminder case study highlighted in **Box 5** shows how Big Data can help to evaluate the effectiveness of disaster response and identify areas for improvement.
- *Identify practical stumbling blocks* before *disaster strikes.* Specific technologies and specific contexts each have their own specific obstacles to implementation, from limited processing power and slow internet connections to weak or non-existent legal frameworks.
- *Reduce the gap between information product suppliers and users.* As mentioned above, in the 2010 earthquake in Haiti, the digital humanitarian volunteers who were processing Big Data online were not always aware of actual needs on the ground.[162] Effective feedback loops would enable quicker and more effective response.

Tap mobile phone data more fully and rapidly

Mobile phones are the only kind of technology available to many people in developing countries. However, adoption of mobile technology is ad hoc, scattered, and slow. Investments in data infrastructure and fit-for-purpose handheld apps are needed.

Synchronize Big Data sources

Knowing where people are located is central to any disaster assessment. Off-the-shelf population distribution data, whether simple headcounts or disaggregated breakdowns, are seldom fit for the purpose. Disaster analysts therefore look to CDRs and social media data to improve demographic mapping. However, they often not able to obtain the data for use in resilience work because of privacy and legal concerns. If they are able to use the data, they seldom have the chance to combine the CDR and social media data with data from other sources. Significant investment is needed to build systems that integrate CDR and social media data to reduce the transaction cost problem for the end user.

References

Aldrich, Daniel P., and Michelle A. Meyer. "Social capital and community resilience." *American Behavioral Scientist* (2014): 0002764214550299.

Alexander, David. 2005. Symbolic and practical interpretations of the Hurricane Katrina disaster in New Orleans. New York: Social Science Research Council.http://understandingkatrina.ssrc.org/http://understandingkatrina.ssrc.org/

Anttila-Hughes, J. K., & Hsiang, S. M. (2013). Destruction, disinvestment, and death: Economic and human losses following environmental disaster. *Available at SSRN 2220501*

Basher, Reid. "Global early warning systems for natural hazards: systematic and people-centred." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 364, no. 1845 (2006): 2167-2182.

Bierkandt, R., Wenz, L., Willner, S.N., Levermann, A. 2014. Acclimate—a model for economic damage propagation. Environ Syst Decis (2014) 34:507–524.

Blumenstock, J.E. 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda, Information Technology for Development, 18:2, 107-125

Blumenstock, J.E. 2014. Calling for better measurement: Estimating an individual's wealth and well-being from mobile phone transaction records. Proceedings of Knowledge Discovery in Data.

Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—harnessing the Web for public health surveillance. *New England Journal of Medicine*, *360*(21), 2153-2157.

C. Smith, A. Mashhadi and L. Capra (2014) Poverty on the Cheap: Estimating Poverty Maps Using Aggregated Mobile Communication Networks. ACM.

Carpena, F., Cole, S., Shapiro, J., & Zia, B. (2011). *The ABCs of Financial Literacy–Experimental Evidence on Attitudes, Behavior and Cognitive Biases*. mimeo, World Bank.

Cavallo, A. (2013). Online and official price indexes: measuring Argentina's inflation. *Journal of Monetary Economics*, *60*(2), 152-165.

Cavallo, A., Cavallo, E., and Rigobon, R. 2013. Prices and Supply Disruptions during Natural Disasters. NBER Working Paper No. 19474 September 2013 JEL No. E20,E30,O57,Q54

Cervone, Guido, and Germana Manca. "Damage Assessment of the 2011 Japanese Tsunami Using High-Resolution Satellite Data." *Cartographica: The International Journal for Geographic Information and Geovisualization* 46, no. 3 (2011): 200-203.

Chan, E. H., Brewer, T. F., Madoff, L. C., Pollack, M. P., Sonricker, A. L., Keller, M., ... & Brownstein, J. S. (2010). Global capacity for emerging infectious disease detection. *Proceedings of the National Academy of Sciences*, *107*(50), 21701-21706.

Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010a). Framework for analytical quantification of disaster resilience. *Engineering Structures*, *32*(11), 3639-3649.

Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010b). Seismic resilience of a hospital system. *Structure and Infrastructure Engineering*, *6*(1-2), 127-144.

Clayton, S. et al., 2015. Psychological research and global climate change. *Nature Climate Change*, 5(7), pp.640–646. Available at: http://www.nature.com/doifinder/10.1038/nclimate2622.

Cochran, Elizabeth S., Jesse F. Lawrence, Carl Christensen, and Ravi S. Jakka. "The quake-catcher network: Citizen science expanding seismic horizons." *Seismological Research Letters* 80, no. 1 (2009): 26-30.

Costello, M. J. (2009). Motivating online publication of data. *BioScience*, *59*(5), 418-427.

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., ... & Woodgate, P. (2012). Digital Earth 2020: towards the vision for the next decade. *International Journal of Digital Earth*, *5*(1), 4-21.

Dashti, S. et al. 2014. Proceedings of the 11th International ISCRAM Conference –University Park, Pennsylvania, USA, May 2014 S.R. Hiltz, M.S. Pfaff, L. Plotnick, and P.C. Shih, eds.

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, S., & Anderson, S. (2014). Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 colorado floods. *Proc. of ISCRAM*.

Dell, M., Jones, B. F., & Olken, B. A. (2014). What Do We Learn from the Weather? The New Climate–Economy Literature. Journal of Economic Literature, 52(3), 740-798.

Eagle, N., Macy, M., Claxton, R. 2010. Network Diversity and Economic Development. Science DOI: 10.1126/science.1186605, 1029; 328.

Eagle, N., Pentland, A.S., and Lazer, D. 2009. Inferring friendship network structure by using mobile phone data. PNAS, vol. 106, no. 36, pp. 15274–15278.

Fekete, Alexander. "Validation of a social vulnerability index in context to river-floods in Germany." *Natural Hazards and Earth System Science* 9.2 (2009): 393-403.

Feng, Maria, Yoshio Fukuda, Masato Mizuta, and Ekin Ozer. "Citizen Sensors for SHM: Use of Accelerometer Data from Smartphones." *Sensors* 15, no. 2 (2015): 2980-2998.

Frazier, A. E., Renschler, C. S., & Miles, S. B. (2013). Evaluating post-disaster ecosystem resilience using MODIS GPP data. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 43-52.

Frias-Martinez, Vanessa, Virseda, J., Rubio, A., and Frias-Martinez, E. (2010). Towards Large Scale Technology Impact Analyses: Automatic Residential Localization from Mobile Phone-Call Data.

Frias-Martinez, Vanessa, Virsesa, J. (2012) On the relationship between socioeconomic factors and cell phone usage, International Conference on Information and Communication Technologies and Development.

Generous, Nicholas, Geoffrey Fairchild, Alina Deshpande, Sara Y. Del Valle, and Reid Priedhorsky. "Global disease monitoring and forecasting with Wikipedia." (2014): e1003892.

Giada, S., De Groeve, T., Ehrlich, D., & Soille, P. (2003). Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *International Journal of Remote Sensing*, *24*(22), 4251-4266.

Gillespie, T.W., Chu, J., Frankenberg, E. and Thomas, D. 2007: Assessment and prediction of natural hazards from satellite imagery. Progress in Physical Geography 31, 459–70.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L. 2009. Understanding individual human mobility patterns.

Haklay, Mordechai. "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets." *Environment and planning. B, Planning & design* 37.4 (2010): 682.

Henriet, D., & Michel-Kerjan, E. O. (2006). *Optimal Risk-Sharing under Dual Asymmetry of both Information and Market Power: A Unifying Approach*. Working Paper.

Hornik, R. 2002. Public health communication: Evidence for behavior change. Lawrence Erlbaum Associates, Mahwah, NJ.

Howe, P.D. & Leiserowitz, A., 2013. Who remembers a hot summer or a cold winter? The asymmetric effect of beliefs about global warming on perceptions of local climate conditions in the

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. "Quantifying the influence of climate on human conflict." Science 341, no. 6151 (2013): 1235367.

Hsieh, C. H., Reiss, C. S., Hunter, J. R., Beddington, J. R., May, R. M., & Sugihara, G. (2006). Fishing elevates variability in the abundance of exploited species. *Nature*, *443*(7113), 859-862.

Humanitarian Openstreetmap Team (UGMHOT), 2012. Evaluation of OpenstreetMap Data in Indonesia-A Final Report. Departmentof Geodetic & Geomatics Engineering, Faculty of Engineering.

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2014a). Processing social media messages in mass emergency: A survey. *arXiv preprint arXiv:1407.7071*.

Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014b). Aidr: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (pp. 159-162). International World Wide Web Conferences Steering Committee.

IPCC SRX - 2012 - Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation

Joyce, K. E., S. V. Samsonov, Shaun R. Levick, J. Engelbrecht, and S. Belliss. "Mapping and monitoring geological hazards using optical, LiDAR, and synthetic aperture RADAR image data." Natural hazards 73, no. 2 (2014): 137-163.

Joyce, Karen E., Kim C. Wright, Sergey V. Samsonov, and Vincent G. Ambrosia. Remote sensing and the disaster management cycle. INTECH Open Access Publisher, 2009b.

Joyce, Karen E., Stella E. Belliss, Sergey V. Samsonov, Stephen J. McNeill, and Phil J. Glassey. "A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters." *Progress in Physical Geography* (2009a).

Karamperidou, C., Cane, M. A., Lall, U., & Wittenberg, A. T. (2014). Intrinsic modulation of ENSO predictability viewed through a local Lyapunov lens. *Climate dynamics*, *42*(1-2), 253-270.

Kates, R. W., Colten, C. E., Laska, S., & Leatherman, S. P. (2006). Reconstruction of New Orleans after Hurricane Katrina: a research perspective. *Proceedings of the National Academy of Sciences*, *103*(40), 14653-14660.

Kéfi, S., Guttal, V., Brock, W. A., Carpenter, S. R., Ellison, A. M., Livina, V. N., ... & Dakos, V. (2014). Early warning signals of ecological transitions: methods for spatial patterns. *PloS one*, *9*(3), e92097.

Kerle, Norman. "Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it?" International Journal of Applied Earth Observation and Geoinformation 12, no. 6 (2010): 466-476.

Kotler, P. and N. Lee. 2008. Social marketing: Influencing behaviors for good. Sage, Los Angeles, CA.

Kousky, Carolyn. "Informing climate adaptation: A review of the economic costs of natural disasters." Energy Economics 46 (2014): 576-592.

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., ... & Thomas, C. J. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability science*, *7*(1), 25-43.

Letouzé, E. "Big Data for Development: Challenges and Opportunities", (2012), UN Global Pulse, http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

Letouzé, E. , P. Meier and P. Vinck. 2013. "Big data for conflict prevention: New oil and old fires", in F. Mancini (ed.), New Technology and the Prevention of Violence and Conflict, International Peace Institute, New York, http://www.ipinst.org/wp-content/uploads/publications/ipi_epub_new_technology_final.pdf; for an overview & comparison of both taxonomies

Letouzé, E. 2014. "Big data for development: Facts and figures", SciDev, www.scidev.net/global/data/feature/big-data-for-development-facts-and-figures.html

Levermann, A. 2014. Make supply chains climate-smart. [SS2]

Lynam, T., De Jong, W., Sheil, D., Kusumanto, T., & Evans, K. (2007). A review of tools for incorporating community knowledge, preferences, and values into decision making in natural resources management. *Ecology and society*, *12*(1), 5.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *science*, *341*(6149), 976-980.

Marlon, J.R., Rosenthal, S., Feinberg, G., Pal, S. and Leiserowitz, A. (2015). Hurricane Attitudes of Coastal Connecticut Residents: A Segmentation Analysis. June 29, 2015. Yale University. New Haven, CT: Yale Project on Climate Change Communication.

Matin, Nilufar, and Richard Taylor. "Emergence of human resilience in coastal ecosystems under environmental change." *Ecology and Society* 20, no. 2 (2015): 43.

Meier, P. 2014a. "Crowd Computing Satellite & Aerial Imagery." In: Meier, P. Digital Humanitarians.

Meier, P. 2014b. "Artificial Intelligence in the Sky." In: Meier, P. Digital Humanitarians.

Miles, S. B., & Chang, S. E. (2011). ResilUS: a community based disaster resilience model. *Cartography and Geographic Information Science*, *38*(1), 36-51.

Missing Maps, 2015. Missing Maps project website. Available at: http://www.missingmaps.org/

Musaev, Aibek, De Wang, and Calton Pu. "LITMUS: Landslide detection by integrating multiple sources." In *The 11th International Conference on Information Systems for Crisis Response and Management*. 2014.

Nakagawa, Y., and Shaw, R. 2004. Social capital: A missing link to disaster recovery. International Journal of Mass Emergencies and Disasters, 22(1), 5-34.

Nativi, S., Mazzetti, P., Saarenmaa, H., Kerr, J., & Tuama, É. Ó. (2009). Biodiversity and climate change use scenarios framework for the GEOSS interoperability pilot process. *Ecological Informatics*, *4*(1), 23-33.

Nex, F., E. Rupnik, I. Toschi, and F. Remondino. "Automated processing of high resolution airborne images for earthquake damage assessment." *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1 (2014): 315-321.

NOAA/NWS. 2013. Hurricane/Post-Tropical Cyclone Sandy, October 22-29, 2012, Service Assessment. Silver Spring, Maryland

OCHA, (2013). Humanitarianism in the network age.

ODI - 2013 - The geography of poverty, disasters and climate extremes in 2030

Ouzounis, G. K., Soille, P., & Pesaresi, M. (2011). Rubble detection from VHR aerial imagery data using differential morphological profiles. In *Proc. 34th Int'l Symp. Remote Sensing of the Environment*.

P. Deville, C. Linard, S. Martin, M. Gilbert, F.R. Stevens, A.E. Gaughan, V.D. Blondel and A.J. Tatem, 2014. Dynamic population mapping using mobile phone data. PNAS, vol. 111, no. 45: 15888–15893.

Patt, A. G., & Schröter, D. (2008). Perceptions of climate risk in Mozambique: implications for the success of adaptation strategies. *Global Environmental Change*, *18*(3), 458-467.

Pham, P. N., & Vinck, P. (2012). Technology fusion and their implications for conflict early warning systems, public health, and human rights. *health and human rights*, *14*(2), 106-117.

Pham, Thi-Thanh-Hiên, Philippe Apparicio, Christopher Gomez, Christiane Weber, and Dominique Mathon. "Towards a rapid automatic detection of building damage using remote sensing for disaster management: The 2010 Haiti earthquake." *Disaster prevention and management* 23, no. 1 (2014): 53-66.

Polli, Diego Aldo, Fabio Dell'Acqua, and Laura Candela. "Mapping Earthquake Damage from Post-Event only VHR SAR Texture Maps: Zooming into Poor Estimation Cases."

Preston, B. L., Yuen, E. J., & Westaway, R. M. (2011). Putting vulnerability to climate change on the map: a review of approaches, benefits, and risks. *Sustainability Science*, *6*(2), 177-202.

Qiu, Shi, Changyong Cao, Bin Zhang, Xi Shao, Wang Xie, Yan Bai, and Chuanrong Li. "Feasibility study of remote sensing using structured light for 3D damage assessments after natural disasters." In *SPIE Asia Pacific Remote Sensing*, pp. 92632R-92632R. International Society for Optics and Photonics, 2014.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, *331*(6018).

Rossi, Esteban, John Rogan, and Laura Schneider. "Mapping forest damage in northern Nicaragua after Hurricane Felix (2007) using MODIS enhanced vegetation index data." *GIScience & Remote Sensing* 50, no. 4 (2013): 385-399.

Sala, S., and Dendena, B. 2014. "GIS in the Global South". In: Robin Mansell, and Peng Hwa Ang (eds.), The International Encyclopedia of Digital Communication and Society, First Edition. Wiley.

Seybolt, T. B., Aronson, J.D., and Fischhoff, B. 2013. Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict.

Shahbazi, Mozhdeh, Jérôme Théau, and Patrick Ménard. "Recent applications of unmanned aerial imagery in natural resource management." *GIScience & Remote Sensing* 51, no. 4 (2014): 339-365.

Shih, F., Seneviratne, O., Liccardi, I., Patton, E., Meier, P., & Castillo, C. (2013). Democratizing mobile app development for disaster management. In *Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities* (pp. 39-42). ACM.

Smith, C., Mashhadi, A., and Capra, L. 2013. Ubiquitous Sensing for Mapping Poverty in Developing Countries. Orange Data for Development Challenge, NetMob 2013. 1-3 May 2013, Cambridge (MA), United States.

Smyrl, L., T. Kern, and J. Allen. 2011. USGS Twitter Earthquake Dispatch (@USGSted). Fort Collins, CO: USGS Fort Collins Science Center. 1 p.

Sosnik, D., M. Dowd, and R. Fournier. 2006. Applebee's America: How successful political, business, and religious leaders connect with the new American community. Simon & Schuster, New York, NY.

Statistics New Zealand (2012). Using cellphone data to measure population movements. Wellington: Statistics New Zealand

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, *108*(4), 811-826.

Stow, Douglas A., Christopher D. Lippitt, Lloyd L. Coulter, and Bruce A. Davis. "Time-Sensitive Remote Sensing Systems for Post-Hazard Damage Assessment." In *Time-Sensitive Remote Sensing*, pp. 13-28. Springer New York, 2015.

Tralli, D.M., Blom, R.G., Zlotnicki, V., Donnellan,A. and Evans, D.L.2005: Satellite remote sensing of earthquake, volcano, fl ood, landslide and coastal inundation hazards. ISPRS Journal of Photogrammetry and Remote Sensing 59, 185–98.

UN Global Pulse, 'Feasibility Study: Supporting Forest and Peat Fire Management Using Social Media', Global Pulse Project Series, no.10, 2014a.

UN Global Pulse, 'Using Mobile Phone Activity For Disaster Management During Floods', Global Pulse Project Series no.2, 2014b.

UNISDR - 2015 - Global Assessment Report on Disaster Risk Reduction

Vaidya, Ranjan. "Trust formation in information systems implementation in developing countries: The role of emancipatory expectations." ACIS, 2014.

Van der Linden, S., 2015. The social-psychological determinants of climate change risk perceptions: Towards a comprehensive model. *Journal of Environmental Psychology*, 41, pp.112–124. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0272494414001170.

Yamazaki, Fumio, Daiki Hanada, and Kentaro Suzuki. "Use of airborne optical and thermal imagery for the detection of building damage due to the 2012 Tsukuba tornado."

Endnotes

[1] Letouzé, 2014
[2] For further information on the challenges in ICT4D see: Kleine, Dorothea, and Tim Unwin. "Technological Revolution, Evolution and New Dependencies: what's new about ict4d?." *Third World Quarterly* 30, no. 5 (2009): 1045-1067.
[3] For further information on the challenges in PGIS see: Chambers, Robert. "Participatory mapping and geographic information systems: whose map? Who is empowered and who disempowered? Who gains and who loses?." *The Electronic Journal of Information Systems in Developing Countries* 25 (2006).
[4] ODI 2013, UNISDR 2015
[5] UNISDR 2015
[6] Udry, various years, Kousky 2013, Hsiang, Burke, and Miguel 2013, Dell, Jones and Olken 2014
[7] IPCC SRX 2012
[8] UNISDR 2012
[9] ODI 2015
[10] https://sustainabledevelopment.un.org/content/documents/7891TRANSFORMING%20OUR%20WORLD.pdf
[11] http://www.unisdr.org/archive/45308;
[12] Seeliger & Turok 2013. http://www.mdpi.com/2071-1050/5/5/2108/htm#B7-sustainability-05-02108 and Turner, B.L. Vulnerability and resilience: Coalescing or paralleling approaches for sustainability science? Global Environ. Change 2007, 20, 570–576. [Google Scholar] [CrossRef]
[13] Cutter, S. L. "Vulnerability to Environmental Hazards." *Progress in Human Geography* 20 (1996): 529–539.
[14] See Anderies et al 2013, Turner et al 2010, as well as Seeliger & Turok 2013.
[15] Letouzé, 2014 and 2015
[16] Pentland, 2012
[17] Letouzé, Vinck and Meier, 2013
[18] As noted in Letouzé et al, 2013 and pointed out by Lea Shanley then, the "active vs. passive dichotomy (…) has been discussed a great deal in the geography literature. The distinction made is whether people are actively providing information about themselves for a specific purpose, or are at least aware of the data collection and don't object, or whether they are unaware of and may object to being observed/tracked and have their information collected and used for purposes other than they originally intended".
[19] For a filler discussion see Letouzé, 2012 and 2015 and Global Pulse, 2013
[20] Carpenter et al. 2012
[21] http://www.unglobalpulse.org/projects
[22] See http://europeandcis.undp.org/blog/2013/03/21/data-dive-measuring-poverty-through-real-time-data/
[23] http://www.cn.undp.org/content/china/en/home/presscenter/pressreleases/2014/08/harnessing-the-power-of-big-data.html
[24] http://europeandcis.undp.org/blog/2014/04/09/proxy-data-and-disaster-resilience/
[25] http://www.am.undp.org/content/armenia/en/home/presscenter/articles/2015/04/06/cracking-coconuts-big-data-gaming-maps-and-portals-.html
[26] http://europeandcis.undp.org/blog/2014/08/29/undp4future-what-were-doing-today/
[27] https://data.hdx.rwlabs.org/
[28] http://www.d4d.orange.com/en/Accueil
[29] Chitturi et al., 2013
[30] Brouwer and Provost, 2015
[31] For further information, see the Proceedings of 2013 edition and 2015 edition, respectively available at: http://perso.uclouvain.be/vincent.blondel/netmob/2013/, http://www.netmob.org/, van den Elzen et al., 2013 and 2015; Trestian et al., 2015
[32] Flooding through the lens of mobile phone activity." Pastor- Escuredo, D., Morales-Guzmán, A. et al, IEEE Global Humanitarian Technology Conference, GHTC 2014. Flooding through the lens of mobile phone activity." Pastor-Escuredo, D., Morales-Guzmán, A. et al, IEEE Global Humanitarian Technology Conference, GHTC 2014.
[33] For further information on the challenges in ICT4D see: Kleine, Dorothea, and Tim Unwin. "Technological Revolution, Evolution and New Dependencies: what's new about ict4d?." *Third World Quarterly* 30, no. 5 (2009): 1045-1067.
[34] For further information on the challenges in PGIS see: Chambers, Robert. "Participatory mapping and geographic information systems: whose map? Who is empowered and who disempowered? Who gains and who loses?." *The Electronic Journal of Information Systems in Developing Countries* 25 (2006).
[35] Toyama, Kentaro. "Technology as amplifier in international development." In Proceedings of the 2011 iConference,

pp. 75-82. ACM, 2011.
[36] Heeks, Richard. Most eGovernment-for-development projects fail: how can risks be reduced?. Manchester: Institute for Development Policy and Management, University of Manchester, 2003.
[37] Pesaresi, 2014: NEW REF: Pesaresi, M. (2014). Global Fine-Scale Information Layers: the Need of a Paradigm Shift. Proceedings of the 2014 conference on Big Data from Space (BiDS'14) p. 8-11
[38] European Space Agency, 2010. NEW REF: European Space Agency (2010) GlobCover 2009 Product Description Manual.
[39] USGS, 2013. NEW REF: United States Geological Survey (USGS) (2013) Land Cover Type.
[40] Klotz et al. 2014 NEW REF: KLOTZ, Martin, et al. "Mapping Global Exposure from Space: A Review of Existing Products and Comparison of Two New Layers of Global Urban Extent.", *Second European Conference on Earthquake Engineering and Seismology.*
[41] Klotz et al. 2014.
[42] Pesaresi et al. 2013. NEW REF: Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., ... & Zanchetta, L. (2013). A global human settlement layer from optical HR/VHR RS data: concept and first results. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*(5), 2102-2131.
[43] Palen et al. 2014
[44] Haklay 2010
[45] Missing Map, 2015
[46] http://hotosm.org/projects/tanzania
[47] http://hotosm.org/projects/indonesia-0
[48] Palen et al. 2014
[49] Deville et al. 2014
[50] Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & Von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS medicine*, *8*(8), 1128.
[51] ITU
[52] Deville et al., 2014
[53] Wesolowski et al. (2012)
[54] Eagle, 2010
[55] Smith-Clarke et al., 2012
[56] Blumenstock, 2014
[57] Frias-Martinez et. al., 2012
[58] Illife et al., 2015
[59] Aldrich et al. 2014
[60] e.g. Eagle et al. 2009, Onnela et al. 2007
[61] UNISDR, 2006
[62] Smyrl et al. 2011
[63] http://bpp.mit.edu
[64] Stow et al. (2015)
[65] Qiu et al., 2014
[66] Rossi et al., 2009
[67] Cervone et al., 2013
[68] Pham et al., 2015
[69] Yamazaki et al., 2013
[70] Polli et al., 2013; Nex et al., 2014
[71] Meier, 2014
[72] Asmolov, 2014
[73] Sarcevic et al.
[74] Roth & Luczak-Roesch
[75] Bengtsson et al. 2011
[76] Blumenstock, 2012
[77] Gonzalez et al., 2008
[78] UN Global Pulse, 2014b
[79] UN Global Pulse, 2014b
[80] Statistics New Zealand (2012)
[81] For a fuller discussion, see http://www.sciencemag.org/content/343/6176/1203.full
[82] Dashti et al. 2014
[83] UN Global Pulse, 2014a
[84] http://www.hubbardresearch.com/cigar-global-intervention-decision-model/
[85] Ibid.
[86] Scheffer et al. 2009

[87] Scheffer et al. 2009
[88] Anderies et al. 2006
[89] Scheffer et al. 2009; Dakos et al. 2015
[90] Hsieh et al. 2006
[91] Karamperidou et al. 2013
[92] See www.early-warning-signals.org
[93] Kates et al. 2006
[94] Blumenstock et al. 2014; Bengtsson et al. 2011
[95] Carpenter et al. 2012; Tompkins and Adger, 2004
[96] ISDR, 2005
[97] Dietz, 2003; Ostrom, 2005
[98] World Bank. 2014. *Open Data for Resilience Field Guide*. Washington, DC: World Bank.
[99] Buchanan-Smith and Davies (1995)
[100] Besley and Burgess, 2001
[101] Groshek (2009)
[102] Miard (2008)
[103] See for example: Cavallo, 2013
[104] boyd and Crawford, 2010, Letouzé, 2012, 2015 , Letouzé, Vinck and Meier, 2013, Pentland, 2014
[105] "Ebola and big data: Waiting on hold," *The Economist (print edition)*, 25 October 2014, available at: http://www.economist.com/news/science-and-technology/21627557-mobile-phone-records-would-help-combat-ebola-epidemic-getting-look.
[106] For a review and discussion, see Letouzé, Vinck and Kammourieh, 2015
[107] http://www.responsible-data.org/data-governance-project.html. See also UN Global Pulse (June, 2013). Big Data for Development: A Primer. Retrieved from http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf
[108] WHO, 2014
[109] http://www.wired.com/2014/10/big-data-new-oil-or-snake-oil/
[110] Letouzé and Vinck, 2015; de Montoye, Kendall and Kerry, 2015.
[111] Imran et al. 2014a
[112] Olteanu et al., 2014
[113] Roth & Luczak-Roesch 2015
[114] Imran et al. 2014a
[115] Kerle, 2010
[116] https://data.nasa.gov/data
[117] Internal validity refers to the degree to which a causal relationship can be established between two variables in a study, technically speaking, "internal validity exists if the observed effects of the independent variable on the dependent variable are real and not caused by extraneous factors"; external validity on the other hand refers to the degree to which the findings can be generalized to the general population outside of the sample; technically speaking it is "the ability to generalize the study results to other groups and settings beyond those in the current experiment". http://www.flashcardmachine.com/internal-vsexternal-validity.html
[118] https://veri.ly/about
[119] https://www.foreignaffairs.com/articles/nepal/2015-06-01/virtual-aid-nepal
[120] Imran et al. 2014b
[121] "Sampling bias is systematic error due to a non-random sample of a population, causing some members of the population to be less likely to be included than others, resulting in a biased sample, defined as a statistical sample of a population (or non-human factors) in which all participants are not equally balanced or objectively represented. (S)ampling bias (…) undermines the external validity of a test (the ability of its results to be generalized to the rest of the population)." Source: Wikipedia
[122] For instance, as noted in Letouzé, Pestre and Zagheni, 2015, 2015, "asking 100 billionaires whether they would prefer to have dinner with friends or be sent to jail would yield results that would look very similar to those found in the general population, because income is not correlated with aversion to jail." Studying the impact of biases in mobile phone ownership on estimates of human mobility in Kenya, Wesolowski et al (2013) for example found that "that mobility estimates are surprisingly robust to the substantial biases in phone ownership across different geographical and socioeconomic groups".
[123] In other words, data from a technology that is used by 99% of the population are more likely to be representative of the general population than from one that is used by 1% of the population, although data from a representative sample of 0.1% the population will be more by design representative that the data from the 99% of the population described above.
[124] See Zagheni and Weber, 2012, Letouzé, Zagheni, Pestre et al, 2015
[125] Pham and Vinck, 2012
[126] Aldrich, 2014

[127] Gutierrez, 2011
[128] Nakagawa and Shaw, 2004
[129] Pelling and High, 2005
[130] van der Windt and Humphreys
[131] http://qz.com/420219/in-just-5-years-submarine-cables-have-brought-a-20-fold-increase-in-bandwidth-in-africa/
[132] http://www.unisdr.org/we/coordinate/hfa
[133] UNISDR, 2006protoc2006
[134] State of Humanitarian Aid, OCHA, 2015
[135] Meesters and van de Walle (2013)
[136] http://docs.hdx.rwlabs.org/hdx-is-a-year-old-our-stats-and-a-look-ahead/
[137] Weber et al., 2014
[138] Joireman et al., 2010; Howe and Leiserowitz, 2013, Akerlof et al., 2013; Krosnick, et al. 2006
[139] Pidgeon, Kasperson, & Slovic, 2003
[140] Patt and Schröter (2008)
[141] Jones et al. 2015
[142] Sterman 2011, Kahneman
[143] http://www.grameenfoundation.org/blog/lessons-learned-applab%E2%80%99s-first-three-years-uganda#.Vafpmzm8qHY
[144] WDR, 2015, Mani et al. 2013
[145] Starbird and Palen (2013)
[146] Cinnamon et al.
[147] Hughes (2014)
[148] http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html
[149] http://www.sciencemag.org/content/347/6221/536.full
[150] See notably boyd and Crawford, 2011
[151] Menlo Report Commission, 2011
[152] Pham and Vinck (2012)
[153] Letouzé and Vinck (2015)
[154] UNSG Independent Expert Advisory Group report A World that Count,
[155] Blumenstock et al.
[156] http://irevolution.net/2013/01/11/disaster-resilience-2-0/
[157] Reflections on Big Data for Conflict Prevention, 25
[158] Patrick Meier, http://irevolution.net/2013/01/11/disaster-resilience-2-0/
[159] See http://www.weforum.org/reports/data-driven-development-pathways-progress
[160] http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf
[161] http://www.undatarevolution.org/report/
[162] Meesters and van de Walle (2013)